



Neighborhood Greenspace and Socioeconomic Risk are Associated with Diabetes Risk at the Sub-neighborhood Scale: Results from the Prospective Urban and Rural Epidemiology (PURE) Study

Blake Byron Walker¹ · Sebastian Tobias Brinkmann · Tim Große · Dominik Kremer · Nadine Schuurman · Perry Hystad · Sumathy Rangarajan · Koon Teo · Salim Yusuf · Scott A. Lear

Accepted: 18 March 2022 / Published online: 12 May 2022
© The Author(s) 2022

Abstract Greenspace and socioeconomic status are known correlates of diabetes prevalence, but their combined effects at the sub-neighborhood scale are not yet known. This study derives, maps, and validates a combined socioeconomic/greenspace index of individual-level diabetes risk at the sub-neighborhood scale, without the need for clinical measurements. In two Canadian cities (Vancouver and Hamilton), we computed 4 greenspace variables from satellite imagery and extracted 11 socioeconomic variables from the Canadian census. We mapped 5125 participants from the Prospective Urban

and Rural Epidemiology Study by their residential address and used age- and sex-dependent walking speeds to estimate individual exposure zones to local greenspace and socioeconomic characteristics, which were then entered into a principal component analysis to derive a novel diabetes risk index (DRI-GLUCoSE). We mapped index scores in both study areas and validated the index using fully adjusted logistic regression models to predict individual diabetes status. Model performance was then compared to other non-clinical diabetes risk indices from the literature. Diabetes prevalence among participants was 9.9%. The DRI-GLUCoSE index was a significant predictor of diabetes status, exhibiting a small non-significant attenuation with the inclusion of dietary and physical activity variables. The final models achieved a predictive accuracy of 75%, the highest among environmental risk models to date. Our combined index of local greenspace and socioeconomic factors demonstrates that the environmental component of diabetes risk is not sufficiently explained by diet and physical activity, and that increasing urban greenspace may be a suitable means of reducing the burden of diabetes at the community scale.

B. B. Walker (✉) · S. T. Brinkmann · T. Große · D. Kremer
Community Health Environments and Social Terrains Lab,
Institute for Geography, Friedrich-Alexander-Universität
Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen,
Germany
e-mail: blake.walker@fau.de

N. Schuurman
Department of Geography, Simon Fraser University,
Burnaby, Canada

P. Hystad
Spatial Health Lab, College of Public Health and Human
Sciences, Oregon State University, Corvallis, USA

S. Rangarajan · K. Teo · S. Yusuf
Population Health Research Institute, McMaster
University, Hamilton, Canada

S. A. Lear
Faculty of Health Sciences, Simon Fraser University,
Burnaby, Canada

Keywords Diabetes · Greenspace · Socioeconomic status · GIS · Indices

Background

Type 2 diabetes is a multifactorial disease with foundations in genetic and behavioural risk factors.

However, with the continuing increases in the prevalence of diabetes among countries of all economic strata, these factors do not comprehensively explain risk for diabetes and suggests societal and environmental factors are also involved, including the socioeconomic composition and geographical context of a neighborhood. This is reflected in the WHO's 2016 statement underscoring the importance of socioeconomic conditions and healthy neighborhood environments as a focus area for prevention policy [1].

Previous studies have identified a higher prevalence of diabetes in subpopulations with low socioeconomic status (SES), most commonly measured using composite indices that capture diverse aspects of socioeconomic (dis)advantage [2, 3]. When measured at the individual- or household-level, socioeconomic risk factors for diabetes include income [4], educational attainment [5, 6], and occupational status [5]. Individual SES is not distributed randomly, and low and high SES individuals cluster together, which can be measured using area-based SES measures, for example, using census tracts. At the neighborhood scale, these measures represent the socioeconomic composition of an area and have been used to identify policy interventions targeting the socioeconomic component of disease risk. In public health planning and prevention programmes, area-based SES indices (e.g., the Vancouver Area Neighborhood Deprivation Index) have demonstrated immense value as correlates for a wide range of diseases, syndromes, and health conditions [7]. Even when controlling for individual- or household-level SES, area-based metrics often remain significant predictors, suggesting a social-environmental/neighborhood-scale compositional component to the effects of SES on health [5, 8, 9].

In addition to the socioeconomic composition of a neighborhood, the geographical/structural context is known to exhibit effects on health. Many recent studies have focussed on neighborhood-scale contextual characteristics relevant to diabetes risk, predominantly greenspace and walkability; both are known to be associated with a lower diabetes incidence/prevalence, as they provide enhanced opportunities for physical activity [10–15]. In addition, greenspace may capture other pathways important to diabetes risk, such as reducing air pollution exposures [16], decreasing stress [16], or increasing neighborhood social capital [17]. As such, policy and

community-based initiatives to improve or enhance urban greenspace are often presented as measures for improving population health [1].

In many locations, the level of neighborhood greenspace is associated or correlated with SES and studies have attempted to disentangle the two and examine their effects in isolation using statistical adjustment [18]. However, we hypothesise that the health-related benefits of urban greenspace may, to a limited degree, mitigate the risk associated with socioeconomic deprivation. As such, area-level predictive models and risk indices should take not only the socioeconomic composition of a neighborhood into account, but also include contextual factors such as greenspace. The effects of neighborhood SES and greenspace are unlikely to be independent but rarely has the combined influence of these exposures been examined. This study therefore aims to examine the combined effects of local SES and walkable greenspace on diabetes risk to develop a novel, high-resolution geographical exposure model. We derived and validated a high-resolution, area-based Diabetes Risk Index of Green Land Use and Community Socioeconomic Environments (DRI-GLUCoSE) in two Canadian cities and their rural peripheries and mapped the results as follows.

Methods

Participant Data

The current cross-sectional study focusses on 5125 participants recruited between 2006 and 2009 in Hamilton and Vancouver, Canada, for whom we had robust address and community-level characteristics. These sites are part of the PURE study, which is an ongoing global cohort study investigating risk factor for chronic disease in 27 countries [19]. All participants provided informed consent and the study was approved by the local institution research ethics boards.

Individual variables collected using surveys included residential address, age, sex, household income range, tobacco and alcohol consumption, history of diagnosed diabetes, and history of diabetes medication use. Additionally, each participant's diet was determined using the alternative healthy eating index (AHEI), a nine-component index on

dietary choices and nutrient intake designed to assess dietary-based chronic disease risk [20], and daily physical activity metabolic equivalent of task (MET) scores were calculated from robust questionnaire instruments as described by Lear et al. [21]. Trained project staff conducted in-person measurements to derive body mass index (BMI, measured as a continuous variable) and waist-to-hip ratio (WHR).

Participants were classified as having diabetes (either type 1 or type 2) if they fulfilled any of the following criteria: self-reported a diagnosis of diabetes; current or past use of diabetes-specific medication; a fasting plasma glucose level greater than or equal to 7.0 mmol/l.

Environmental Data

Data for Canadian census dissemination areas (DA, the smallest available census area with an average population of 400–800 residents) for the census year 2006 were acquired from Statistics Canada. The SES variables included in this analysis are listed in Table 1. Each DA was categorised as urban/suburban/rural using a previously validated, Canada-specific classification method based on active transportation and population density [7, 22, 23].

The Normalized Difference Vegetation Index (NDVI) provides a greenspace metric often used in health research [24]. Derived from satellite imagery, higher NDVI values indicate a greater intensity of greenness. We acquired cloud-free LANDSAT 5 satellite images for Hamilton (09.05.2006) and Vancouver (23.07.2006) through the United States Geological Survey's EarthExplorer platform. Both images were preprocessed and NDVI scores were calculated and mapped as 30-m pixels (native resolution) in the study areas.

Spatial Modelling

Data preparation and analysis were conducted using R (v.4), and mapping was completed using QGIS (v.3.10) on a Linux system (AMD Ryzen 9 3900×CPU, 64 GB DDR4). All code including detailed documentation are available at <https://github.com/CHEST-Lab/DRIGLUCoSE>.

In order to estimate each participant's potential exposures to greenspace and local SES, we mapped

age- and sex-specific walkable buffer zones around each individual's residential address using street and path networks derived from OpenStreetMap data (Fig. 1). Using age- and sex-specific walking speeds (average male–female difference = 0.13 km/h; [25]), we then mapped each participant's walkable areas in 2-min increments from their residential address, with a maximum walking time of 20 min. The 20-min maximum parameter was derived through sensitivity analysis, in which walking areas were calculated in 2-min intervals and iteratively entered into the successive statistical models described below. The 20-min parameter was selected as it (i) featured the highest predictive performance, (ii) is a heuristically realistic representation of movement patterns in the city, in that it adequately captures activity spaces of persons in the study (this was ascertained through informal discussion with residents of the study area, in which several of the authors reside), and (iii) the 20-min zone also approximates the radius of structurally homogenous spaces/neighborhoods in the study area and may therefore be the most suitable proxy for composition of the built environment. Each participant was thereby assigned ten concentric walking zones from 0 to 20 min walking time (Fig. 1a). The concentric walking zones were then overlaid on the SES and greenspace data described above (Fig. 1b). A logit weighting function was applied to each buffer zone's mean distance to derive distance-based variable weights, such that the estimated effect of an SES or greenspace variable decreases as distance from the home increases (Fig. 1c). A logit function was selected as it heuristically approximates a suitable distance-decay function [26] and various parameterisations were assessed through sensitivity analysis. The zone-distance-weighted mean of each of the 15 SES variables and 4 greenspace variables was then assigned to each participant for index derivation. Equations and documented code are provided on our GitHub page, linked above.

In order to increase the amount of built environment information contained in the model, we calculated 4 separate metrics from the NDVI data: median NDVI score; standard deviation; and 95th and 5th percentiles. These were selected as they respectively represent overall local greenspace levels, variability in the amount and intensity of greenspace within each walking zone, the intensity of the most green areas (P_{95}) that may serve as

Table 1 Study participants and neighborhood characteristics by diabetes status

	No diabetes (<i>N</i> =4616, 90.1%)	Diabetes (<i>N</i> =509, 9.9%)	Total (<i>N</i> =5125)	Bivariate odds ratio OR (95% CI, <i>p</i> -value)
City				
Hamilton	2307 (87.5%)	331 (12.5%)	2638	
Vancouver	2309 (92.8%)	178 (7.2%)	2487	0.54 (0.44–0.65, <i>p</i> <0.001)
BMI				
Mean (SD)	27.3 (5.4)	32.1 (6.4)	27.8 (5.7)	1.12 (1.11–1.14, <i>p</i> <0.001)
Median (Q1; Q3)	26.4 (23.8; 29.8)	30.9 (27.5; 35.7)	26.8 (24.0; 30.4)	
Waist to hip ratio				
Mean (SD)	85.2 (9.0)	93.6 (8.5)	86.0 (9.3)	1.11 (1.10–1.12, <i>p</i> <0.001)
Median (Q1; Q3)	85.2 (78.6; 91.6)	94.2 (87.8; 99.6)	86.0 (79.3; 92.6)	
Obesity (WHR)				
Low and moderate	2787 (96.0%)	115 (4.0%)	2902	
High	1829 (82.3%)	394 (17.7%)	2223	5.22 (4.21–6.48, <i>p</i> <0.001)
Age (years)				
Mean (SD)	52.7 (9.4)	56.9 (8.3)	53.1 (9.3)	1.29 (1.23–1.36, <i>p</i> <0.001)
Median (Q1; Q3)	53.0 (45.0, 60.0)	58.0 (51.0, 63.0)	53.0 (46.0, 61.0)	
Sex				
Male	2076 (87.6%)	294 (12.4%)	2370	
Female	2540 (92.2%)	215 (7.8%)	2755	0.60 (0.50–0.72, <i>p</i> <0.001)
Household income range				
> 90 k	1793 (93.3%)	129 (6.7%)	1922	0.75 (0.71–0.80, <i>p</i> <0.001)
45 k–65 k	759 (89.8%)	86 (10.2%)	845	
30 k–45 k	564 (85.2%)	98 (14.8%)	662	
65 k–90 k	1010 (92.7%)	80 (7.3%)	1090	
20 k–30 k	304 (81.3%)	70 (18.7%)	374	
< 20 k	186 (80.2%)	46 (19.8%)	232	
AHEI score				
Mean (SD)	37.7 (10.0)	35.0 (9.5)	37.4 (9.9)	0.97 (0.96–0.98, <i>p</i> <0.001)
Median (Q1; Q3)	37.6 (30.7, 44.8)	34.8 (28.1, 41.4)	37.3 (30.4, 44.5)	
Physical activity MET score				
< 525	1721 (87.3%)	250 (12.7%)	1971	
≥ 525	2065 (92.3%)	172 (7.7%)	2237	0.57 (0.47–0.70, <i>p</i> <0.001)
Ever smoked				
No	2087 (92.3%)	174 (7.7%)	2261	
Yes	1699 (87.3%)	248 (12.7%)	2937	1.75 (1.43–2.15, <i>p</i> <0.001)
Daily drinker				
≥ 1 drink/day	1157 (91.0%)	114 (9.0%)	1271	

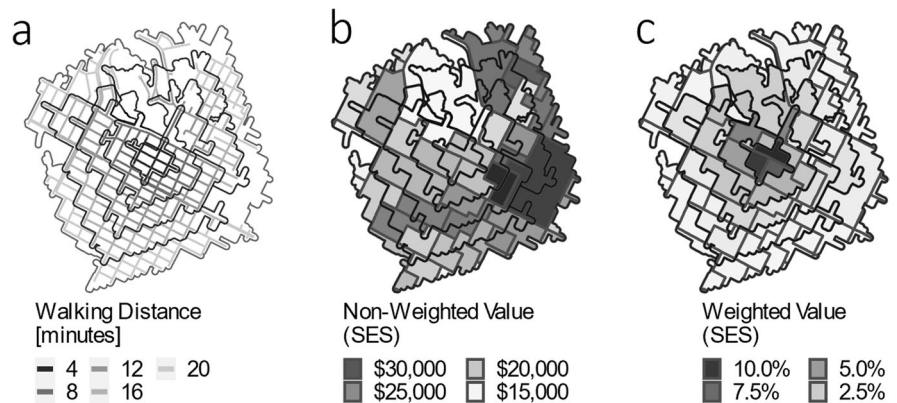
Table 1 (continued)

	No diabetes (<i>N</i> =4616, 90.1%)	Diabetes (<i>N</i> =509, 9.9%)	Total (<i>N</i> =5125)	Bivariate odds ratio OR (95% CI, <i>p</i> -value)
<1 drink/day	2629 (89.5%)	308 (10.5%)	2937	1.19 (0.95–1.49, <i>p</i> =0.133)
Participants' neighborhood socioeconomic status				
Neighborhood type				
Suburban/rural	3607 (89.9%)	405 (10.1%)	4012	0.92 (0.73–1.15, <i>p</i> =0.459)
Urban	1009 (90.7%)	104 (9.3%)	1113	
Individual mean income (CAD/1000)				
Mean (SD)	37.9 (13.9)	34.3 (12.1)	37.5 (13.8)	0.97 (0.96–0.98, <i>p</i> <0.001)
Median (Q1; Q3)	35.0 (29.0; 42.0)	32.0 (27.0; 38.0)	35.0 (29.0; 42.0)	
Household median income (CAD/1000)				
Mean (SD)	65.5 (21.5)	60.6 (20.5)	65.0 (21.4)	0.99 (0.98–0.99, <i>p</i> <0.001)
Median (Q1; Q3)	61.0 (51.0; 77.0)	57.0 (45.0; 71.0)	61.0 (51.0; 76.0)	
Prevalence of low income (%)				
Mean (SD)	9.4 (6.9)	11.3 (7.6)	9.6 (7.0)	1.04 (1.02–1.05, <i>p</i> <0.001)
Median (Q1; Q3)	8.1 (4.3; 13.7)	10.1 (5.2; 16.3)	8.2 (4.4; 14.0)	
Commute walking/bicycle (%)				
Mean (SD)	7.5 (6.9)	6.8 (5.8)	7.4 (6.8)	0.98 (0.97–1.00, <i>p</i> =0.027)
Median (Q1; Q3)	5.1 (3.0; 9.9)	5.3 (3.2; 8.6)	5.2 (3.0; 9.7)	
Labour force participation rate (%)				
Mean (SD)	66.6 (7.7)	65.3 (7.6)	66.5 (7.7)	0.98 (0.97–0.99, <i>p</i> <0.001)
Median (Q1; Q3)	66.8 (61.8; 71.7)	64.9 (60.7; 70.5)	66.6 (61.7; 71.6)	
Gov't transfer payments (%)				
Mean (SD)	9.9 (5.2)	12.4 (6.3)	10.1 (5.4)	1.08 (1.06–1.09, <i>p</i> <0.001)
Median (Q1; Q3)	8.8 (5.7; 13.0)	11.5 (7.2; 16.1)	9.0 (5.8; 13.3)	
Unemployment rate (%)				
Mean (SD)	5.5 (2.5)	6.3 (2.9)	5.6 (2.6)	1.11 (1.08–1.15, <i>p</i> <0.001)
Median (Q1; Q3)	5.4 (3.8; 7.0)	5.9 (4.4; 7.7)	5.4 (3.8; 7.0)	
Lone parent families (%)				
Mean (SD)	14.8 (6.5)	17.3 (7.5)	15.0 (6.6)	1.05 (1.04–1.07, <i>p</i> <0.001)
Median (Q1; Q3)	14.0 (10.9; 18.4)	16.1 (11.9; 21.8)	14.1 (11.0; 18.8)	
Education—no degree (%)				
Mean (SD)	19.6 (9.2)	23.7 (10.4)	20.0 (9.4)	1.04 (1.03–1.05, <i>p</i> <0.001)
Median (Q1; Q3)	17.7 (12.8; 24.9)	21.0 (15.8; 31.6)	17.9 (13.1; 25.6)	
Private dwellings—owned (%)				
Mean (SD)	73.9 (17.8)	72.9 (17.6)	73.8 (17.7)	1.00 (0.99–1.00, <i>p</i> =0.222)
Median (Q1; Q3)	77.1 (62.1; 88.5)	75.6 (61.7; 87.0)	77.0 (62.1; 88.4)	

Table 1 (continued)

	No diabetes (<i>N</i> =4616, 90.1%)	Diabetes (<i>N</i> =509, 9.9%)	Total (<i>N</i> =5125)	Bivariate odds ratio OR (95% CI, <i>p</i> -value)
Private dwellings—rented (%)				1.00 (1.00–1.01, <i>p</i> =0.166)
Mean (SD)	25.3 (17.2)	26.4 (17.0)	25.4 (17.2)	
Median (Q1; Q3)	22.1 (11.2; 36.9)	24.0 (12.9; 37.7)	22.4 (11.4; 36.9)	
Participants' neighborhood greenspace: Normalized Difference Vegetation Index (NDVI); OR per 0.1 NDVI-unit increase				
NDVI—median				0.68 (0.61–0.77, <i>p</i> <0.001)
Mean (SD)	0.341 (0.089)	0.314 (0.084)	0.338 (0.089)	
Median (Q1; Q3)	0.336 (0.291; 0.379)	0.313 (0.266; 0.360)	0.333 (0.288; 0.377)	
NDVI—standard deviation				0.79 (0.55–1.13, <i>p</i> =0.201)
Mean (SD)	0.090 (0.026)	0.088 (0.024)	0.090 (0.026)	
Median (Q1; Q3)	0.086 (0.071; 0.105)	0.084 (0.072; 0.100)	0.086 (0.071; 0.105)	
NDVI—Min				0.75 (0.67–0.83, <i>p</i> <0.001)
Mean (SD)	0.206 (0.097)	0.182 (0.086)	0.204 (0.096)	
Median (Q1; Q3)	0.202 (0.135; 0.262)	0.173 (0.118; 0.243)	0.198 (0.134; 0.260)	
NDVI—Max				0.69 (0.62–0.77, <i>p</i> <0.001)
Mean (SD)	0.495 (0.093)	0.465 (0.093)	0.492 (0.093)	
Median (Q1; Q3)	0.484 (0.435; 0.546)	0.459 (0.401; 0.514)	0.482 (0.432; 0.543)	

Fig. 1 Spatial weight-
ing procedure to account
for diminishing effects of
distance in which (a) a
road network was used to
compute isochrones, (b)
an unweighted variable is
mapped over the network,
and (c) a distance-decay
weighting scheme is applied
to the variable



local attractants or forested areas within a walking zone, and non-green areas (P_5) that are characteristic of a dense urban or industrial built environment (greyspace) and bare land, hypothesised to exhibit negative effects.

Index Derivation

Using participant data and residential address locations, we first derived the index as follows: in a subsequent step, we calculated and mapped the index across both study areas.

In order to manage class imbalance (i.e., ratio of diabetes to non-diabetes participants) for the index derivation [27], we used a combination of

undersampling and the SMOTE algorithm [28] to iteratively derive model calibration sets, as described in our code/documentation. To derive the DRI-GLUCoSE index, we used principal component analysis (PCA) [29–32]. The 11 SES and 4 greenspace variables listed in Table 1 were normalised and tested for non-sphericity and suitability using Bartlett's test and for sampling adequacy using Kaiser–Meyer–Olkin before being entered into a PCA model. The resulting loadings from the first component were then applied as variable weights for the 15 input variables, the sum of which was then rescaled to a range of $-1, 1$ and taken as the participant DRI-GLUCoSE score. The 4 SES variables with negligible Eigenvalues or absent from the first component were removed, as these had no detectable influence on model accuracy, resulting in a final index with 11 SES variables and 4 greenspace variables. The index is scaled such that high index values correspond with socially deprived areas and low greenspace.

For model comparison, we derived three variants of the index: (a) the combined DRI-GLUCoSE index described above; (b) a variant using only the SES variables; (c) a variant only using the greenspace variables. These are described in more detail on our GitHub page (linked above).

In order to map index scores across both study areas, we first removed non-residential land use from the maps, then overlaid a 50-m pixel grid. The 50-m parameter was selected to achieve a balance between computational load and spatial precision following sensitivity analysis of variable pixel sizes. From the centre of each 50-m pixel, we assume a standard participant with a 20-min walking zone and a fixed walking speed of 4.7 km/h [25]. The DRI-GLUCoSE index scores for each pixel were then calculated using the weighting procedure described above. This resulted in a DRI-GLUCoSE index map for both study areas.

Statistical Modelling and Validation

Using individual-level diabetes as a binary response variable, three sets of logistic regression models were run: (a) bivariate models with the DRI-GLUCoSE index and all control variables (age, sex, BMI, household income, urban/rural household, AHEI, recreational physical activity, smoking status, and alcohol consumption); (b) semi-adjusted multivariable

models using the DRI-GLUCoSE index, age, sex, BMI, household income, and urbanicity; and (c) fully adjusted multivariable models using all variables from model B, plus tobacco and alcohol consumption, diet, and physical activity. For comparison, we also computed a series of bivariate and (semi-/fully-) adjusted models using various combinations of the 4 greenspace variables and the univariate predictors without PCA, including testing for interaction effects between greenspace and SES predictors. Odds ratios (OR) with 95% confidence intervals (CI) and p -values were reported for all models. Experimental models and models using WHR as the adiposity measure are reported in the online documentation.

To assess model performance and test for overfitting, we randomly split the data into training (80%) and testing (20%) subsets, preserving the within-group diabetes prevalence. In-sample statistics (j -index, sensitivity, specificity, AROC, and accuracy) of the training subset were calculated through tenfold cross-validation and compared to the out-of-sample statistics of the testing subset.

Results

Sample Characteristics

Of the total 5125 participants included in this analysis, 10% were classified as having diabetes (Table 1). The mean BMI and WHR were 27.8 kg/m² and 0.86, respectively, and participant median age was 53 years (IQR 46, 61) at the time of data collection. Approximately 28% of participants reported an average of more than one unit of alcohol consumed per day, and 58.8% of household annual incomes were above CAD \$65,000. Participants were equally divided between the two cities, with 78% of participants residing in suburban or rural neighborhoods.

DRI-GLUCoSE Index

Bartlett's test results confirmed strong non-sphericity ($p < 0.001$) of the predictor variables, and Kaiser–Meyer–Olkin returned an overall measure of sampling adequacy of 0.74, indicating a reasonable degree of suitability. The PCA results indicate that 49.3% of variance was explained in the first component, whose factor loadings for the derived index are

Table 2 PCA results. Factor loadings in PC1 correspond to variable weights for the DRI-GLUCoSE index

Variable	Factor loadings (PC1)	Factor loadings (PC2)
Government transfer payments	-0.35	0.19
Lone parent families	-0.32	0.07
Household median income	0.31	0.18
Prevalence of low income	-0.30	-0.09
Unemployment rate	-0.29	0.15
Education—no degree	-0.29	0.35
NDVI—median	0.26	-0.12
NDVI—5th percentile	0.26	0.07
NDVI—95th percentile	0.25	-0.29
Individual mean income	0.24	-0.04
Private dwellings—owned	0.24	0.45
Private dwellings—rented	-0.24	-0.44
Labour force participation rate	0.12	-0.08
Commute active	-0.10	-0.32
NDVI—standard deviation	-0.02	-0.41
Total variance explained	49.3%	14.5%

Bartlett's test for sphericity $p < 0.001$

Kaiser-Meyer-Olkin test for measure of sampling adequacy = 0.74

shown in Table 2. The remaining components respectively explained 14.5%, 9.6%, 6.8%, and 5.2% of variance and were discarded. The factor loadings for the first component were then used as variable weights and entered a weighted linear combination to derive index scores.

The mapped index values (Fig. 2) highlight areas where the local SES and greenspace conditions predict a higher (dark) or lower (light) diabetes risk.

Regression Models

The results of the logistic regression models based on the testing subset are shown in Fig. 3. Participants' local DRI-GLUCoSE score exhibited a consistent positive association across all models. A small, non-significant attenuation of this effect was observed in the fully adjusted models, where individual-level lifestyle factors were included (diet, physical activity, smoking, and alcohol consumption). Increased odds of diabetes were observed for obese participants, current/former smokers, and persons who consume an average of less than one unit of alcohol per day. A lower risk is observed for urban residence, higher

household incomes, and physical activity scores, but healthy eating score was non-significant (odds ratios shown in online appendix: <https://github.com/CHEST-Lab/DRIGLUCoSE>).

The multivariable models' accuracies were 75% for the semi-adjusted model (Youden Index = 0.41, sensitivity = 0.78, specificity = 0.62, AROC = 0.76) and 75% for the fully adjusted model (Youden Index = 0.41, sensitivity = 0.75, specificity = 0.71, AROC = 0.78). Model diagnostics indicated no overfitting. Experimental models based on alternative configurations of greenspace variables, univariate (non-PCA) predictor variables, and interactions between greenspace and SES predictors consistently featured lower accuracy and/or featured poor model fit.

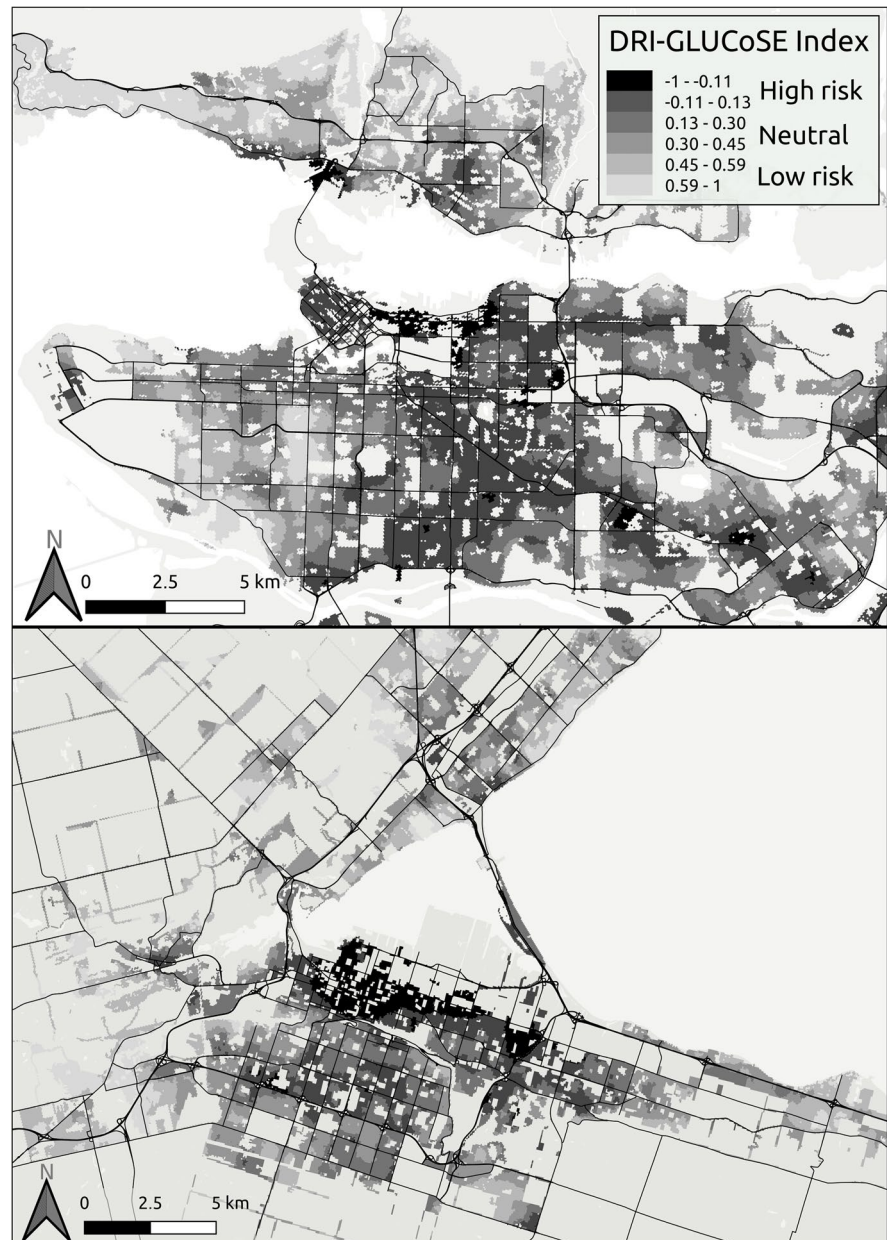
The final models were compared to identically parameterised logistic models using two additional variants of the DRI-GLUCoSE index: SES only and greenspace only. The combined SES and greenspace model achieved the highest prediction accuracy of 75%, compared to 72% for SES only (see Appendix).

Conclusions

Numerous studies since the 1970s have demonstrated the power of SES for identifying and predicting health risks and outcomes. In this study, we augmented the utility of SES indices by including a novel model for greenspace, which is known to be a strong predictor for physical activity and positive health status/outcomes. Our regression model results indicate the DRI-GLUCoSE index is significantly associated with reduced odds of individual-level diabetes risk in two test cities in Canada, and its use of both greenspace and SES provides the strongest environmental risk model to date (75% prediction accuracy) for predicting individual diabetes risk using only non-invasive data. These results indicate that one's diabetes risk is associated with their local neighborhood greenspace and SES and contribute to a growing recognition of geographical factors as important predictors of disease risk.

The index results are concurrent with the literature, in that higher neighborhood SES and local greenspace exhibit a protective effect [2, 5, 11, 12, 14, 15]. In terms of urban/regional planning policy and diabetes prevention, low SES, unhealthy food

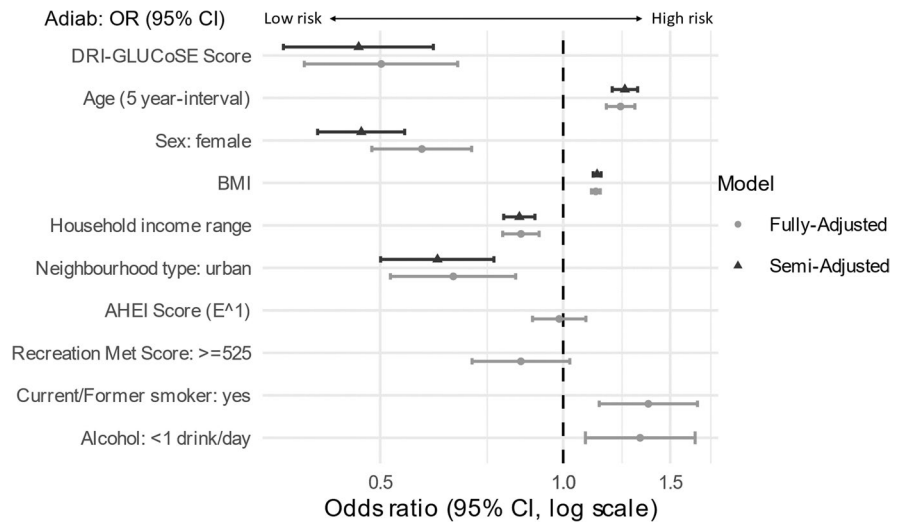
Fig. 2 DRI-GLUCoSE scores for Vancouver (top) and Hamilton (bottom), ranging from low risk (light) to high risk (dark)



environments, and lack of infrastructure for physical activity have been underscored as key priorities [1]. Markevych et al. [33] and Astell-Burt et al. [15] categorized the mechanisms linking greenspace to health benefits into three domains: mitigation (e.g., better air quality); restoration (e.g., stress recovery); and instoration (e.g., promotion of physical activity) and emphasise the interconnectedness of mediation pathways, which has not yet been thoroughly addressed in the literature [34]. Our modelling results

largely correspond with previous studies, in that low household income [4] and tobacco use [2, 35] were significant risk factors. Urban place of residence was associated with decreased diabetes odds in our models, as also reported by Basiak et al. [36], but contrary to Dagenais et al. [6], who observed elevated diabetes risk in urban areas. Recreational physical activity [6, 37] and healthy eating [38] exhibited a minor attenuation in the effect of the DRI-GLUCoSE index on diabetes risk. Also observed was an association between

Fig. 3 Forest plot showing significant effects for both semi-adjusted and fully adjusted multivariable logistic models. DRI-GLUCoSE, Diabetes Risk Index-Green Land Use and Community Socioeconomic Environments; BMI, body mass index; AHEI, Alternative Healthy Eating Score



low/no alcohol consumption and increased diabetes odds, a result concordant with the literature [2, 39] and likely explained by patients being advised against alcohol consumption following a diabetes diagnosis [2, 36].

By accounting for environmental characteristics in the DRI-GLUCoSE index, we observed strong predictive performance based primarily on environmental characteristics. Other diabetes risk indices have been developed using a larger number of non-clinical participant-level variables and no environmental factors and have achieved similar predictive performance with an AROC of 0.78 [8] and 0.745 [40]. Within similar accuracy, the DRI-GLUCoSE index provides a free and easy-to-use risk estimation and mapping tool for estimating the combined effects of SES and greenspace on diabetes risk at the sub-neighborhood scale without the need for clinical measurements or individual-level data. The required socioeconomic data and satellite imagery are freely available for most regions of the globe and require minimal pre-processing, and the index calculation tool with documentation is freely available via GitHub, enabling public health analysts and researchers to calculate and map the index in the region of their choice. By mapping areas of interest using DRI-GLUCoSE, zones with a higher potential risk can be identified and targeted for urban renewal policy, diabetes prevention, patient counselling, and health services planning. However, it is important to note that the effect size exhibited by greenspace characteristics is significantly smaller than those of local and household

SES and individual-level risk factors such as age, sex, and modifiable risk. The inclusion of greenspace to the SES-derived index only resulted in a 3% increase in overall accuracy in the fully adjusted models. This is indicative of (i) the relatively weak effects exhibited by local greenspace, and (ii) the importance of health-related behaviours in mitigating diabetes risk, regardless of geographical setting.

This study uses participant data that did not differentiate between type 1 and type 2 diabetes; given that the environmental effect on type 1 diabetes is likely to be low, it may be that our index and models underestimate the associations observed. While this study benefited from using two cities and their surrounding regions, both study areas are located in a high-income global region, and the index and modelling results may therefore not be generalisable to other global regions. Further research will focus on index refinement for middle- and low-income countries. Our study is limited by an exclusive focus on participants' place of residence as their primary exposure region. Importantly, this study used a cross-sectional design, preventing any inference of causality. Finally, despite being demographically representative, the study cohort may exhibit some selection bias, as participation was voluntary. However, as the logistic models presented herein are similar to others from the literature, we believe these results to be adequately representative for the study population. Our results provide strong evidence for an environmental component to diabetes risk that is not accounted for by the selected covariates, but a dedicated study design to analyse the

potential roles of, e.g., the food environment and air/noise/light pollution, may elucidate these effects in more detail.

The DRI-GLUCoSE index differs in that neighborhoods are defined not by administrative boundaries, rather, by age- and sex-specific walking zones. While this novel method for estimating exposure to greenspace and socioeconomic settings provides a more realistic representation of a person's activity space, it is important to note that the downscaling of census data to a higher spatial resolution suffers from the assumption of demographic and socioeconomic homogeneity within the census units (i.e., the attributes of a census dissemination area are assigned to each individual resident within it, regardless of whether they reside in the centre of that area or on the border). So while this technique sought to mitigate facets of the ecological fallacy induced through spatial containerisation and boundary effects, it is not able to overcome these limitations entirely.

In this study, we derived and validated the DRI-GLUCoSE index as a high-resolution tool for quantifying and mapping the combined socioeconomic and greenspace component of local diabetes risk using epidemiological cohort data in two Canadian cities. The index remained a significant predictor of diabetes risk after controlling for individual-level modifiable risk factors. In the absence of individual-level data, neighborhood-level indices like DRI-GLUCoSE can provide a useful means for identifying areas of higher environmental risk. This is invaluable for planning prevention policy, designing healthy neighborhoods, and targeting patient counselling guidelines.

Acknowledgements The authors wish to thank all participants and staff in the PURE Study, as well as the volunteers working with the OpenStreetMap Foundation and those involved in the R Project. We also extend our thanks to the United States Geological Survey for providing free satellite imagery.

Prof. Dr. BB Walker is supported by the German Federal Ministry for Education and Research (BMBF). Dr. S Yusuf is supported by the Mary W. Burke endowed chair of the Heart and Stroke Foundation of Ontario. The PURE study is an investigator-initiated study that is funded by the Population Health Research Institute, Hamilton Health Sciences Research Institute (HHSRI), the Canadian Institutes of Health Research, Heart and Stroke Foundation of Ontario, Support from Canadian Institutes of Health Research's Strategy for Patient Oriented Research, through the Ontario SPOR Support Unit, as well as the Ontario Ministry of Health and Long-Term Care and through unrestricted grants from several pharmaceutical

companies [with major contributions from AstraZeneca (Canada), Sanofi-Aventis (France and Canada), Boehringer Ingelheim (Germany and Canada), Servier, and GlaxoSmithKline], and additional contributions from Novartis and King Pharma and from various national or local organisations in participating countries.

Author Contributions B.B.W. and S.A.L. conceptualised and supervised this study. S.Y., K.T., S.A.L., and S.R. lead the study through which participant data were collected. S.T.B., T.G., B.B.W., S.A.L., and P.H. conducted the literature review. S.T.B. and T.G. prepared the datasets. S.T.B., B.B.W., D.K., and T.G. conducted the statistical analyses. S.T.B. produced the tables and figures. B.B.W., S.A.L., P.H., K.T., S.Y., S.T.B., T.G., D.K., N.S., and S.R. interpreted the results and wrote the manuscript. Project financing was acquired and managed by S.A.L., N.S., and B.B.W. B.B.W. is the guarantor of this article.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Health Organization. *Global report on diabetes*. Geneva: WHO Press World Health Organization; 2016.
2. Jeyam A, McGurnaghan SJ, Blackburn LAK, McKnight JM, Green F, Collier A, et al. Diabetic neuropathy is a substantial burden in people with type 1 diabetes and is strongly associated with socioeconomic disadvantage: a population-representative study from Scotland. *Dia Care*. 2020;43:734–42.
3. Anjana RM, Mohan V, Rangarajan S, Gerstein HC, Venkatesan U, Sheridan P, et al. Contrasting associations between diabetes and cardiovascular mortality rates in low-, middle-, and high-income countries: cohort study data from 143,567 individuals in 21 countries in the PURE Study. *Dia Care*. 2020;43:3094–101.
4. Bijlsma-Rutte A, Rutters F, Elders PJM, Bot SDM, Nijpels G. Socio-economic status and HbA1c in type 2 diabetes: a systematic review and meta-analysis. *Diabetes Metab Res Rev*. 2018;34:e3008.

5. Consolazio D, Koster A, Sarti S, Schram MT, Stehouwer CDA, Timmermans EJ, et al. Neighbourhood property value and type 2 diabetes mellitus in the Maastricht study: a multilevel study. *PLOS ONE*. 2020;15:e0234324.
6. Dagenais GR, Gerstein HC, Zhang X, McQueen M, Lear S, Lopez-Jaramillo P, et al. Variations in diabetes prevalence in low-, middle-, and high-income countries: results from the prospective urban and rural epidemiological study. *Diabetes Care*. 2016;39:780–7.
7. Walker BB, Schuurman N, Wen CK, Shakeel S, Schneider L, Finley C. Cancer resection rates, socioeconomic deprivation, and geographical access to surgery among urban, suburban, and rural populations across Canada. *PLoS one*. 2020;15:e0240444.
8. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust*. 2010;192:197–202.
9. Fairley L, Dundas R, Leyland AH. The influence of both individual and area based socioeconomic status on temporal trends in Caesarean sections in Scotland 1980–2000. *BMC Public Health*. 2011;11:330.
10. Kartschmit N, Sutcliffe R, Sheldon MP, Moebus S, Greiser KH, Hartwig S, et al. Walkability and its association with prevalent and incident diabetes among adults in different regions of Germany: results of pooled data from five German cohorts. *BMC Endocr Disord*. 2020;20:7.
11. Müller G, Harhoff R, Rahe C, Berger K. Inner-city green space and its association with body mass index and prevalent type 2 diabetes: a cross-sectional study in an urban German city. *BMJ open*. 2018;8:e019062.
12. Dalton AM, Jones AP, Sharp SJ, Cooper AJM, Griffin S, Wareham NJ. Residential neighbourhood greenspace is associated with reduced risk of incident diabetes in older people: a prospective cohort study. *BMC Public Health*. 2016;16:1171.
13. Creator MI, Glazier RH, Moineddin R, Fazli GS, Johns A, Gozdyra P, et al. Association of neighborhood walkability with change in overweight, obesity, and diabetes. *JAMA*. 2016;315:2211–20.
14. Bodicoat DH, O'Donovan G, Dalton AM, Gray LJ, Yates T, Edwardson C, et al. The association between neighbourhood greenspace and type 2 diabetes in a large cross-sectional study. *BMJ open*. 2014;4:e006076.
15. Astell-Burt T, Feng X, Kolt GS. Is neighborhood green space associated with a lower risk of type 2 diabetes? Evidence from 267,072 Australians. *Diabetes Care*. 2014;37:197–201.
16. Howell NA, Tu JV, Moineddin R, Chen H, Chu A, Hystad P, et al. The probability of diabetes and hypertension by levels of neighborhood walkability and traffic-related air pollution across 15 municipalities in Southern Ontario, Canada: a dataset derived from 2,496,458 community dwelling-adults. *Data Brief*. 2019;27:104439.
17. Flôr CR, Baldoni NR, Aquino JA, Baldoni AO, Fabro ALD, Figueiredo RC, et al. What is the association between social capital and diabetes mellitus? A systematic review. *Diabetes Metab Syndr*. 2018;12:601–5.
18. Mitchell R, Popham F. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet*. 2008;372:1655–60.
19. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J*. 2009;158:1-7.e1.
20. McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr*. 2002;76:1261–71.
21. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *Lancet*. 2017;390:2643–54.
22. Walker BB, Shashank A, Gasevic D, Schuurman N, Poirier P, Teo K, et al. The local food environment and obesity: evidence from three cities. *Obesity (Silver Spring, Md)*. 2019;28:40–5.
23. Gordon DLA, Janzen M. Suburban nation? Estimating the size of Canada's suburban population. *J Archit Plan Res*. 2013;30:197–220.
24. James P, Banay RF, Hart JE, Laden F. A review of the health benefits of greenness. *Curr Epidemiol Rep*. 2015;2:131–42.
25. Dewulf B, Neutens T, van Dyck D, de Bourdeaudhuij I, van de Weghe N. Correspondence between objective and perceived walking times to urban destinations: influence of physical activity, neighbourhood walkability, and socio-demographics. *Int J Health Geogr*. 2012;11:43.
26. Jia P, Wang F, Xierali IM. Differential effects of distance decay on hospital inpatient visits among subpopulations in Florida, USA. *Environ Monit Assess*. 2019;191:381.
27. Liu S, Ong M, Kin L, Mun K, Yao J, Motani M. Early prediction of sepsis via SMOTE upsampling and mutual information based downsampling. *CinC*. 2019;6:1–4.
28. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *JAIR*. 2002;16:321–57.
29. Lampert T, Hoebel J, Kuntz B, Müters S. Messung des sozioökonomischen Status und des subjektiven sozialen Status in KiGGS Welle 2. *J Health Monitor*. 2018;3(1):114–33.
30. Meijer M, Engholm G, Grittner U, Gritter U, Bloomfield K. A socioeconomic deprivation index for small areas in Denmark. *Scand J Public Health*. 2013;41:560–9.
31. Bell N, Hayes MV. The Vancouver Area Neighbourhood Deprivation Index (VANDIX): a census-based tool for assessing small-area variations in health status. *Can J Public Health*. 2012;103:S28–32.
32. Havard S, Deguen S, Bodin J, Louis K, Laurent O, Bard D. A small-area index of socioeconomic deprivation to capture health inequalities in France. *Soc Sci Med (1982)*. 2008;67:2007–16.
33. Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, et al. Exploring pathways linking greenspace to health: theoretical and methodological guidance. *Environ Res*. 2017;158:301–17.

34. Dzhambov AM, Browning MHEM, Markevych I, Hartig T, Lercher P. Analytical approaches to testing pathways linking greenspace to health: a scoping review of the empirical literature. *Environ Res.* 2020;186:109613.
35. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA.* 2007;298:2654–64.
36. Basiak A, Róžańska D, Poltyn-Zaradna K, Wołyniec M, Szuba A, Zatońska K. Comparison of intake of food groups between participants with normoglycemia, impaired fasting glucose, and type 2 diabetes in PURE Poland population. *Int J Diabetes Dev Ctries.* 2019;39:315–24.
37. Kriska AM, Rockette-Wagner B, Edelstein SL, Bray GA, Delahanty LM, Hoskin MA et al. The impact of physical activity on the prevention of type 2 diabetes: evidence and lessons learned from the Diabetes Prevention Program, a long-standing clinical trial incorporating subjective and objective activity measures. *Dia Care.* 2020;44:43–9.
38. Lee Y, Park K. Adherence to a vegetarian diet and diabetes risk: a systematic review and meta-analysis of observational studies. *Nutrients.* 2017;9:603.
39. Lai Y-J, Hu H-Y, Lee Y-L, Ko M-C, Ku P-W, Yen Y-F, et al. Frequency of alcohol consumption and risk of type 2 diabetes mellitus: a nationwide cohort study. *Clin Nutr.* 2019;38:1368–72.
40. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study. *Fam Pract.* 2008;25(3):191–6.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.