



OPEN ACCESS

# Neighbourhood looking glass: 360° automated characterisation of the built environment for neighbourhood effects research

Quynh C Nguyen,<sup>1</sup> Mehdi Sajjadi,<sup>2</sup> Matt McCullough,<sup>3</sup> Minh Pham,<sup>4</sup> Thu T Nguyen,<sup>5</sup> Weijun Yu,<sup>6</sup> Hsien-Wen Meng,<sup>6</sup> Ming Wen,<sup>7</sup> Feifei Li,<sup>4</sup> Ken R Smith,<sup>8</sup> Kim Brunisholz,<sup>9</sup> Tolga Tasdizen<sup>2</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2017-209456>).

For numbered affiliations see end of article.

## Correspondence to

Dr Quynh C Nguyen,  
Department of Epidemiology  
and Biostatistics, University  
of Maryland School of Public  
Health, College Park, MD  
20742, USA; quynh.ctn@gmail.com

Received 12 May 2017

Revised 2 October 2017

Accepted 18 December 2017

Published Online First

15 January 2018

## ABSTRACT

**Background** Neighbourhood quality has been connected with an array of health issues, but neighbourhood research has been limited by the lack of methods to characterise large geographical areas. This study uses innovative computer vision methods and a new big data source of street view images to automatically characterise neighbourhood built environments.

**Methods** A total of 430 000 images were obtained using Google's Street View Image API for Salt Lake City, Chicago and Charleston. Convolutional neural networks were used to create indicators of street greenness, crosswalks and building type. We implemented log Poisson regression models to estimate associations between built environment features and individual prevalence of obesity and diabetes in Salt Lake City, controlling for individual-level and zip code-level predisposing characteristics.

**Results** Computer vision models had an accuracy of 86%–93% compared with manual annotations. Charleston had the highest percentage of green streets (79%), while Chicago had the highest percentage of crosswalks (23%) and commercial buildings/apartments (59%). Built environment characteristics were categorised into tertiles, with the highest tertile serving as the referent group. Individuals living in zip codes with the most green streets, crosswalks and commercial buildings/apartments had relative obesity prevalences that were 25%–28% lower and relative diabetes prevalences that were 12%–18% lower than individuals living in zip codes with the least abundance of these neighbourhood features.

**Conclusion** Neighbourhood conditions may influence chronic disease outcomes. Google Street View images represent an underused data resource for the construction of built environment features.

## INTRODUCTION

Neighbourhood characteristics influence physical and mental health, health behaviours, and the ability of individuals and families to access necessary resources for maintaining good health. Neighbourhood quality has been connected with mortality risk,<sup>1–5</sup> life expectancy,<sup>6</sup> mental health,<sup>7</sup> self-rated health, obesity<sup>8–10</sup> and diabetes<sup>11 12</sup>—even after accounting for compositional characteristics of residents. Adverse conditions concentrate in poor and

minority neighbourhoods,<sup>9 13–15</sup> thereby increasing health disparities. In a recent study of Seattle, San Diego and Baltimore, neighbourhoods with lower socioeconomic status and higher proportions of racial/ethnic groups had poorer aesthetics (eg, unmaintained buildings, graffiti, broken windows and litter). Conversely, neighbourhoods with higher socioeconomic status and lower proportions of minorities had better pedestrian amenities in terms of sidewalks, crosswalks and intersection control features.<sup>16</sup>

Currently, detailed neighbourhood data come from neighbourhood surveys, administrative data such as the census and systematic inventories of neighbourhood features. Subjective assessments of neighbourhoods from community residents help identify factors that residents believe are most important to their health and increase understanding of how individuals differentially use and interact with their environment.<sup>17</sup> However, self-reported neighbourhood data can be encumbered by several biases, including same-source bias (when two variables appear correlated because information was gathered using self-reported data on both variables) and social desirability bias (inaccuracies caused by a desire by the respondent to give answers that they think will be viewed favourably).<sup>18 19</sup>

Neighbourhood audits of built environmental features have heretofore been conducted using costly, time-consuming onsite visits or manual annotation of samples of street segment images. The development of data algorithms that can automatically process street segments to create indicators of the physical environment would dramatically lower costs and offer new resources for neighbourhood research. In this study, we leverage publicly available and pervasive Google Street View images to derive indicators of neighbourhood built environments. Google Street View had captured 20 petabytes of data, equivalent to 5 million miles of road around the world.<sup>20</sup> We limit this study to the USA. We obtained permission from Google Street View to use their image data for this research project.

## Study aims

The purpose of this study is to construct built environment indicators using computer vision techniques and publicly available Google Street View images, and to examine relationships between neighbourhood built environments, demographic



**To cite:** Nguyen QC, Sajjadi M, McCullough M, et al. *J Epidemiol Community Health* 2018;**72**:260–266.

characteristics of residents and health outcomes. We chose the following three cities that differ in landscape characteristics, population density and demographic composition: Salt Lake City, Utah; Chicago, Illinois; and Charleston, West Virginia.

## METHODS

### Street view image collection

We built algorithms to process image data for Chicago, Charleston and Salt Lake City between December 2016 and February 2017. Using Google's Street View Image API, we obtained image data on all road intersections and a random sample of street segments approximately 50 m apart. At each search location, we obtained street view images to the west, east, north and south to capture features of the environment. Image resolution was 640×640 pixels. In total, 227 000 images were obtained for Chicago, 150 000 for Salt Lake City and 53 000 for Charleston.

### Image data processing

Convolutional neural networks<sup>21</sup> achieve state-of-the-art accuracy for several computer vision tasks including object recognition, object detection and scene labelling.<sup>22–24</sup> To create a training dataset for our computer vision models, we manually annotated 14 000 images. Each image had three binary labels for the following neighbourhood characteristics: (1) presence of a crosswalk (yes/no), (2) building type (single-family detached house vs other) and (3) street greenness/landscaping (street trees and street landscaping comprised at least 30% of the image (yes/no). A cut-point of approximately 30% was used to assist with inter-rater reliability in manual annotations of street greenness. Moreover, we found that most images had some street landscaping and aimed to create a neighbourhood indicator to distinguish between ample and sparse street landscaping. Images were manually labelled by four of the authors. Inter-rater agreement was above 85% for all neighbourhood indicators (86%–96% for crosswalks; 85%–90% for building type; 85%–95% for green30). We randomly divided each labelled image dataset into a training set (80%), used to calibrate the model, and a test set (20%), used to evaluate the trained model's accuracy. We used a deep convolutional network (Visual Geometry Group 16 model) that is commonly used for object recognition. Three separate networks were trained, one for each indicator of interest. We obtained 84.59%, 85.40% and 93.03% accuracy for the 'commercial buildings/apartments,' 'green-30%' and 'crosswalk' recognition tasks, respectively.

We next experimented with semisupervised learning. In addition to the labelled dataset, we used an unlabelled set containing 226 879 images captured from Google Street View of different neighbourhoods in Chicago. The following steps were used to process the unlabelled images. We resized each unlabelled image to 256×256 and then replicate the image four times. Next, we cropped each replication at a random location into 224×224. Except the first replication, we also matched the histogram of each replicated image to the next three unlabelled images. This approach increased the accuracy to 86.07%, 86.55% and 93.37% for the 'commercial buildings/apartments,' 'green-30%' and 'crosswalk' recognition tasks, respectively. (Note: we were unable to leverage unlabelled images in the deep convolutional network approach because unlabelled images can only be used with semisupervised learning approaches.) Image values of neighbourhood features (eg, crosswalks) were then aggregated to produce small area summaries. We used zip code boundaries as a working definition of neighbourhoods. The neighbourhood built environment datasets developed for this manuscript can be

downloaded at <https://hashtaghealth.github.io/geoportals/start.html>

### Individual-level health data

We merged our neighbourhood data with individual outcome data from the Utah Population Database (UPDB). Most families in Utah are represented in the UPDB. The central component of UPDB is the extensive set of Utah family histories with links to demographic and medical data. The UPDB contains 19 million records for over 7.3 million individuals. The UPDB is a dynamic database updated annually by data contributors. Multiple records for each individual are linked by a person ID. The UPDB can only be used for biomedical and health-related research. Confidentiality and privacy of individuals are strictly protected. The UPDB is provided by the Pedigree and Population Resource under the direction of Dr Ken Smith.

### Health outcomes

#### Cohort of adults

Our analytic sample consisted of adults aged 20 years and older in 2015 living in Salt Lake City.

#### Obesity

Obesity was defined dichotomously as body mass index (BMI) ( $\text{kg}/\text{m}^2$ )  $\geq 30$  mg/dL. Self-reported height and weight were extracted from driver licence data. Utah law requires individuals to renew their driver licence every 5 years and to renew in person every 10 years. BMI data were drawn from the most recent driver licence data available, between 2011 and 2017.

*Diabetes* was assessed via diagnoses by a healthcare provider using International Classification of Diseases, Ninth Revision (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'). Sources of the clinic data included the University of Utah Health Sciences Center Enterprise Data Warehouse and Utah's Department of Health inpatient and ambulatory surgery data.

#### Covariates

We included individual sociodemographic characteristics along with health/disease predispositions to adjust for potential confounding of the relationship between neighbourhood environments and adult chronic disease. Covariates included age (years), sex (male/female), marital status (married vs divorced/single), non-White race, ethnicity and education (less than high school, high school, some college, college or greater). We used education information appearing on the Utah birth certificates of their children (if any).

#### Zip code characteristics

We used zip code boundaries as a working definition of neighbourhoods. To examine how Google-derived neighbourhood variables relate to more traditional neighbourhood variables, we merged our image dataset with 5-year estimates (2011–2015) from the American Community Survey which comprised the following zip code-level characteristics: population density, percentage of the population 65 years and older, percentage of Hispanics, percentage of blacks, median household income and percentage of householder living in current residence for 5 or

**Table 1** Descriptive characteristics of neighbourhood characteristics

	Green streets	Crosswalk present	Commercial/apartment building
	Mean (SD)	Mean (SD)	Mean (SD)
Salt Lake City	59.0 (49.2)	8.0 (27.1)	38.5 (48.7)
Chicago, Illinois	71.2 (45.3)	22.5 (41.8)	55.8 (49.7)
Charleston, West Virginia	78.6 (41.0)	3.4 (18.1)	44.9 (49.7)
N	2 26875	1 50 300	53 360

Neighbourhood characteristics derived from street images collected between December 2016 and February 2017 from Google's Street View Image API.

more years. We investigated property values as a covariate but excluded it from the model because it was highly correlated ( $r=0.93$ ) with median household income.

**Analytic approach**

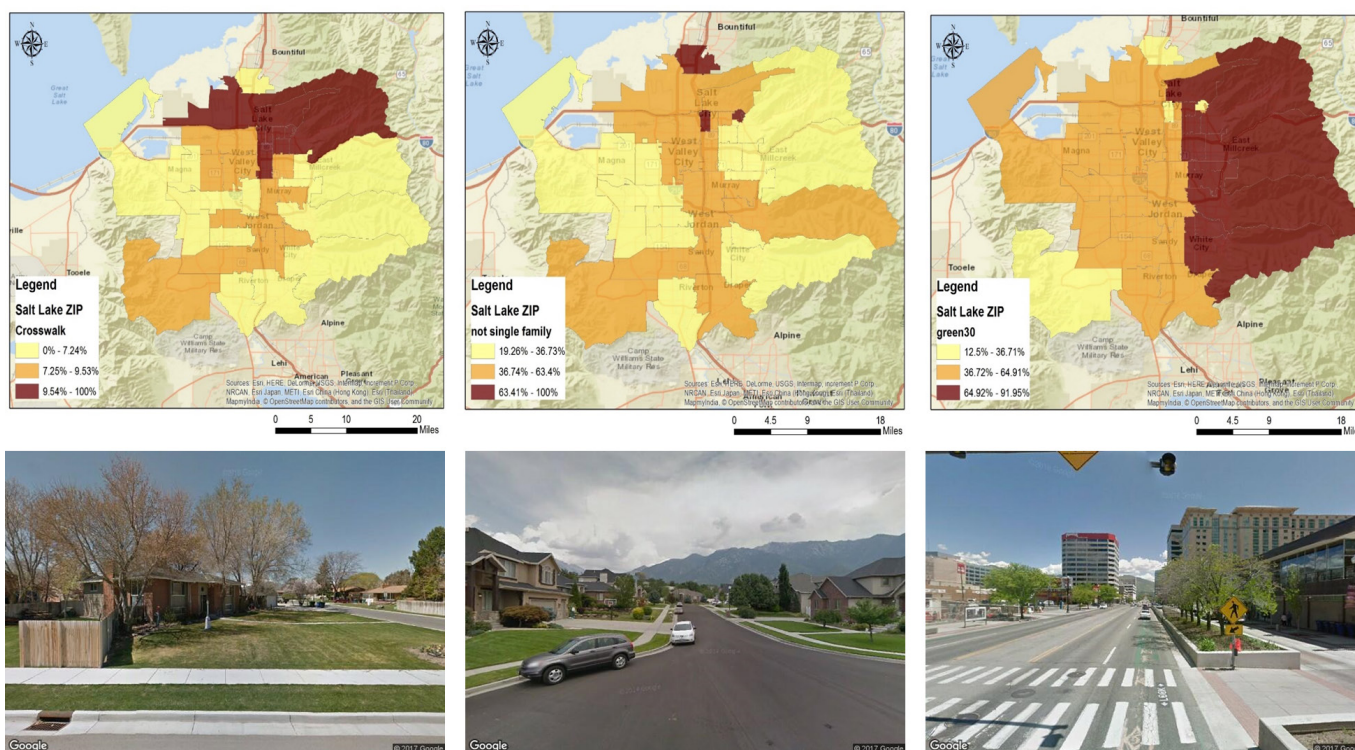
Choropleth maps were created using ArcGIS Desktop software (ESRI) and the 2016 U.S. Census TIGER/Line Shapefiles. We used tertile breaks in displaying spatial data patterns. A one-to-many spatial join was performed with the image locations and zip codes in order to aggregate the image results into geographical areas. We implemented adjusted linear regression models to examine associations between area-level built environment characteristics and other area-level characteristics (ie, demographics and economic characteristics). 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. We implemented log Poisson regression models to estimate prevalence ratios (95% CIs)

representing associations between tertiles of built environment characteristics and individual chronic disease prevalence. SEs were adjusted for clustering of values at the zip code level. Statistical analyses were implemented with Stata MP V.13 (Stata).

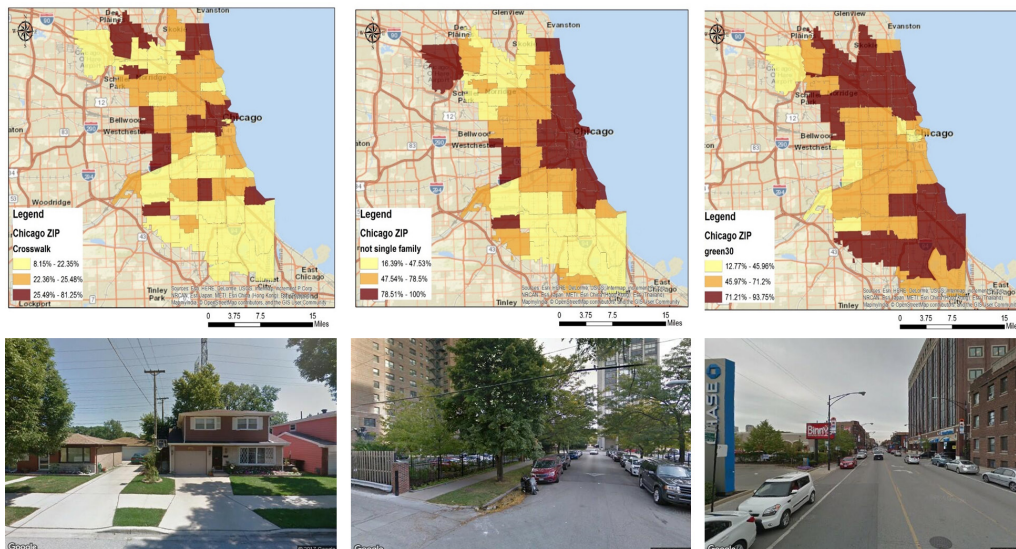
**RESULTS**

There was much variation in the built environment characteristics across the three cities (table 1). Charleston had the highest percentage of green streets (79%), while Chicago had the highest percentage of streets with crosswalks (23%) and commercial buildings/apartments (56%). Figures 1–3 display the zip code distribution of the three built environment characteristics in the three cities as well as sample street view images. (Note: full-page figures are also reproduced in the online supplemental files, Figures S1–S9). In Salt Lake City, crosswalks were most common in central regions, commercial buildings/apartments were found in high density in a few small areas and street greenness was abundant in the eastern region. In Chicago, crosswalks and commercial buildings/apartments were prevalent across the city, with somewhat lower percentages for the most southern areas. In Chicago, green space was prominent in the most northern and southern areas. In Charleston, crosswalks and green space could be found across the region, while commercial buildings/apartments were concentrated in the south-western region.

In Salt Lake City, greater population density was marginally related to fewer commercial buildings/apartments and more green streets (see online supplemental table S1). Areas with more senior citizens had more green streets in Salt Lake City. Percentage of blacks was correlated with *more* crosswalks and more commercial buildings/apartments. Percentage of Hispanics was correlated with *fewer* crosswalks and *fewer* commercial buildings/apartments (online supplemental table S1). In Chicago,



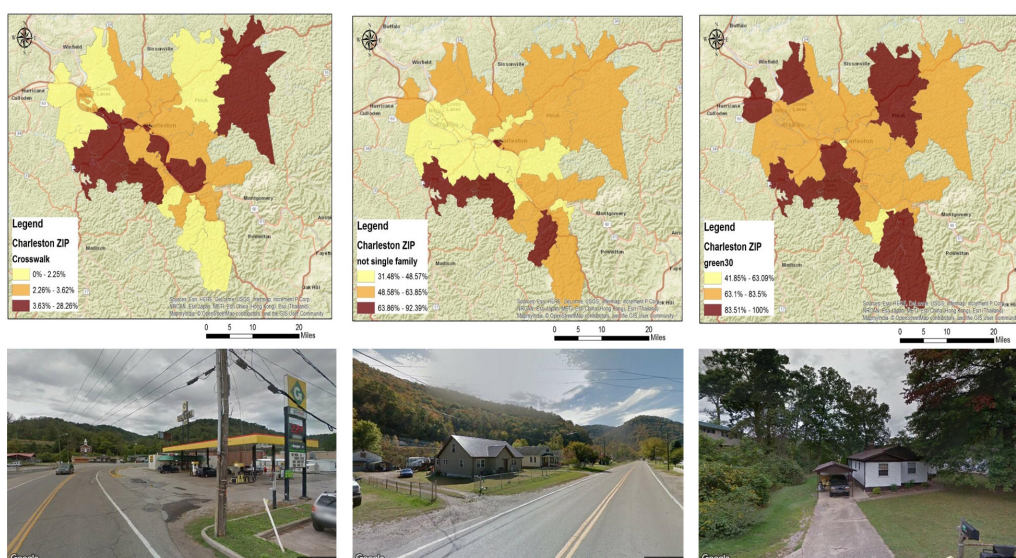
**Figure 1** Zip code distribution of built environment characteristics in Salt Lake City, Utah. For each zip code, the following was calculated: percentage of street images with (A) presence of a crosswalk, (B) building type other than a single-family detached house and (C) street landscaping comprising at least 30% of the image. Neighbourhood characteristics were categorised into tertiles and spatially mapped. Darker colours represent higher tertile values. Data source: Google Street View images.



**Figure 2** Zip code distribution of built environment characteristics in Chicago, Illinois. For each zip code, the following was calculated: percentage of street images with (A) presence of a crosswalk, (B) building type other than a single-family detached house and (C) street landscaping comprising at least 30% of the image. Neighbourhood characteristics were categorised into tertiles and spatially mapped. Darker colours represent higher tertile values. Data source: Google Street View images.

zip codes with more seniors and higher median household income had greener streets and fewer commercial buildings/apartments (online supplemental table S2). In Charleston, population density was negatively associated with green streets. Zip codes with higher population density and higher percentage of young residents were less likely to have commercial buildings/apartments. Percentage of blacks was related to more commercial buildings/apartments and marginally more crosswalks (see online supplemental table S3). Analyses examining relationships at the census tract level generally showed similar associations as seen at the zip code level (online supplemental tables S4–S6), although at the tract level, it becomes more apparent that higher median household income was associated with more green streets and fewer commercial buildings/apartments.

Merging built environment indicators in Salt Lake City with individual-level health outcomes from the Utah Population Database, we examined the potential influences of neighbourhood characteristics on adult obesity and diabetes. Comparing the third tertile with the first tertile, more green streets (prevalence ratio (PR)=0.73; 95% CI 0.63 to 0.85), crosswalks (PR=0.76; 95% CI 0.69 to 0.85) and presence of commercial buildings/apartments (PR=0.79; 95% CI 0.67 to 0.94) were associated with lower individual obesity prevalence (table 2). More green streets (PR=0.86; 95% CI 0.77 to 0.96), crosswalks (PR=0.87; 95% CI 0.80 to 0.95) and presence of commercial buildings/apartments (PR=0.81; 95% CI 0.67 to 0.98) were also associated with lower individual diabetes prevalence (table 2). In sensitivity analyses, we adjust for birth weight, a predisposing factor



**Figure 3** Zip code distribution of built environment characteristics in Charleston, West Virginia. For each zip code, the following was calculated: percentage of street images with (A) presence of a crosswalk, (B) building type other than a single-family detached house and (C) street landscaping comprising at least 30% of the image. Neighbourhood characteristics were categorised into tertiles and spatially mapped. Darker colours represent higher tertile values. Data source: Google Street View images.

**Table 2** Built environment predictors of adult obesity and diabetes,\* Salt Lake City

Built environment characteristics	Obese	Diabetes
	Prevalence ratio (95% CI)†	Prevalence ratio (95% CI)†
Green streets		
Third tertile (highest)	0.73 (0.63 to 0.85)	0.86 (0.77 to 0.96)
Second tertile	0.99 (0.92 to 1.06)	1.03 (0.97 to 1.08)
Crosswalks		
Third tertile (highest)	0.76 (0.69 to 0.85)	0.87 (0.80 to 0.95)
Second tertile	1.02 (0.97 to 1.07)	1.01 (0.95 to 1.06)
Commercial buildings/apartments		
Third tertile (highest)	0.79 (0.67 to 0.94)	0.81 (0.67 to 0.98)
Second tertile	0.93 (0.86 to 1.01)	0.91 (0.84 to 0.99)
N	727 737	736 218

\*Data source for health outcomes: Utah Population database.

†Adjusted Poisson models were run for each outcome separately. Models controlled for individual-level age, sex, race, ethnicity, education and marital status as well as zip code-level population density, percentage of the population 65 years and older, percentage of Hispanics, percentage of blacks, median household income and percentage of householder living in current residence for 5 years or more. Built environment characteristics were categorised into tertiles, with the lowest tertile serving as the referent group. SEs were adjusted for clustering of values at the zip code level.

for obesity and diabetes.<sup>25 26</sup> Results were qualitatively similar (online supplemental table S7), but the sample size was smaller because analyses used only adults born in Utah. Analyses using continuous BMI also suggest that green streets and crosswalks were related to lower BMI (online supplemental table S8).

## DISCUSSION

In this study, we demonstrate the feasibility and accuracy of computer vision models for automatic detection and characterisation of neighbourhood built environment features. Moreover, in piloting our algorithms across three visually and geographically distinct US cities, we found that the models were generally robust to varying landscapes. Group demographic characteristics correlated with built environment features, either suggesting differential access or preference for different types of neighbourhoods. Greater median family income was related to greener streets and fewer commercial buildings/apartments. In Chicago, greater population density was related to more commercial buildings/apartments, but in Charleston and Salt Lake City, the opposite was true, and suggested different forms of dominant housing and building types for more populous areas. Furthermore, in Salt Lake City, we found that green streets, crosswalks and commercial buildings/apartments were associated with lower individual prevalence of obesity and diabetes, controlling for predisposing characteristics.

### Study findings in context

Previous studies have used Google Street View to perform virtual audits to assess walkability,<sup>27</sup> physical disorder,<sup>28</sup> retail alcohol stores<sup>29</sup> and urban greenery<sup>30</sup> among other neighbourhood characteristics. Researchers have also developed a Computer Assisted Neighborhood Visual Assessment System to allow users to conduct virtual neighbourhood audits of Google Street View images.<sup>31</sup> Studies comparing the results of onsite audits and virtual audits have found good agreement across a range of measures.<sup>32 33</sup> A study of 37 block faces in New York City found high levels of agreement for pedestrian safety, motorised traffic, parking and infrastructure. Assessment for physical disorder and

neighbourhood characteristics that had fine details or temporal variability had lower levels of agreement.<sup>34</sup>

Moreover, research teams have begun to use automated approaches to extract neighbourhood features from Google Street View images such as pedestrian count<sup>35</sup> and visual enclosure (ie, proportion sky visible from a point on the street)<sup>36</sup>—measures associated with walkability. In this paper, we use street view images to extract additional neighbourhood characteristics and investigate their associations with chronic disease. Our findings are consistent with prior literature. Mix land uses, connected streets, higher residential density and attractive scenery promote physical activity.<sup>37–39</sup>

### Study strengths and limitations

The neighbourhood built environment can impact health and health behaviours by locating either risks or resources for health. Thus far, characterising and investigating built environment features has generally been limited to local studies due to the resource-intensive nature of (1) site visits to create community inventories of features and (2) manual annotations of street images. In this study, we leveraged recent advancements in computer vision and the emergence of massive sources of street image data. We developed a data collection strategy using geographic information systems to assemble a collection of street intersection and street segment images for three cities in the USA. We implemented computer vision algorithms to produce neighbourhood summaries of built environment that have been theoretically and empirically identified to be important for health outcomes. We then investigated the potential impact of neighbourhood environments on chronic diseases using administrative records and medical records from a large healthcare provider, accounting for predisposing characteristics.

Nonetheless, our study has limitations. Like other modes of data collection, image data can only capture a subset of community features. For instance, we are unable to create indicators for noise (eg, hearing gunshots) and witnessing criminal activity. Even when the goal is a built environment characteristic, not all types of characteristics lend themselves to easy extraction from images and processing by computer vision algorithms. In preliminary analyses, we attempted to construct indicators of litter, but manual annotations of street images had low reliability between raters given the small size of litter and its resemblance to other small objects such as leaves. We also had difficulty obtaining enough training data for dilapidated buildings and graffiti because of their relative rarity. Ratings of neighbourhood aesthetics ('beauty') varied greatly between raters. Characterising subjective qualities of a neighbourhood may require amassing resident and visitor perceptions to capture variability. César Hidalgo, Ramesh Raskar and colleagues used crowdsourcing techniques to obtain labelled data on perceived safety of street views from cities across the USA and from these labelled data, they built machine learning algorithms to predict how safe the street image looks to a human observer.<sup>40</sup>

Besides the *type* of neighbourhood features that can easily lend themselves to automatic extraction via computer vision models, the *depth* of neighbourhood features that can be extracted may be limited. Neighbourhood inventories created from onsite visits provide a wealth of fine-grain information about hundreds of different features. Building a computer vision model to accurately extract each of these hundreds of features would be a difficult task. Well-known neighbourhood audit instruments include the Irvine-Minnesota Inventory,<sup>41</sup> the Pedestrian Environment Data Scan<sup>42</sup> and the Maryland Inventory of Urban Design Qualities.<sup>43</sup>

Another limitation was that analyses were cross-sectional, and thus the study was unable to evaluate longitudinal trends. The analyses did not take into account residential histories and the length of time individuals lived in the communities. We find links between built environment characteristics and chronic disease prevalence; however, our study was unable to identify the direction of causality. For instance, study participants' chronic disease diagnoses may have preceded their time in a neighbourhood or their diagnoses may have prompted them to move into a different neighbourhood.

## CONCLUSIONS

The epidemic rise in chronic health conditions is recent. This suggests that its cause is social, cultural and constructed rather than purely biological. Nonetheless, the dearth of data on contextual factors limits understanding of the influence of neighbourhood characteristics on health. We harness the underused potential of street image data to characterise built environments. Our substantive investigation of the impact of built neighbourhood characteristics on chronic conditions suggests that green streets, crosswalks and mix building types were related to lower obesity and diabetes prevalence. Results and data products can be used by city agencies and public health practitioners to inform strategies for improving community health.

### What is already known on this subject

Neighbourhood characteristics have been associated with mental and physical health outcomes including chronic disease. However, the bulk of the literature has focused on small geographical locations due to the time and cost demands of neighbourhood audits of built environment characteristics.

### What this study adds

In this study, we demonstrate the feasibility and accuracy of computer vision models for the automatic detection and characterisation of neighbourhood built environment features across three visually and geographically distinct cities. Investigating health outcomes in Salt Lake City, we found that green streets, crosswalks and commercial buildings/apartments were associated with lower individual obesity and diabetes prevalence, controlling for individual-level and zip code-level predisposing characteristics.

### Author affiliations

<sup>1</sup>Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, Maryland, USA

<sup>2</sup>Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah, USA

<sup>3</sup>Department of Geography, University of Utah, Salt Lake City, Utah, USA

<sup>4</sup>School of Computing, University of Utah, Salt Lake City, Utah, USA

<sup>5</sup>Department of Epidemiology and Biostatistics, University of California San Francisco School of Medicine, San Francisco, California, USA

<sup>6</sup>Department of Health, Kinesiology, and Recreation, University of Utah, Salt Lake City, Utah, United States

<sup>7</sup>Department of Sociology, University of Utah, Salt Lake City, Utah, USA

<sup>8</sup>Department of Family and Consumer Studies and Population Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA

<sup>9</sup>Institute for Healthcare Delivery Research, Intermountain Healthcare, Salt Lake City, Utah, USA

**Acknowledgements** We thank Dr James VanDerslice and Heidi Hanson for their guidance on the project.

**Contributors** QCN took the lead in designing the study, implementing analyses and writing the manuscript. MS and TT devised and implemented computer vision approaches to process the Google Street View images. MM and MP obtained Google Street View images via the Google Street View API, mapped the data and edited the manuscript. TXN, WY, H-WM, MW, FL, KRS and KB assisted with the design of the study and edited the manuscript.

**Funding** This work was supported by NIH grants 5K01ES025433 and 3K01ES025433-03S1 (Dr Nguyen, PI).

**Competing interests** None declared.

**Patient consent** Our study involved analyses of de-identified records.

**Ethics approval** The University of Utah Institutional Review Board approved the study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Zip code and census tract level indicators of the built environment indicators developed for this manuscript can be downloaded at our project website: <https://hashtaghealth.github.io/geoportal/start.html>

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

- 1 Wing S, Barnett E, Casper M, *et al*. Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women. *Am J Public Health* 1992;82:204–9.
- 2 Tyroler HA, Wing S, Knowles MG. Increasing inequality in coronary heart disease mortality in relation to educational achievement: profile of places of residence, United States, 1962–87. *Annals of Epidemiology* 1993;3:S51–S4.
- 3 Morris JN, Blane DB, White IR, mortality Lof. Levels of mortality, education, and social conditions in the 107 local education authority areas of England. *J Epidemiol Community Health* 1996;50:15–17.
- 4 Eames M, Ben-Shlomo Y, Marmot MG. Social deprivation and premature mortality: regional comparison across England. *BMJ* 1993;307:1097–102.
- 5 Townsend P, Phillimore P, Beattie A. *Health and deprivation: inequality and the North*. London, England: Routledge, 1988.
- 6 Clarke CA, Miller T, Chang ET, *et al*. Racial and social class gradients in life expectancy in contemporary California. *Soc Sci Med* 2010;70:1373–80.
- 7 Truong KD, Ma S. A systematic review of relations between neighborhoods and mental health. *J Ment Health Policy Econ* 2006;9:137–54.
- 8 Mujahid MS, Diez Roux AV, Shen M, *et al*. Relation between neighborhood environments and obesity in the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 2008;167:1349–57.
- 9 Black JL, Macinko J, Dixon LB, *et al*. Neighborhoods and obesity in New York City. *Health Place* 2010;16:489–99.
- 10 Heinrich KM, Lee RE, Regan GR, *et al*. How does the built environment relate to body mass index and obesity prevalence among public housing residents? *Am J Health Promot* 2008;22:187–94.
- 11 Lysy Z, Booth GL, Shah BR, *et al*. The impact of income on the incidence of diabetes: a population-based study. *Diabetes Res Clin Pract* 2013;99:372–9.
- 12 Grigsby-Toussaint DS, Lipton R, Chavez N, *et al*. Neighborhood socioeconomic change and diabetes risk: findings from the Chicago childhood diabetes registry. *Diabetes Care* 2010;33:1065–8.
- 13 Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216–22.
- 14 Macintyre S, Maciver S, Sooman A, Area SA. Area, Class and Health: should we be focusing on places or people? *J Soc Policy* 1993;22:213–33.
- 15 Duncan C, Jones K, Moon G, Context MG. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;46:97–117.
- 16 Thornton CM, Conway TL, Cain KL, *et al*. Disparities in pedestrian streetscape environments by income and race/ethnicity. *SSM Popul Health* 2016;2:206–16.
- 17 Blacksher E, Lovasi GS. Place-focused physical activity research, human agency, and social justice in public health: taking agency seriously in studies of the built environment. *Health Place* 2012;18:172–9.
- 18 Raudenbush SW, Sampson RJ. Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Social Methodol* 1999;29:1–41.
- 19 Sampson RJ, Raudenbush SW. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *Am J Sociol* 1999;105:603–51.

- 20 Farber D. Google takes Street View off-road with backpack rig. *CNET* 2012.
- 21 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 2012;1097–105.
- 22 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv* 2014;14091556.
- 23 Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2015:1–9.
- 24 Zhou B, Lapedriza A, Xiao J, et al. Learning deep features for scene recognition using places database. *Advances in neural information processing systems* 2014:487–95.
- 25 Yu ZB, Han SP, Zhu GZ, et al. Birth weight and subsequent risk of obesity: a systematic review and meta-analysis. *Obes Rev* 2011;12:525–42.
- 26 Davey RX. Birthweight and risk for diabetes. *Diabetes Care* 2002;25:25.
- 27 Brookfield K, Tilley S. Using virtual street audits to understand the walkability of older adults' route choices by gender and age. *Int J Environ Res Public Health* 2016;13:1061.
- 28 Mooney SJ, Bader MD, Lovasi GS, et al. Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit. *Am J Epidemiol* 2014;180:626–35.
- 29 Less EL, McKee P, Toomey T, et al. Matching study areas using google street view: a new application for an emerging technology. *Eval Program Plann* 2015;53:72–9.
- 30 Li X-j ZC, Li W, et al. Assessing street-level urban greenery using google street view and a modified green view index. *Urban Forestry & Urban Greening* 2015;14:675–85.
- 31 Bader MD, Mooney SJ, Lee YJ, et al. Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health Place* 2015;31:163–72.
- 32 Badland HM, Opat S, Witten K, et al. Can virtual streetscape audits reliably replace physical streetscape audits? *J Urban Health* 2010;87:1007–16.
- 33 Gullón P, Badland HM, Alfayate S, et al. Assessing walking and cycling environments in the streets of madrid: comparing on-field and virtual audits. *J Urban Health* 2015;92:923–39.
- 34 Rundle AG, Bader MD, Richards CA, et al. Using google street view to audit neighborhood environments. *Am J Prev Med* 2011;40:94–100.
- 35 Yin L, Cheng Q, Wang Z, et al. 'Big data' for pedestrian volume: exploring the use of google street view images for pedestrian counts. *Appl Geogr* 2015;63:337–45.
- 36 Yin L, Wang Z. Measuring visual enclosure for street walkability: using machine learning algorithms and google street view imagery. *Appl Geogr* 2016;76:147–53.
- 37 Saelens BE, Glanz K, Sallis JF, et al. Nutrition Environment Measures Study in restaurants (NEMS-R): development and evaluation. *Am J Prev Med* 2007;32:273–81.
- 38 Humpel N, Owen N, Leslie E. Environmental factors associated with adults' participation in physical activity: a review. *Am J Prev Med* 2002;22:188–99.
- 39 Zuniga-Teran AA, Orr BJ, Gimblett RH, et al. Neighborhood design, physical activity, and wellbeing: applying the walkability model. *Int J Environ Res Public Health* 2017;14:76.
- 40 Naik N, Kominers SD, Raskar R, et al. Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major US cities. *National Bureau of Economic Research* 2015.
- 41 Day K, Boarnet M, Alfonzo M, et al. The Irvine-Minnesota inventory to measure built environments: development. *Am J Prev Med* 2006;30:144–52.
- 42 Clifton KJ, Livi Smith AD, Rodriguez D. The development and testing of an audit for the pedestrian environment. *Landsc Urban Plan* 2007;80:95–110.
- 43 Ewing R, Handy S, Brownson RC, et al. Identifying and measuring urban design qualities related to walkability. *J Phys Act Health* 2006;3:S223–S240.