



OPEN ACCESS

Leveraging computer vision for predicting collision risks: a cross-sectional analysis of 2019–2021 fatal collisions in the USA

Quynh C. Nguyen ,¹ Mitra Alirezaei,² Xiaohe Yue,¹ Heran Mane,¹ Dapeng Li,³ Lingjun Zhao,⁴ Thu T Nguyen,¹ Rithik Patel,¹ Weijun Yu ,¹ Ming Hu ,⁵ D. Alex Quistberg ,^{6,7} Tolga Tasdizen²

¹Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, Maryland, USA

²Department of Electrical and Computer Engineering, Scientific Computing and Imaging Institute, The University of Utah, Salt Lake City, Utah, USA

³Department of Geography and the Environment, The University of Alabama, Tuscaloosa, Alabama, USA

⁴Department of Computer Science, University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland, USA

⁵School of Architecture, University of Notre Dame, Notre Dame, Indiana, USA

⁶Urban Health Collaborative, Drexel University School of Public Health, Philadelphia, Pennsylvania, USA

⁷Department of Environmental and Occupational Health, Drexel University School of Public Health, Philadelphia, Pennsylvania, USA

Correspondence to

Dr Quynh C. Nguyen, Department of Epidemiology & Biostatistics, University of Maryland School of Public Health, College Park, MD 20742, USA; qtnghuyen@umd.edu

Received 23 October 2023
Accepted 19 April 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Nguyen QC, Alirezaei M, Yue X, *et al.* *Inj Prev* Epub ahead of print: [please include Day Month Year]. doi:10.1136/ip-2023-045153

ABSTRACT

Objective The USA has higher rates of fatal motor vehicle collisions than most high-income countries. Previous studies examining the role of the built environment were generally limited to small geographic areas or single cities. This study aims to quantify associations between built environment characteristics and traffic collisions in the USA.

Methods Built environment characteristics were derived from Google Street View images and summarised at the census tract level. Fatal traffic collisions were obtained from the 2019–2021 Fatality Analysis Reporting System. Fatal and non-fatal traffic collisions in Washington DC were obtained from the District Department of Transportation. Adjusted Poisson regression models examined whether built environment characteristics are related to motor vehicle collisions in the USA, controlling for census tract sociodemographic characteristics.

Results Census tracts in the highest tertile of sidewalks, single-lane roads, streetlights and street greenness had 70%, 50%, 30% and 26% fewer fatal vehicle collisions compared with those in the lowest tertile. Street greenness and single-lane roads were associated with 37% and 38% fewer pedestrian-involved and cyclist-involved fatal collisions. Analyses with fatal and non-fatal collisions in Washington DC found streetlights and stop signs were associated with fewer pedestrians and cyclists-involved vehicle collisions while road construction had an adverse association.

Conclusion This study demonstrates the utility of using data algorithms that can automatically analyse street segments to create indicators of the built environment to enhance understanding of large-scale patterns and inform interventions to decrease road traffic injuries and fatalities.

INTRODUCTION

Each year, 1.19 million people are killed on roadways around the world.¹ Globally, motor vehicle collisions are the 12th-leading cause of death overall and the leading cause of death for young people aged 5–29 years.¹ The USA also has higher rates of fatal collisions than most high-income countries.² Reduction in collision rates would have powerful societal impacts by protecting young people and ensuring their safety on the roads so that they can contribute economically, politically and socially to their communities.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The USA has higher rates of fatal vehicle collisions than most other high-income countries. However, there are few studies that systematically identify specific features of the built environment that contribute to motor vehicle collisions and pedestrian injuries and fatalities across geographical areas larger than cities.

WHAT THIS STUDY ADDS

⇒ This study used our national collection of Google Street View images and computer vision models to extract built environment features and to examine how the built environment may impact collision risk.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study has implications for public health and urban planning working to create environments that foster health and reduce road traffic collisions at the population level. Results from this study can guide population-based strategies to improve the safety of roadways.

Poor road infrastructure and neighbourhood design are important contributors to rising numbers of road traffic injuries and deaths,^{3,4} but most studies examining the role of the built environment are limited to smaller geographical areas and often only certain locations within cities due to the challenges obtaining these data.^{5–8} Most frequently, detailed neighbourhood data come from neighbourhood surveys, administrative data (such as census data) and in-person or virtual audits of the built environment for small areas.⁹ Prior studies with data on a select few neighbourhoods or cities may not be generalisable or relevant to neighbourhoods across the USA.

This study advances public health research and practice by producing national built environment indicators of motor vehicle collision risk using computer vision models. Previous studies have used Google Street View (GSV) images for characterising built environment features (eg, walkability indicators), emphasising the strength of using GSV imagery in facilitating large-scale studies.^{10,11} For example, Quistberg *et al* used GSV images and

trained neural networks to successfully identify built environment features relevant to pedestrian safety such as medians, crosswalks and pedestrian signals.^{12 13}

Study aims and hypotheses

This study aims to examine whether built environment characteristics derived from GSV images are related to motor vehicle collisions in the USA. We hypothesise that areas with street designs that safeguard pedestrian and cyclist movements and possess speed-reducing features, will have fewer vehicle collisions and fewer collision-related injuries and fatalities.

METHODS

Study sample

Motor vehicle collision outcomes

Fatal motor vehicle collision data came from the 2019, 2020 and 2021 Fatality Analysis Reporting System (FARS) national datasets produced by the National Highway Traffic Safety Administration.¹⁴ FARS is a national yearly census of fatal motor vehicle traffic collisions. Collision-related outcomes examined include (1) total collisions, (2) total fatalities (i.e., total number of deaths across collisions), (3) total vehicles involved in collisions and (4) total collisions involving pedestrians and cyclists. To enable reliable estimates for census tracts across the USA, we pooled fatal collision data from 2019 to 2021 and calculated the annual average number of such collisions per 10 000 population for each census tract.

To enable examination of both fatal and non-fatal motor vehicle collisions, 2019–2022 Washington DC motor vehicle collision data were obtained from the District Department of Transportation.¹⁵

Demographic and socioeconomic data

The analyses accounted for census tract median age, per cent male, per cent Black, per cent Hispanic, per cent owner-occupied housing and Child Opportunity Index (composite of 29 neighborhood-level indicators in the areas of education; health and environment; social and economic conditions).¹⁶ All variables had correlations of <0.70. Covariate information was obtained from the American Community Survey (ACS) 2018 5-year estimates, except for population size, which was obtained from the 2010 US Census.

Street view image collection

We used the Street View Static API to collect GSV images. Sampling points were generated for all primary and secondary roads, street intersections, and locations along road segments at 100 m intervals. For the sampling point, GSV images from four directions (facing west, east, north and south) were collected to capture 360° angles of the built environment. In total about 164 million images were collected from across the USA. Dates of the available street view images varied immensely (median year=2012; date range 2007–2019). For analyses involving Washington DC, a subset of 103 476 images were used. We performed spatial join analysis to determine the census tract for each GSV image based on the image's geocoordinates.

Built environment indicators

Four indicators of the built environment were selected that have been theoretically and empirically linked to vehicle collision risk (sidewalks, streetlights, street greenness, and single-lane roads) and can be robustly detected with computer vision models.

Sidewalks give pedestrians a dedicated safe space on which to travel, separate from high-speed vehicles.¹⁷ Streetlights can increase safety of pedestrians and cars at night by increasing visual awareness of surroundings, road conditions and weather changes.¹⁸ Street trees create a visual narrowing of the roadway thereby encouraging reduced speeds.^{19 20} Street greenness can help redirect a driver's attention, especially for long-distance driving²¹ and can act as a barriers protecting pedestrians.²² Another built environment indicator, single-lane roads limit the amount of vehicular traffic and can reduce speed variability, lane changes and collisions caused by overtaking and multiple-lane pedestrian crossings.²³ Road construction and work zones can increase collision rates²⁴ with some studies reporting a twofold increase.^{25 26} Factors that increase collision risk in construction zones can include narrower lanes; unexpected detours and complicated road geometries.^{24 27 28}

Image data analysis

The following built environment characteristics were derived in image analysis including the presence of a (1) sidewalk (at least one side of the road); (2) streetlight; (3) single-lane road; (4) road construction and (5) street greenness (if $\geq 30\%$ of image was of street trees or other landscaping). Most street view images had some street greenness so this indicator was used to distinguish between sparse and more amble street greenery. Additionally, this threshold achieved inter-rater reliabilities >85% in manual annotations of images.²⁹

To create the training and test datasets for the computer vision models, 18 000 GSV images from the national data collection were manually annotated. Labelers included the principal investigator and three graduate research assistants. Inter-rater agreement was above 85% for all neighbourhood indicators. This dataset was then divided into a training (80%) and test set (20%). A standard deep convolutional neural network architecture, Visual Geometry Group-19³⁰ or ResNet-18,³¹ in TensorFlow³² was trained with sigmoid cross entropy with logits as the loss function. The accuracy of the recognition tasks (agreement between manually labelled images and computer vision predictions) was as follows: sidewalk (84%); streetlight (88%); street greenness (89%); single-lane road (88%) and road construction (96%).

Due to time and resource constraints, two additional built environment characteristics were extracted: (1) presence of stop signs and (2) road construction for Washington DC images only. Fifteen hundred images were annotated from the Washington DC street view dataset, and randomly split into 80% for training, 10% for validation and 10% for testing. A Visual Question Answering model (Vision-and-Language Transformer Without Convolution or Region Supervision, ICML 2021) was fine-tuned. After applying a learning rate of 1e-6 and 30 epochs, the model was selected based on the best accuracy on the validation set. The model achieved an F1 score of 88.9%, and an accuracy of 90.0% on the test set.

Statistical analyses

GSV images were collected every 100 m on road networks, thus census tracts with more road networks would have more GSV images collected. To account for this variation, we calculated, for each census tract, the percentage of total number of images that contained a given built environment indicator (e.g., per cent with a street light = (number of images with a streetlights/total number of images) $\times 100$).

Table 1 Descriptive statistics of neighbourhood characteristics and motor vehicle-related collisions, census tract

	N	Mean (SD)
Built environment characteristics		
% with presence of a sidewalk	70359	43.96 (30.72)
% with presence of a streetlight	70318	15.66 (14.96)
% street greenness	70359	85.54 (15.39)
% with presence of a single-lane road	70359	67.11 (14.57)
USA collision outcomes		
Total number of fatal collisions	72531	3.08 (114.31)
Total number of deaths	72531	3.29 (120.21)
Total vehicles involved	72531	5.09 (201.33)
Total pedestrian-involved and cyclist-involved collisions	72531	0.66 (31.16)
Washington DC collision outcomes		
Total number of collisions (fatal and non-fatal)	154	320 (273)
Pedestrian-involved collisions	154	45.64 (46.71)
Cyclist-involved collisions	154	25.31 (39.24)
Census tract sociodemographic characteristics		
Number of individuals in census tract	72864	4237 (1973)
Median age	72578	7.01 (3.73)
Per cent male	72578	49.18 (4.05)
Per cent Black	72578	13.83 (22.29)
Per cent Hispanic	72578	15.27 (20.82)
Per cent owner-occupied housing	72472	64.32 (22.50)
Child Opportunity Index, range 0–100	72213	49.15 (28.61)

Data sources: Built environment characteristics were derived from Google Street View images from the USA in 2019. Vehicle-related collisions were obtained from the National Fatality System for vehicle collisions involving at least one fatality to produce per capita yearly averages (per 10 000 population) for the time period 2019–2021. Vehicle-related collisions were obtained from the District Department of Transportation for collisions on the roadway blocks network of Washington, DC to produce per capita yearly averages per 10 000 population for the time period 2019–2022. Census tract sociodemographics were obtained from the American Community Survey (ACS) 2018 5-year estimates and the 2015 Child Opportunity Index.

Adjusted Poisson regression models were run separately for each collision-related outcome. These models were controlled for census tract median age, per cent male, per cent Black, per cent Hispanic, per cent owner-occupied housing and Child Opportunity Index. Built environment characteristics were categorised into tertiles, with the lowest tertile serving as the

reference group. Tertiles were chosen to ease interpretation of results and allow for non-linearities in the association between area characteristics and collision-related outcomes. Rate ratios (RRs) and 95% CIs were derived from these models to represent associations between tertiles of built environment characteristics and per capita motor vehicle collisions. Statistical analyses were implemented using Stata MP V.16 (StataCorp).

RESULTS

Across census tracts, on average, about 44% of GSV images had sidewalks, 16% had streetlights, 86% were deemed green streets and 67% had single-lane roads (table 1). A little over half of census tracts (N=34 715) had at least one fatal collision between 2019 and 2021. Across all census tracts, the 3-year average of fatal collisions was 3.08 per 10 000 population. About 20% of fatal collisions involved pedestrians or cyclists (0.66 per 10 000 population). For Washington DC census tracts, on average, 320 fatal and non-fatal collisions (per 10 000 population) occurred per year over the 2019–2022 time period.

The presence of sidewalks, streetlights, street greenness and single-lane roads was associated with marked reductions in collisions-related outcomes (table 2). For example, census tracts in the third (highest) tertile for the presence of sidewalks had a 70% lower rate of fatal collisions (RR 0.30; 95% CI 0.27 to 0.33) compared with census tracts in the lowest tertile. Census tracts in the second tertile of sidewalks had a 52% lower rate of fatal collisions (RR 0.48; 95% CI 0.45 to 0.52). Census tracts in the highest tertile of single-lane roads, streetlights and street greenness had 50%, 30% and 26% lower rates of fatal collisions, respectively.

For fatal collisions involving pedestrians and cyclists, a 15%, 37% and 38% lower rate was observed in census tracts in the third tertile for sidewalks, street greenness and single-lane roads, respectively, compared with the lowest tertile. However, some adverse associations were also observed. Census tracts in the second tertile of sidewalks had 14% higher rates of fatal collisions involving pedestrians and cyclists compared with the lowest tertile. Additionally, while streetlights were associated with fewer total fatal vehicle collisions and fatalities, they were

Table 2 Census tract level built environment predictors of fatal motor vehicle-related collisions in the USA, 2019–2022

	Total number of fatal vehicle collisions	Total fatalities (number of deaths)	Total number of vehicles involved	Total number of collisions with pedestrians and cyclists involved
Built environment characteristics	Rate ratio (95% CI)	Rate ratio (95% CI)	Rate ratio (95% CI)	Rate ratio (95% CI)
Sidewalks				
3rd tertile (highest)	0.30 (0.27 to 0.33)	0.28 (0.26 to 0.31)	0.30 (0.28 to 0.33)	0.85 (0.76 to 0.94)
2nd tertile	0.48 (0.45 to 0.52)	0.46 (0.42 to 0.50)	0.50 (0.46 to 0.54)	1.14 (1.02 to 1.28)
Streetlights				
3rd tertile (highest)	0.70 (0.65 to 0.77)	0.67 (0.61 to 0.73)	0.73 (0.67 to 0.80)	1.52 (1.25 to 1.85)
2nd tertile	0.55 (0.51 to 0.58)	0.53 (0.49 to 0.57)	0.56 (0.52 to 0.61)	1.11 (0.93 to 1.33)
Street greenness				
3rd tertile (highest)	0.74 (0.69 to 0.79)	0.72 (0.67 to 0.77)	0.71 (0.66 to 0.75)	0.63 (0.58 to 0.69)
2nd tertile	0.77 (0.73 to 0.81)	0.75 (0.71 to 0.80)	0.76 (0.72 to 0.81)	0.85 (0.78 to 0.92)
Single-lane road				
3rd tertile (highest)	0.50 (0.48 to 0.53)	0.49 (0.46 to 0.52)	0.50 (0.47 to 0.52)	0.62 (0.58 to 0.67)
2nd tertile	0.69 (0.65 to 0.73)	0.68 (0.63 to 0.72)	0.69 (0.65 to 0.73)	0.77 (0.71 to 0.83)

Data sources: Built environment characteristics were derived from Google Street View images from the USA in 2019. Vehicle-related collisions were obtained from the Fatality Analysis Reporting System for fatal vehicle collisions. Collision data were summed across 2019–2022 to produce per capita averages per 10 000 population.

Adjusted Poisson regression models were run for each outcome separately. Models controlled for median age, per cent male, per cent Black, per cent Hispanic, per cent owner-occupied housing and Child Opportunity Index at the census tract level. Built environment characteristics were categorised into tertiles, with the lowest tertile serving as the referent group. N=65 159 census tracts.

Table 3 Census tract level built environment predictors of motor vehicle-related collisions (fatal and nonfatal) in Washington DC

Built environment characteristics	Total collisions	Pedestrian involved	Cyclist involved
	Rate ratio (95% CI)	Rate ratio (95% CI)	Rate ratio (95% CI)
Sidewalks			
3rd tertile (highest)	0.76 (0.59 to 0.99)	0.91 (0.68 to 1.21)	0.71 (0.50 to 1.01)
2nd tertile	1.07 (0.80 to 1.41)	1.26 (0.92 to 1.72)	1.16 (0.77 to 1.73)
Street lights			
3rd tertile (highest)	0.86 (0.63 to 1.16)	0.78 (0.53 to 1.15)	0.68 (0.43 to 1.07)
2nd tertile	0.96 (0.69 to 1.33)	0.81 (0.54 to 1.22)	0.78 (0.47 to 1.30)
Street greenness			
3rd tertile (highest)	0.48 (0.35 to 0.64)	0.46 (0.31 to 0.66)	0.32 (0.21 to 0.48)
2nd tertile	0.59 (0.46 to 0.76)	0.58 (0.43 to 0.77)	0.51 (0.35 to 0.75)
Stop signs			
3rd tertile (highest)	0.73 (0.54 to 0.99)	0.74 (0.52 to 1.07)	0.51 (0.32 to 0.80)
2nd tertile	0.94 (0.65 to 1.35)	0.79 (0.51 to 1.24)	0.63 (0.38 to 1.04)
Construction			
3rd tertile (highest)	1.39 (1.09 to 1.76)	1.61 (1.25 to 2.07)	1.47 (1.02 to 2.12)
2nd tertile	1.32 (1.02 to 1.70)	1.21 (0.92 to 1.60)	0.99 (0.62 to 1.58)

Data sources: Built environment characteristics were derived from Google Street View images from Washington DC collected in 2018–2021. Motor vehicle-related collisions were obtained from the District Department of Transportation for collisions on the roadway blocks network of Washington, DC to produce per capita yearly averages per 10 000 population for the time period 2019–2022.

Adjusted Poisson regression models were run for each outcome separately. Models controlled for median age, per cent male, per cent Black, per cent Hispanic, per cent owner-occupied housing and Child Opportunity Index at the census tract level. Built environment characteristics were categorised into tertiles, with the lowest tertile serving as the referent group. N=151 census tracts.

associated with higher rates of pedestrian/cyclist-involved collisions (table 2).

Table 3 presents regression analyses for Washington DC and includes data for both fatal and non-fatal collisions. In national and Washington DC analyses, sidewalk presence and street greenness were associated with a lower rate of total collisions in the third tertile, particularly those involving pedestrians and cyclists (table 3). However, only in Washington DC were streetlights connected with lower vehicle collisions involving pedestrians and cyclists; a higher frequency of streetlights was associated with a lower rate of pedestrians and cyclists involved in vehicle collisions in Washington DC (table 3). Additionally, in Washington DC, the highest tertile of stop signs had 27% fewer vehicle collisions, 26% fewer collisions involving pedestrians and 49% fewer collisions involving cyclists compared with the lowest tertile of stop signs. The highest tertile of road construction experienced 39% more vehicle collisions, 61% more collisions involving pedestrians and 47% more collisions involving cyclists compared with the lowest tertile of road construction.

DISCUSSION

Study findings in context

It is estimated that fatal and non-fatal collisions will cost the global economy US\$1.8 trillion between 2015 and 2030.³³ Recent literature suggests that built environment and road conditions significantly impact collision risk. Areas with mixed land use and smaller block sizes can lessen the need for vehicle usage while also promoting pedestrian activity.³⁴ Roads designed to be forgiving and minimise unexpected events can enhance overall road safety and mitigate vehicle collisions, especially for vulnerable users such as youth and the elderly.^{34 35}

Neighbourhood evaluations of built environmental features have traditionally relied on existing geographic information systems data or costly, labour-intensive onsite visits, or manual annotation of selected street segment images (i.e., virtual audits). Due to the resource-intensive nature of onsite visits and manual

image annotations, previous studies tend to involve only a few geographies or use locally available data.

Using computer vision models that can automatically analyse street imagery to create indicators of the built environment could dramatically reduce costs and time and provide a valuable data resource. These advancements would enhance understanding of large-scale patterns and inform interventions to decrease road traffic injuries and fatalities. The contribution of this study was to develop a national data repository that provides collision risk profiles for areas across the USA. Additionally, drawing on national collision data from the US FARS, we examined associations between built environment characteristics and fatal vehicle collisions. Supplemental analyses with Washington DC data confirmed robustness of findings on fatal and non-fatal vehicle collisions. Built environment characteristics under investigation—sidewalks, streetlights, single-lane roads and greener streets—were associated with dramatic reductions in vehicle collisions, particularly those involving pedestrians and cyclists (although only for the third tertile for sidewalks).

There were some unexpected results in this study. Contrary to the guiding hypotheses in this study, census tracts with a higher number of streetlights did not see fewer fatal collisions involving pedestrians or cyclists; instead, there was an increase. This phenomenon could be due to various factors; neighbourhoods with more streetlights might have more traffic from vehicles, pedestrians and cyclists than those with fewer streetlights, thus increasing the possibility for more collisions. A prior study³⁶ found a statistically significant dose–response association between higher average luminance of streetlights and improved urban road safety. While the current study assessed the presence of streetlights, it did not evaluate their adequacy. Furthermore, finding that streetlights were associated with fewer collisions in Washington DC could indicate that the relationship with road safety could vary with geographical factors.

Fatal collisions data from the time period 2019–2021 were used for this study. It should be noted that despite reduced motor

vehicle traffic (i.e., fewer miles travelled) during 2020 and 2021 due to the COVID-19 pandemic, fatal collisions and road traffic deaths in the USA continued to follow the prepandemic trend of increasing number of fatal collisions.^{37,38} Thus, our study is as relevant and important as ever given the need for additional data resources and empirical findings on how to prevent vehicle collisions.

Despite the study's strengths, this study is subject to certain limitations. While this study investigated important built environment characteristics that have been theoretically and empirically linked to vehicle collision risk, the examination did not involve all possible built environment characteristics and additional future research is warranted on traffic calming and cycling infrastructure. Traffic calming measures such as raised crosswalk, speed bumps, traffic circles and reduced speed limits can decrease both the occurrence and severity of collisions.³⁹ Dedicated spaces for cyclists (such as bike lanes) and pedestrians (such as safe crossings and refuge islands or medians) can also provide extra protection against collisions and collision-induced injuries.⁴⁰ Moreover, many of the built environment features we modelled were derived from dichotomous measures, such as street greenness. Future research into different characterisations that include continuous measures could further help investigate dose–response patterns.

Additionally, future studies should build on these findings by exploring the impact of changes over time. In our dataset, there were insufficient fatal collisions to see a change over this short period of time. Also, built environments tend to change slowly,⁴¹ and GSV imagery may have time lags of multiple years between image updates. Using a longer time scale would enable longitudinal analyses, further strengthening causal inference. Furthermore, human behaviours such as substance use, seat belt wearing and distracted driving can influence the likelihood of a collision. Future studies can potentially incorporate indicators of neighbourhood alcohol availability (such as liquor stores, restaurants, pubs) as well as vehicle design into collision risk models.

While this study endeavours to provide a comprehensive overview of the associations between built environment characteristics and traffic collisions, the results should be interpreted with caution, considering GSV image updates are less frequent in rural areas as they are in urban locales, with the implication that built environment characterisation in rural areas may not be as current compared with urban areas.⁴² Moreover, although FARS provides a robust dataset for fatal collisions, it is well-established that non-fatal collisions can be under-reported^{43,44} or are not documented in a manner that ensures their inclusion in public databases.⁴⁵

CONCLUSIONS

Motor vehicle collisions pose a significant public safety concern, causing physical, mental and economic harm to individuals, families, and communities. This study harnesses the underutilized potential of Google Street View image data to capture built environment characteristics predictive of fatal vehicle collisions nationally. The findings reveal that sidewalks, streetlights, street greenness and single-lane roads can help reduce motor vehicle collisions and protect pedestrians and cyclists. Motor vehicle collisions do not need to be an inevitable consequence of highly mobile societies. This study has implications for public health and urban planning working to create environments that foster health and improve the safety of roadways.

Contributors QCN is serving as the corresponding author and guarantor for the study. QCN, TT, and DAQ developed the original idea for the study and helped draft

the manuscript. Data curation and analyses was performed by MA, LZ, QCN, XY, HM, and DL. Manuscript writing and editing was performed by TTN, RP, WY, XY, HM and MH.

Funding Research reported in this publication was supported by the National Library of Medicine under Award Number R01LM012849 (QCN) and the National Institute on Minority Health and Health Disparities R01MD015716 (TTN), R01MD016037 (QCN).

Disclaimer 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.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the University of Maryland Institutional Review Board (IRB project # 2062932).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. The Fatality Analysis Reporting System (FARS) data can be accessed at: <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars> Fatal and non-fatal traffic collisions in Washington DC can be accessed at: <https://opendata.dc.gov/datasets/crashes-in-dc/about> Google Street View neighborhood-level data can be accessed in the geoportal or by request from the coauthors: <https://arcg.is/88nK40>.

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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Quynh C. Nguyen <http://orcid.org/0009-0006-1184-1560>

Weijun Yu <http://orcid.org/0000-0002-0771-5581>

Ming Hu <http://orcid.org/0000-0003-2583-1161>

D. Alex Quistberg <http://orcid.org/0000-0001-9730-2686>

REFERENCES

- World Health Organization (WHO). Global status report on road safety 2023. Geneva, Switzerland The World Health Organization (WHO); 2023.
- Sauber-Schatz EK, Ederer DJ, Dellinger AM, et al. Vital signs: motor vehicle injury prevention—United States and 19 comparison countries. *MMWR Morb Mortal Wkly Rep* 2016;65:672–7.
- Ameratunga S, Hijar M, Norton R. Road-traffic injuries: confronting disparities to address a global-health problem. *Lancet* 2006;367:1533–40.
- Stevenson M, Thompson J, de Sá TH, et al. Land use, transport, and population health: estimating the health benefits of compact cities. *Lancet* 2016;388:2925–35.
- Smit W, Hancock T, Kumaresen J, et al. Toward a research and action agenda on urban planning/design and health equity in cities in low and middle-income countries. *J Urban Health* 2011;88:875–85.
- Quistberg DA, Koepsell TD, Boyle LN, et al. Pedestrian signalization and the risk of pedestrian-motor vehicle collisions in Lima, Peru. *Accid Anal Prev* 2014;70:273–81.
- Quistberg DA, Koepsell TD, Johnston BD, et al. Bus stops and pedestrian-motor vehicle collisions in Lima, Peru: a matched case-control study. *Inj Prev* 2015;21:e15–22.
- Quistberg DA, Koepsell TD, Miranda JJ, et al. The walking environment in Lima, Peru and pedestrian-motor vehicle collisions: an exploratory analysis. *Traffic Inj Prev* 2015;16:314–21.
- Sisiopiku VP, Akin D. Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data. *Transp Res F: Traffic Psychol Behav* 2003;6:249–74.
- 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.
- Biljecki F, Ito K. Street view imagery in urban analytics and GIS: a review. *Landsc Urban Plan* 2021;215:104217.
- Quistberg A, Gonzalez CI, Arbeláez P, et al. *430 training neural networks to identify built environment features for pedestrian safety*. BMJ Publishing Group Ltd, 2022.
- González C, Ayobi N, Escallón F, et al. STRIDE: street view-based environmental feature detection and pedestrian collision prediction. 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW): 2023: IEEE; 2023:3222–34.
- Fatality Analysis Reporting System (FARS). Available: <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>
- Crashes in DC. Available: <https://opendata.dc.gov/datasets/crashes-in-dc/about>

- 16 Acevedo-Garcia D, Noelke C, McArdle N, *et al.* Racial and ethnic inequities in children's neighborhoods: evidence from the new child opportunity index 2.0: study uses the child opportunity index 2.0 to examine geographic and racial/ethnic inequities children are exposed to in the one hundred largest metropolitan areas of the United States. *Health Aff* 2020;39:1693–701.
- 17 Yu C-Y. Built environmental designs in promoting pedestrian safety. *Sustainability* 2015;7:9444–60.
- 18 Steinbach R, Perkins C, Tompson L, *et al.* The effect of reduced street lighting on road casualties and crime in England and Wales: controlled interrupted time series analysis. *J Epidemiol Community Health* 2015;69:1118–24.
- 19 Zhu M, Sze NN, Newnam S. Effect of urban street trees on pedestrian safety: a micro-level pedestrian casualty model using multivariate bayesian spatial approach. *Accid Anal Prev* 2022;176:106818.
- 20 Harvey C, Aultman-Hall L. Urban streetscape design and crash severity. *Transp Res Rec* 2015;2500:1–8.
- 21 Calvi A. Does roadside vegetation affect driving performance?: driving simulator study on the effects of trees on drivers' speed and lateral position. *Transp Res Rec* 2015;2518:1–8.
- 22 Kadir MAA, Othman N. Towards a better tomorrow: street trees and their values in urban areas. *Procedia - Social and Behavioral Sciences* 2012;35:267–74.
- 23 Wang J, Cicchino JB. Safety effects of roundabout conversions in camel, Indiana, the roundabout city. *J Safety Res* 2022;82:159–65.
- 24 Sorock GS, Ranney TA, Lehto MR. Motor vehicle crashes in roadway construction workzones: an analysis using narrative text from insurance claims. *Accid Anal Prev* 1996;28:131–8.
- 25 Doege TC, Levy PS. Injuries, crashes, and construction on a superhighway. *Am J Public Health* 1977;67:147–50.
- 26 Pigman JG, Agent KR. Highway accidents in construction and maintenance work zones. Transportation Research Record; 1990.
- 27 Shehab T, Phu L. *Accident patterns in road construction work zones*. The American Institute of Constructors, 2015:46–57.
- 28 Li Y, Bai Y. Comparison of characteristics between fatal and injury accidents in the highway construction zones. *Safety Science* 2008;46:646–60.
- 29 Nguyen QC, Sajjadi M, McCullough M, *et al.* Neighbourhood looking glass: 360° automated characterisation of the built environment for neighbourhood effects research. *J Epidemiol Community Health* 2018;72:260–6.
- 30 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014.
- 31 He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition: 2016; 2016:770–8.
- 32 Abadi M, Barham P, Chen J, *et al.* Tensorflow: a system for large-scale machine learning. 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16): 2016; 2016:265–83.
- 33 Chen S, Kuhn M, Prettner K, *et al.* The global macroeconomic burden of road injuries: estimates and projections for 166 countries. *Lancet Planet Health* 2019;3:e390–8.
- 34 Cervero R. *Land-use mixing and suburban mobility*. University of California Transportation Center, 1989.
- 35 Nitsche P, Stütz R, Helfert M. A simulation-based concept for assessing the effectiveness of forgiving roadside treatments. *Adv Transp Stud* 2011.
- 36 Jackett M, Frith W. Quantifying the impact of road lighting on road safety—A new Zealand study. *IATSS Research* 2013;36:139–45.
- 37 Fatality facts 2021: yearly Snapshot. n.d. Available: <https://www.iihs.org/topics/fatality-statistics/detail/yearly-snapshot>
- 38 Continuation of research on traffic safety during the COVID-19 public health emergency: January – June 2021. n.d. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/2021-10/Traffic-Safety-During-COVID-19_Jan-June2021-102621-v3-tag.pdf
- 39 Gonzalo-Orden H, Rojo M, Pérez-Acebo H, *et al.* Traffic calming measures and their effect on the variation of speed. *Transp Res Proc* 2016;18:349–56.
- 40 Liu J, Hainen A, Li X, *et al.* Pedestrian injury severity in motor vehicle crashes: an integrated spatio-temporal modeling approach. *Accid Anal Prev* 2019;132:105272.
- 41 Hirsch JA, Grengs J, Schulz A, *et al.* How much are built environments changing, and where?: patterns of change by neighborhood sociodemographic characteristics across seven US metropolitan areas. *Soc Sci Med* 2016;169:97–105.
- 42 Bader MDM, Mooney SJ, Bennett B, *et al.* The promise, practicalities, and perils of virtually auditing neighborhoods using google street view. *Ann Am Acad Political Soc Sci* 2017;669:18–40.
- 43 Watson A, Watson B, Vallmuur K. How accurate is the identification of serious traffic injuries by police? The concordance between police and hospital reported traffic injuries. Proceedings of the 2013 Australasian road safety research, policing and education conference: 2013: Australasian College of Road Safety (ACRS); 2013:1–10.
- 44 Yamamoto T, Hashiji J, Shankar VN. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accid Anal Prev* 2008;40:1320–9.
- 45 Calma E, Jackson C. Observed disparities between 911 calls and crash reports. In: *Criminal justice, transportation*. D.C. Policy Center, 2021.