



Estimating multiple greenspace exposure types and their associations with neighbourhood premature mortality: a socioecological study

DOI:

[10.1016/j.scitotenv.2021.147919](https://doi.org/10.1016/j.scitotenv.2021.147919)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Labib, SM., Lindley, S., & Huck, J. (2021). Estimating multiple greenspace exposure types and their associations with neighbourhood premature mortality: a socioecological study. *Science of the Total Environment*, 789, Article 147919. <https://doi.org/10.1016/j.scitotenv.2021.147919>

Published in:

Science of the Total Environment

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact openresearch@manchester.ac.uk providing relevant details, so we can investigate your claim.



Estimating multiple greenspace exposure types and their associations with neighbourhood premature mortality: a socioecological study

S. M. Labib; Sarah Lindley; Jonny J Huck

Abstract

Background: Greenspace exposures are often measured using single exposure metrics, which can lead to conflicting results. Existing methodologies are limited in their ability to estimate greenspace exposure comprehensively. We demonstrate new methods for estimating single and combined greenspace exposure metrics, representing multiple exposure types that combine impacts at various scales. We also investigate the association between those greenspace exposure types and premature mortality.

Methods: We used geospatial data and spatial analytics to model and map greenspace availability, accessibility and eye-level visibility exposure metrics. These were harmonised and standardised to create a novel composite greenspace exposure index (CGEI). Using these metrics, we investigated associations between greenspace exposures and years of potential life lost (YPLL) for 1673 neighbourhoods applying spatial autoregressive models. We also investigated the variations in these associations in conjunction with levels of socioeconomic deprivation based on the index of multiple deprivations.

Results: Our new CGEI metric provides the opportunity to estimate spatially explicit total greenspace exposure. We found that a 1-unit increase in neighbourhood CGEI was associated with approximately a 10-year reduction in YPLL. Meaning a 10% or 0.1 increase in the CGEI is associated with an approximately one year lower premature mortality value. A single 1-unit increase in greenspace availability was associated with a YPLL reduction of 9.8 years, whereas greenness visibility related to a reduction of 6.14 years. We found no significant association between greenspace accessibility and YPLL. Our results further identified divergent trends in the relations between greenspace exposure types (e.g. availability vs. accessibility) and levels of socioeconomic deprivation (e.g. least vs. most).

Conclusion: Our methods and metrics provide a novel approach to the assessment of multiple greenspace exposure types, and can be linked to the broader exposome framework. Our results showed that a higher composite greenspace exposure is associated with lower premature mortality.

Key words: Green Space, Greenness, Pathways, Mortality, Exposome, Geographic Information Systems

Highlights:

- Existing greenness exposure metrics are inadequate in estimating multiple pathways
- We developed novel metrics of greenspace availability, accessibility, and visibility
- We generated composite greenspace exposure combining three individual metrics
- We found inverse associations between composite greenspace and premature mortality
- The positive impact of greenspace exposure is more pronounced in deprived neighborhoods

[Author Accepted Manuscript- in *Science of the total environment*]

1. Introduction

The connection between human health and the natural environment is widely recognised (Frumkin *et al.*, 2017; Díaz *et al.*, 2018; Bratman *et al.*, 2019). Several recent reviews have demonstrated that human interactions with nature, particularly exposure to green and blue spaces (commonly referred to jointly as ‘greenspaces’) provide multiple pathways enabling positive health benefits (Hartig *et al.*, 2014; Markevych *et al.*, 2017). The importance of these pathways is also recognised in the increasing emphasis on urban green infrastructure as a core part of spatial planning (Lindley *et al.*, 2019; Dennis *et al.*, 2020). Existing literature indicates that exposure to greenspace reduces health risk factors associated with all-cause mortality, adverse pregnancy outcomes, elevated blood pressure, type 2 diabetes, among others via improved physical health (Kondo *et al.*, 2018; Rojas-Rueda *et al.*, 2019; Jimenez *et al.*, 2021). Studies have also reported positive effects related to greenspace exposure on mental health. Positive effects on mental health include the reduction of anxiety levels and stress levels as well as alleviating depression, improving sleep, enhancing cognitive development, and improving overall life satisfaction (Dadvand *et al.*, 2015; Fong, Hart and James, 2018; Kondo *et al.*, 2018). Nonetheless, a few studies have also found null, mixed, or adverse effects related to greenspace exposure (Picavet *et al.*, 2016; Dzhambov *et al.*, 2020). Noting this, some scholars have argued that the possible reasons for these inconsistent findings might be attributed to differences in measuring exposures in terms of both spatial and temporal variations as well as in the context of specific health and medical conditions (e.g., allergies) that may be negatively impacted by some types of green space exposure (Markevych *et al.*, 2017; Davis *et al.*, 2021; Stas *et al.*, 2021).

The conceptualisation of greenspace exposure used to frame previous studies involving objective measurement can be broadly classified into three categories; greenspace availability, accessibility, and visibility (Dadvand and Nieuwenhuijsen, 2019). These three categories represent the most commonly used spatial measures of greenspace exposure (Ekkel and de Vries, 2017; Labib, Lindley and Huck, 2020b). Each of these three greenspace exposure categories are linked to different, though often overlapping, mechanistic pathways influencing health (Hartig *et al.*, 2014; James *et al.*, 2015; Markevych *et al.*, 2017; Bratman *et al.*, 2019). For example, availability of greenspace has been

associated with the reduction of environmental stressors (e.g., air and noise pollution, heat mitigation) (Frumkin *et al.*, 2017; Lindley *et al.*, 2019). Access to greenspaces may encourage physical activities and increase social cohesion, linked to the capacity building pathway (Hartig *et al.*, 2014; Markevych *et al.*, 2017; Nieuwenhuijsen *et al.*, 2017). Visibility of greenspace or ‘greenness’ visibility (Labib, Huck and Lindley, 2021) may be associated with stress recovery and attention restoration (Ulrich, 1984; Kaplan and Kaplan, 1989). Additionally, spatial measures of greenspace exposure may act as proxies for the “cumulative opportunity” of exposure to nature (Frumkin *et al.*, 2017), the actual exposure will depend on the usage and time spent in contact with greenspaces (Bratman *et al.*, 2019; Holland *et al.*, 2021). It should be noted that, unlike some other external environmental exposure types (Turner *et al.*, 2017), the precise impact and value of greenspace exposure influences are, as yet, unclear. This is due at least in part to the multiple mechanistic pathways involved in health benefits, the heterogeneous exposure assessment methods applied in existing studies, and the difficulty in accounting for actual exposure in terms of amount of time spent in nature (Turner *et al.*, 2017; Labib, Lindley and Huck, 2020b; Holland *et al.*, 2021).

Existing approaches to the objective measurement of greenspace exposure typically only consider *one* of these three types of exposure (availability, accessibility, and visibility) and often associate them individually with health indicators. Accordingly, the aggregate value of the impact of greenspace on human health is not yet well understood (Frumkin *et al.*, 2017; Bratman *et al.*, 2019; Labib, Lindley and Huck, 2020b). Understanding the influence of aggregate exposure is crucial because many of the non-communicable disease and health indicators considered in relation to greenspace exposure can be attributed to the aggregate effects of multiple exposure types (Frumkin *et al.*, 2017; Silva, Rogers and Buckley, 2018; Labib, Lindley and Huck, 2020b). Additionally, individuals are exposed to varying amounts and types of greenspace at different locations (e.g., home, office, streets), and this variability may result in differing health effects both on any given individual and from individual to individual over time (Helbich, 2018). Such variations in exposure over different spatial and temporal extents also relates to the uncertain geographic context problem (UGCoP) noted by (Kwan, 2012). For greenspace exposure, the UGCoP indicates that there is spatial uncertainty in the

actual areas over which the greenspace exposure can impact health outcomes along multiple pathways and contexts (Pearce, 2018; Labib, Lindley and Huck, 2020a). The pre-existing methodologies applied to greenspace exposure assessments are usually considered at a fixed spatial scale. As a result, there are uncertainties in estimating the aggregate effects of greenspace exposure at multiple locations related to varying temporal and spatial scales (Kwan, 2012; Helbich, 2018; Labib, Lindley and Huck, 2020b). To better understand exposure duration, exposure sequences, and exposure accumulation, measurement of aggregate greenspace exposures at multiple scales therefore requires careful attention.

In this regard, new methodological solutions are required that allow the measurement of composite greenspace exposure based on hierarchical spatial scale values. Turner et al. (Turner *et al.*, 2017) indicated that conducting assessments of greenspace exposures that represent the totality of all such measurements in aggregate may have only moderate feasibility for the study of larger populations. Turner et al. also noted that it would be difficult to infer this exposure at the individual-level in such populations. However, the rapid advancement in spatial science disciplines (e.g., Geographical Information Science- GIS) and related location-based technologies, such as remote sensing, Global Navigation Satellite Systems, and environmental and personal sensors, in conjunction with increased computing capabilities for big data analysis could help overcome at least some of the aforementioned challenges (Jia, 2019; Labib, Lindley and Huck, 2020b).

Considering the above arguments, the primary aim of the present study was to develop a new index that allows estimation of aggregate greenspace exposures at hierarchal spatial scales by applying tools and techniques related to GIS, satellite images, and other spatial data. We present a new composite greenspace exposure index (CGEI) that integrates three spatially explicit objective greenspace exposure types (i.e., availability, accessibility, and visibility) for the provision of city- or region-wide measurements of greenspace exposure at fine spatial resolutions. We also investigate statistical associations between single and composite greenspace exposure estimates at the neighbourhood scale in relation to years of potential life lost (YPLL) as a measure of premature mortality. Previous studies indicated that the availability of greenspace reduced mortality and morbidity for diverse populations (Coutts and Horner, 2016; James *et al.*, 2016; Fong, Hart and James, 2018; Rojas-Rueda *et al.*, 2019),

and that greenspace availability varies with the level of social deprivation and inequality in any given neighbourhood (Mitchell and Popham, 2008; Labib, Lindley and Huck, 2020a). We therefore hypothesise that greenspace exposure characteristics (and by extension YPLL) for socially deprived neighbourhoods will differ depending on whether composite or single greenspace exposure measurements are used. We also expect that overall greenspace exposure in such neighbourhoods will reduce YPLL in those neighbourhoods.

2. Methods and Materials

2.1 Study settings

We undertook this study in the Greater Manchester (GM) area of the United Kingdom. Greater Manchester is a post-industrial city-region with an area of 1,276 km² and an estimated population of 2.8 million as of 2016 (Dennis *et al.*, 2018). The city-region has a diverse landscape pattern, including flat plain areas surrounded by hills (the Pennine Chain) that rise to 500 m in the north and east. It has varying greenspaces of different sizes and types such as urban parks, pastoral areas, river corridors, and multiple areas of urban to rural transition (Dennis *et al.*, 2018; Labib, Lindley and Huck, 2020a). In the context of greenspace exposure, the differing characteristics in the area's natural environment is of particular interest for understanding the variability and importance of different types of greenspace exposure. From the perspective of both human health and socioeconomic inequalities, Greater Manchester is also an important area of study in a UK context, because neighbourhoods in the city-region have a range of social deprivation characteristics, from the particularly affluent through to extremely deprived. These levels, in turn, are associated with both varied and elevated levels of illness and disability (GMCA, 2019; Dennis *et al.*, 2020).

2.2 Single greenspace exposure modelling

The composite greenspace exposure index we created in this study represents a combined measurement of three single greenspace exposure metrics (i.e., availability, accessibility, and visibility). We developed each of these individual metrics as distinct measures of greenspace exposure, with each following a different methodological approach using multiple sources of spatial data, tools, and

analytical frameworks. The following sections summarise and discuss the methods applied in developing the single exposure metrics.

2.2.1 Greenspace availability exposure

We created a greenspace availability exposure index (GAVI) as a multi-scale, multi-metric map combining three commonly used greenspace metrics i.e., Normalised Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and Land Use-Land Cover (LULC) at five spatial scales (i.e., 100, 200, 300, 400, and 500 m spatial resolutions - which can be thought of as equivalent to differing buffer distances commonly used to estimate exposure) following the methodology described in Labib et al. (Labib, Lindley and Huck, 2020a). These thresholds were carefully selected considering the spatial scale and resolution of the input data, and are thus less vulnerable to scaling effects (Labib, Lindley and Huck, 2020a). We considered these three metrics to account for different characteristics of greenspace representing photosynthetically active vegetation. NDVI characterises greenspace density, LAI measures greenspace volume, and LULC accounts for the overall presence or absence of greenspace (James *et al.*, 2015; Engemann *et al.*, 2019). NDVI and LAI were obtained from 10 m resolution Sentinel-2 satellite images (collected on 4th July 2018, with cloud cover < 2%). Sentinel-2 satellite images were selected due to their higher overall accuracy in identifying vegetation compared to moderate resolution Landsat images (Markevych *et al.*, 2017; Labib and Harris, 2018). The LULC greenspace data were obtained from Dennis et al. (Dennis *et al.*, 2018). Dennis et al. (2018) combined a supervised classification of Sentinel 2A data from the Copernicus Scientific Data Hub with a range of supplementary data on land cover and land use, including from the Ordnance Survey Greenspace Layer. The LULC data included five land cover types (i.e., urban, water, tree canopy, forbs-shrubs, and grass), with a spatial resolution of 10 m.

For each greenspace metric, we calculated mean values in each grid cell of the map at five spatial scales (i.e., resolution) and combined these five scale maps using a weighted average method to produce multi-scale greenspace availability maps based upon NDVI, LAI and LULC (*Figure S1, Supplementary document*). The multi-scale weighting, based upon lacunarity analysis, accounts for scale sensitivity effects observed at varying buffer distances, a detailed description of which is available

in Labib et al. (Labib, Lindley and Huck, 2020a). Finally, we took the mean of the three maps to produce our multi-metric, multi-scale availability index using equation 1.

$$GAVI_j = \frac{(GS_{NDVI_j} + GS_{LAI_j} + GS_{LULC_j})}{T} \text{ (eq. 1)}$$

In (eq1), $GAVI_j$ is the greenspace availability index value for cell j , and GS_{NDVI_j} , GS_{LAI_j} , and GS_{LULC_j} are multi-scale greenspace metric ‘exposure’ values for corresponding cell j . T is the number of metrics (in our case, $T = 3$). In this case, $GAVI$ values ranged between 0 to 1, where 1 indicated the highest availability of greenspace, and 0 denoted the lowest level of, or no available, greenspace. It should be noted that we removed negative NDVI values from our computation, as these values represented bodies of water or other non-green features. We then set the negative NDVI values to zero. LAI values were normalised using minimum-maximum normalisation to obtain values consistent with other metrics. Details of LAI transformation and calculation can be found in Labib et al. (Labib, Lindley and Huck, 2020a). For this research, we re-sampled the $GAVI$ map to 5 m resolution using nearest neighbour technique in ArcGIS Pro (v 2.4) to achieve consistency with other exposure layers.

2.2.2 Greenspace accessibility exposure

We computed the greenspace accessibility exposure index following a four-step method. First, we selected publicly accessible greenspaces within the study area. Second, we identified the access points for these greenspaces. Third, we ran a network analysis to measure and map the shortest network distance from the access points to any location within the study area. Fourth, we normalised the access distance to produce an accessibility raster map for the study area. The details of these steps follow.

Step 1: Selecting accessible greenspaces

To determine accessible greenspaces, we used two GIS datasets; UK Ordnance Survey (OS) Open Greenspace (scale: 1:2500), and the OS Master Map Accessible Natural Greenspace layer (scale: 1:2500). We selected greenspaces that were publicly accessible to any users within or surrounding the GM area (up to 10 km). This distance was selected following Natural England’s ‘Accessible Natural Greenspace Standards’ (ANGSt) model, which details standard access distances for varying greenspaces sizes in the UK see Comber et al. (Comber, Brunsdon and Green, 2008). We considered

the greenspaces outside the boundary of the study area to account for edge effects, whereby people living by the boundary may have access to nearby greenspaces that are just outside of, or on, the boundary. Our selection included public parks and open gardens, national parks, country parks, local nature reserves, and Countryside and Rights of Way Act (CRoW) accessible greenspaces. A detailed description of the greenspaces is listed in *Table S1, Supplementary document*. We selected these greenspace types based on previous studies evaluating greenspace accessibility in the UK (i.e., including spaces which are generally open to the public; (Pauleit *et al.*, 2003; Comber, Brunsdon and Green, 2008). The size of our selected greenspaces ranged between 0.04 ha (400 m²) to more than 500 ha (5 km²), in order to ensure that we accounted for many forms of greenspace, from small ‘pocket parks’ through to large country parks.

Step 2: Access point identification

In urban areas like GM, greenspaces are usually entered through specific access points (Comber, Brunsdon and Green, 2008). To account for this, we identified the access points for greenspaces using two approaches. Firstly, using the OS Open Greenspace access points layer, we selected the access points that intersect the boundary of the greenspaces. However, this data layer did not contain access points for many natural greenspaces (e.g., country parks). To identify the missing access points, we performed a geometric intersection between the greenspace polygon layer (Step 1) and a road and path network layer (OS MasterMap ITN Layer Urban Paths, Scale 1: 1250) using ArcGIS Pro (v2.5). This intersection operation identified the points where streets and paths intersected the greenspace boundaries and, thus, could be access points. The two datasets were combined for use in this analysis.

Step 3: Network distance analysis

To determine the accessibility of a greenspace from a given location, we calculated the network distance, as this recognises that individuals generally need to follow designated road or path networks in order to reach greenspaces in urban areas (Labib, Lindley and Huck, 2020b). This was achieved by undertaking a ‘service area’ analysis in QGIS (v 3.14) using the QNEAT3 toolbox to measure and map the network distance from greenspace access points to all other locations within the study area, using

the QNEAT3 Isochrone Area (Iso-Area) algorithm. In Iso-Area, the starting points were represented by greenspace access points (identified in Step 2). Using the road and path network layer (OS ITN Layer Urban Paths; details: <https://bit.ly/31GaYlu>) the Iso-Area algorithm produced a network distance surface (raster) from the access points by applying the TIN-interpolation method (toolbox details: <https://root676.github.io/IsoAreaAlgs.html>) to a maximum distance of 10 km based on the upper limit in the ANGSt model (Pauleit *et al.*, 2003; Comber, Brunsdon and Green, 2008). The output raster layer mapped the network distances at a spatial resolution of 10 m for computational efficiency. We masked the raster layer to exclude the extra cells generated outside the Greater Manchester boundary and re-sampled to 5 m resolution to achieve consistency with other layers and datasets.

Step 4: Accessibility index calculation

Utilising the raster grid layer produced in Step 3 (*Figure S2, Supplementary document*), we computed the greenspace accessibility index for each raster grid cell within the study area using equation 2.

$$GACI_j = \frac{(ND_{max} - ND_j)}{ND_{max}} (eq. 2)$$

In (eq2), $GACI_j$ is the greenspace accessibility index expressed as a continuous value for cell j , ranging between 0 and 1, where 0 indicates the lowest accessibility and 1 indicates the highest. ND_{max} is the maximum network distance from the nearest access point within the entire study area (in our case the maximum distance was 5932.7 m); ND_j is the network distance of a cell from the nearest greenspace access points.

2.2.3 Greenness visibility exposure

We measured and mapped eye-level greenness visibility exposure for an observer located at ground level by applying viewshed analysis using new software we developed as part of our analysis (available at: <https://github.com/jonnyhuck/green-visibility-index>). We combined a binary greenness layer produced using LULC data from Dennis *et al.*, (Dennis *et al.*, 2018) with LiDAR based digital surface and digital terrain model data (Environmental Agency, 2016) to perform viewshed analysis at a 800 m viewing distance at 5 m intervals for >86 million observer locations covering the whole study

area. We used a line of sight (LOS) algorithm, in conjunction with an empirical distance decay model, in order to calculate a distance-weighted ratio of visible greenness to visible area from each given observation point. The details of the decay model and LOS algorithm are reported in Labib et al. (Labib, Huck and Lindley, 2021). Based on the LOS and decay weight values, we used equation 3 to compute the viewshed greenness visibility index for the study area. The details of the data, algorithms, steps, and modelling process can be found in Labib et al. (Labib, Huck and Lindley, 2021).

$$VGVI_j = \frac{\sum_1^n G_i \times d_i}{(\sum_1^n G_i \times d_i) + (\sum_1^n V_i \times d_i)} \quad (eq.3)$$

Where, $VGVI_j$ is the ‘viewshed greenness visibility index’ value for the observer cell j ; G_i is the visible green cell, V_i is the visible non-green cell, and d_i is distance decay weight corresponding to visible cell i . The estimated VGVI values ranged between 0 and 1, where 0 = no visible green, and 1 = all visible green (Labib, Huck and Lindley, 2021).

2.4 Composite Greenspace Exposure Index (CGEI)

We combined the three single exposure indices into a composite index value for greenspace exposure. The resulting composite greenspace exposure index is produced by calculating the mean of the GAVI, GACI, and VGVI datasets. We used (equation 4) to produce the CGEI map.

$$CGEI_j = \frac{(GAVI_j \times w_{GAVI} + GACI_j \times w_{GACI} + VGVI_j \times w_{VGVI})}{(w_{GAVI} + w_{GACI} + w_{VGVI})} \quad (eq.4)$$

In (eq4), $CGEI_j$ is the composite greenspace exposure index value for cell j ; $GAVI_j$, $GACI_j$, and $VGVI_j$ represent the respective availability, accessibility, and visibility index values for cell j obtained from single exposure maps; w represents the weighting value for each metric. The resulting operation produced the composite greenspace exposure index map at 5 m spatial resolution for the study area. Note that in this study we have used an equal weighting ($w = 1$) for each of the three single exposure metrics. Equal weighting was used in this case as currently there is no empirical evidence supporting alternative weightings for these single greenspace metrics in relation to their potential effects on premature mortality. However, the CGEI can also be estimated using differing weights for each individual metric. To illustrate the CGEI estimation process using different weightings for individual

metrics we experimented with several hypothetical weight values see *Supplementary note-1* for further details.

2.5 Assessment of CGEI and single exposure indices

The potential implications of the composite greenspace exposure metric reported in the present study were evaluated in two ways. Firstly, we investigated how the composite greenspace exposure measurement and its individual components each related to neighbourhood socioeconomic deprivation. Secondly, we compared each of the metrics to neighbourhood YPLL. We used English Lower Super Output Area (LSOA) Census units due to the availability of population-level socioeconomic deprivation and YPLL data, as well as several confounding variables. LSOAs contain a mean population of around 1500 and range from 1000 to 3000 and are widely used to represent neighbourhood units in national analyses (Mitchell and Popham, 2008; Daras *et al.*, 2019). In our analysis, we examined 1673 LSOAs with a total population of 2,682,528 and a mean population of 1603 (standard deviation 394).

2.5.1 Exposure variations at different deprivation levels

We obtained socioeconomic and health data from the English Indices of Multiple Deprivation (IMD) 2015 (DCLG, 2015). The IMD dataset includes social deprivation scores for the LSOAs across seven sub-domains (i.e., income, health, employment, crime, education, living environment, and barriers to housing and services). In this study, we used IMD deciles to investigate variations of greenspace exposure in LSOAs with differing deprivation levels. An IMD decile value of 1 indicated the highest level of deprivation and 10 indicated the lowest level of deprivation. We used a box plot to explore the differences in greenspace exposures visually and then conducted an ANOVA with post hoc Tukey Honestly Significant Difference analysis to identify whether the differences in exposure values among the IMD deciles were statistically significant or not.

2.5.2 Associations of greenspace exposures with YPLL

The years of potential life lost outcome variable was extracted from the IMD 2015 health and disability sub-domain. YPLL measures ‘premature death,’ defined as death before the age of 75 from any cause, including death due to disease or other external causes. This indicator was estimated based

on mortality data for the period 2008 to 2012 as the numerator, and the denominator was the 2008-2012 population estimate. The details of YPLL estimation can be found in the IMD technical report (DCLG, 2015). YPLL is age and sex standardised, therefore it is not vulnerable to biases such as disproportionate representation by a certain demographic section, or double counting (DCLG, 2015; Dennis *et al.*, 2020). Additionally, it is also weighted by the age of the individual who has died, reflecting the higher impact of an unexpected death of a younger person, than an older one. It should be noted that YPLL is a population-level inference of premature mortality, which has been aggregated to the neighbourhood level (LSOA) for the purposes of this analysis. The associations that we observe are therefore at the population-level and the effects are not transferable to individuals or to different spatial scales of analysis.

Pearson correlation coefficients (r) were calculated to assess correlations among greenspace exposure metrics, YPLL, and neighbourhood socioeconomic variables. Additionally, a hierarchical cluster-based 'heat map' visualisation was produced based on the correlation coefficients to identify clustering among correlated variables (method details in Zhang *et al.*, 2017). Most previous studies have used non-spatial models to explain the associations between exposure and health outcomes (Mitchell and Popham, 2008; Dadvand *et al.*, 2015; Dennis *et al.*, 2020). In this study, we have applied both non-spatial and spatial models to investigate the associations between greenspace exposures and YPLL based on frequentist statistical inferences. Results from non-spatial regression models indicating associations between greenspace exposures and YPLL are presented in *Note-2, Supplementary document*. In our non-spatial models, we applied the generalised linear model (GLM) with a gamma distribution and an identity link function to account for positive values of YPLL that could skew distributions. We have also performed multiple linear regressions to aid comparison between the non-spatial and spatial models. While the GLM allows non-normal error distribution of the residuals, the residuals may have spatial autocorrelation, which might violate the assumption of independence of error distribution. We therefore computed Moran's I statistic (Moran, 1950) for the model residuals in order to identify the presence of spatial autocorrelation in non-spatial models. All the non-spatial models indicated significant spatial autocorrelation in the regression residuals (*Note-2, Supplementary*

document), thus demonstrating spatial dependency and violating the model assumption of independence of observations and inflating the model parameter estimates. As a result, these results are not reported here.

We fitted spatially explicit models to account for potential spatial autocorrelations and spatial dependencies among the variables in the models. As such, we used a ‘spatial autoregressive with autoregressive error’ (SARAR) model, which accounts for spatial dependence in both independent and dependent variables (Kelejian and Prucha, 2010; Bivand, Pebesma and Gómez-Rubio, 2013). As both our dependent and independent variables were spatially explicit, SARAR provides a better fit for the model than other spatial regression models such as spatial lag and spatial error models. The details of SARAR modelling can be found in *Note-3, Supplementary document* and Kelejian and Prucha (Kelejian and Prucha, 2010). In our modelling we used a first order queen contiguity spatial weight matrix, this decision was informed by a previous study see Anselin and Rey., (Anselin and Rey, 2014) and our experimental observations of model fit.

We formulated several regression models sequentially for both the spatial approaches, in order to explore the associations between YPLL and greenspace exposure for all of the LSOAs (all population model). In this regard we considered unadjusted and adjusted models. Model 1 is un-adjusted where each greenspace exposure metric was entered in separate models; Model 2 was adjusted for confounders. Model 2 included variables from Model 1 and the confounders. The selection of confounders was informed by a previous systematic review of the relevant literature (Labib, Lindley and Huck, 2020b) and reference to several original studies (Richardson and Mitchell, 2010; Wheeler *et al.*, 2015; Dennis *et al.*, 2020). In the fully adjusted models, we included income scores, crime deprivation scores, barriers to housing-service scores, distance to nearest general medical practice (GPs), the Shannon diversity score, and annual average PM₁₀ concentration as confounders. Income, crime, and barriers to housing and service scores acted as proxy indicators for income levels, crime rates, physical and financial accessibility of housing and local services together. Distance to GP represents access to health services (Ensor and Cooper, 2004; Daras *et al.*, 2019). The Shannon diversity score indicated the overall land cover mix within the study area as a proxy for vegetation quality (Dennis

et al., 2020). Finally, the PM₁₀ concentration indicated one element of the ambient air quality of these neighbourhoods since there may be an additional health burden from air quality which is not accounted for by greenspace benefits (Medina *et al.*, 2004; Jeanjean, Monks and Leigh, 2016; Khomenko *et al.*, 2021). We also tested sensitivity to other confounders such as employment and the education deprivation score, both of these demonstrated high multicollinearity with income deprivation whilst not improving the models, and so were not included. We assessed each model's explanatory power (pseudo R²), Akaike Information Criterion (AIC), and likelihood ratio test (LR) in evaluating the model's goodness of fit, prior to determining the optimum selection of confounders and most robust models.

In addition to population analysis for all the neighbourhoods (LSOAs) in the study area, we explored whether the association between aggregate greenspace exposure and YPLL varied by sub-groups at different levels of deprivation. We used the IMD score to classify each LSOA, and hence its population, into multiple deprivation quintile groups (*Table S2, Supplementary document*). For each group, we applied the same modelling structure and adjusted for the same confounders as previously described for the pooled analysis. Statistical significance was set at $p < 0.05$, and the v.3.6 R (R Team, others and R Core Team, 2013) statistical package was used for all statistical analyses. The code and results can be found in GitHub (<https://github.com/labiblm/CompositeGreenspaceExposure>).

3. Results

3.1 Greenspace exposure distribution and variations

The greenspace exposure index (single and composite) maps are presented in Fig. 1. It is clear from Fig. 1 that greenspace availability (Fig. 1A) and visibility (Fig. 1C) have a similar broad spatial pattern: exhibiting high exposure values at the rural periphery of the study area and along riparian corridors; and low exposure values in the urban cores. However, whilst the broad patterns are similar, the detailed patterns differ considerably due to the presence of buildings or other obstructing features that affect greenspace visibility. Fig. 1 also illustrates that the accessibility index (Fig. 1B) exhibits a rather different spatial pattern to the other two measures. We found relatively higher accessibility exposure values in and around dense urban cores, likely because these areas typically have many small

to medium-size public parks and gardens (e.g., sized 2-20 ha), and good transport links. However, these urban parks and gardens did not result in higher availability or visibility exposure values due to their fragmented pattern and low connectivity. Fig. 1, therefore, demonstrates the differences in exposure values depending on how exposure is conceptualised and measured. Fig. 1(D) illustrates the composite greenspace exposure index map. This map is a combined output of the three single indices, and each cell of this map indicates the mean exposure value of the three exposure surfaces as our measure of aggregate greenspace exposure.

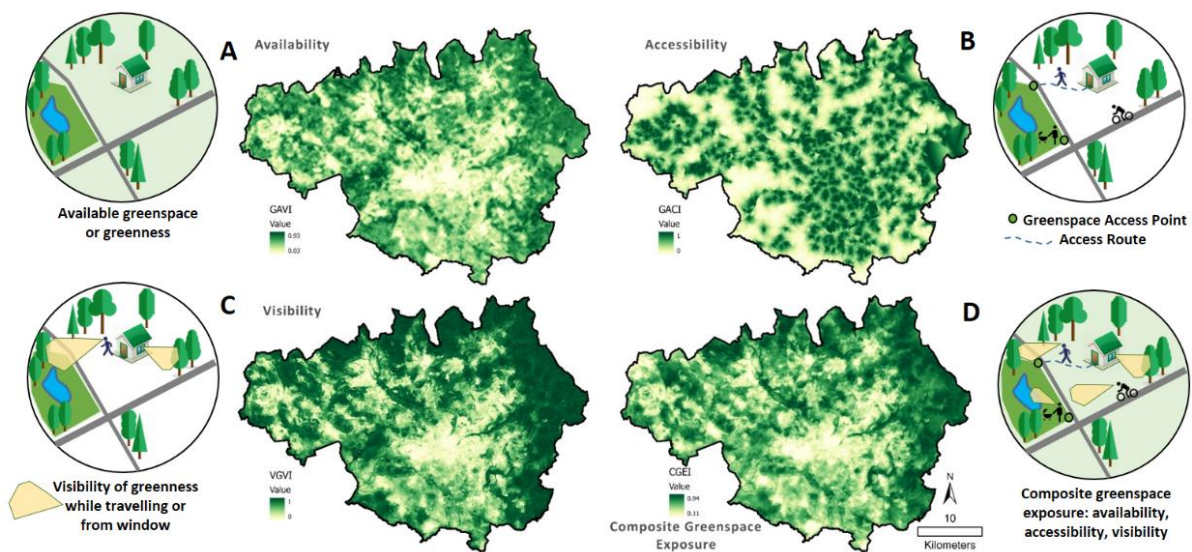


Fig. 1: Maps showing the single and composite greenspace exposure metrics (A) availability, (B) accessibility, and (C) visibility. (D) is the composite greenspace exposure map calculated by averaging three single exposure maps listed in (A-C). The conceptual representation of each type of exposure showed differences in measuring and modelling single exposure type and the composite is a summary of all three exposures together.

The relationship between the three input indices and the resulting composite index is also illustrated for a terraced house located near to a park (Fig. 2). Due to the close proximity of the park (only a few streets away), the house has a very high accessibility exposure value (0.957). Conversely it has a much lower visibility value (0.065) as the park cannot be seen due to the local terrain and intervening buildings. Finally, the house has a relatively low value for greenspace availability (0.225) because of the densely built-up nature of the surrounding area. The composite greenspace exposure

index value (CGEI = 0.416) provides an aggregate overall moderate score for greenspace exposure for this particular house. Therefore, the composite greenspace exposure metric represents a comprehensive measurement that accounts for all three objective exposure types together.

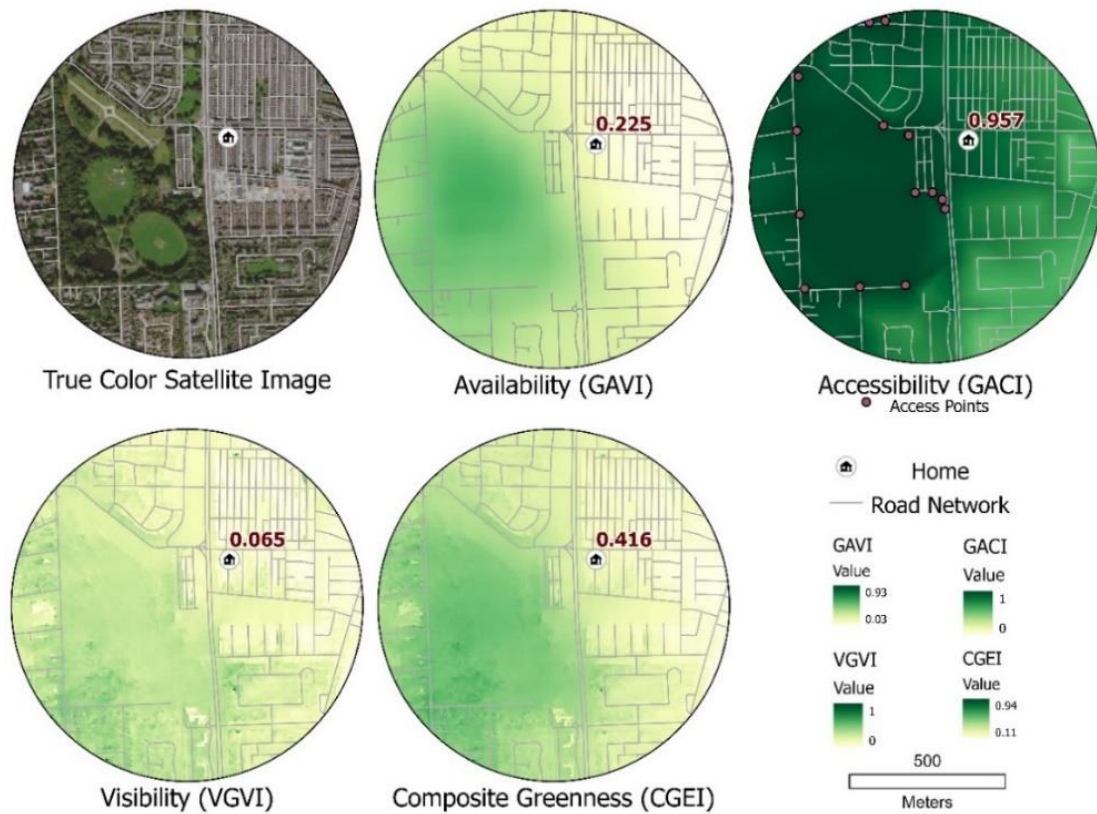


Fig. 2: Differences in greenspace exposure values for an example residence. GAVI shows availability exposure value, GACI shows accessibility exposure value, VGVI presents visibility exposure and CGEI shows the composite greenspace exposure value for the example residence.

3.2 Deprivation and greenspace exposures

The analysis of different greenspace exposure metrics and social deprivation deciles at the neighbourhood (LSOA) scale indicates variations in how greenspace exposures co-vary with social deprivation (Fig. 3). Figure 3 illustrates that the least deprived areas have greater greenspace exposure compared to the most deprived areas in the domains of greenspace availability, visibility, and composite greenspace exposures. In the case of accessibility, however, we observed higher exposure values for more deprived areas, although with a smaller level of variation across the IMD deciles (*summary*

descriptive statistics Table S3, Supplementary document). This result illustrates accessibility exposure in our case study area has a different spatial distribution than availability and visibility based on existing greenspaces, parks and gardens. In the case of Greater Manchester, the most deprived neighbourhoods often have public parks and gardens in close proximity, but these neighbourhoods have low greenspace visibility and availability exposure because they are more densely built-up areas with low provision of private gardens and streets with low levels of canopy coverage. Conversely, the least deprived neighbourhoods usually exhibit lower building density, greater provision of private gardens, and streets with greater levels of canopy coverage, but with less provision of public parks and gardens in proximity. These results demonstrate that apparent inter-group variations in individual greenspace exposure measures can be misleading, and therefore the benefit of this composite measure.

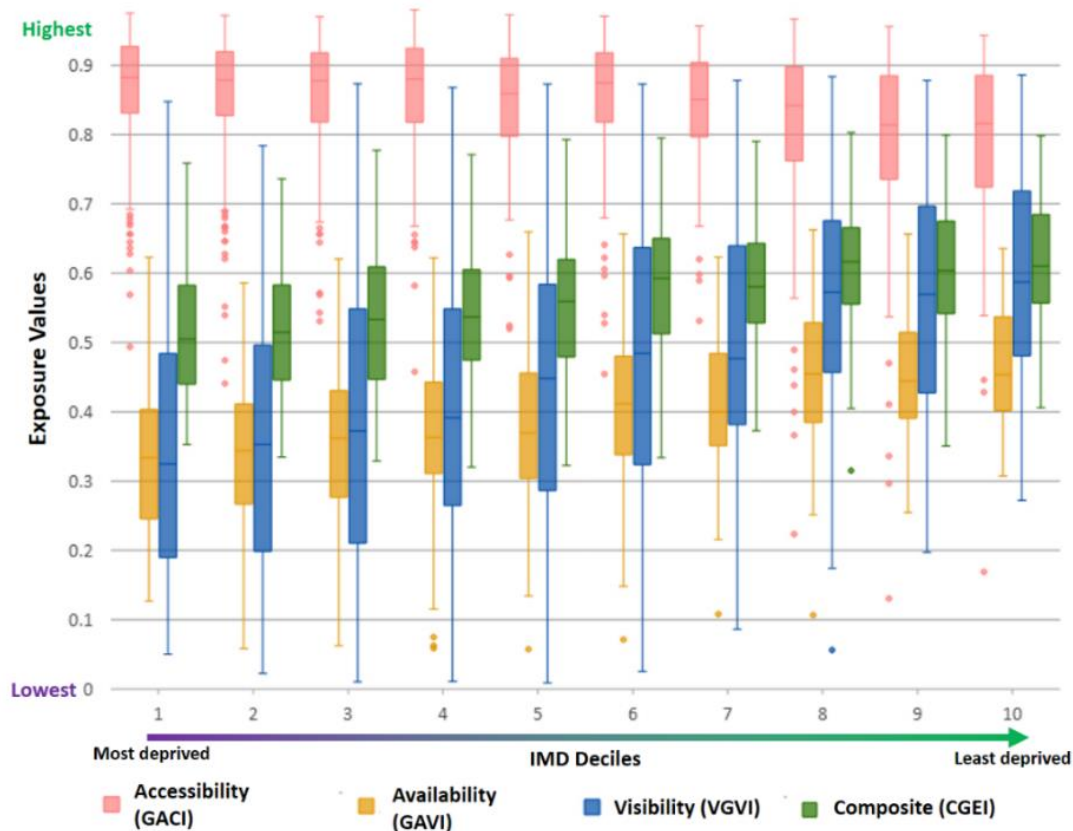


Fig. 3: Greenspace exposure indices compared to different IMD deciles. Dots indicate outliers in the measurements of greenspace exposures.

The ANOVA analysis demonstrates that the differences in greenspace exposures among different deprivation deciles are statistically significant for all greenspace exposures (*Table S4*,

Supplementary document). The post-hoc analysis identifies an interesting pattern in which differences are usually only significant in the case of the difference between the most and least deprived deciles. However, among similar deprivation deciles (e.g., 1-3 or 7-9) the differences in greenspace exposures are not statistically significant (*Table S4, Supplementary document*). This result implies that neighbourhoods that share similar deprivation levels usually demonstrate similar characteristics in terms of greenspace exposure.

3.3 Associations among greenspace exposure types and YPLL

Correlation analyses showed that there were very strong positive correlations among composite, availability and visibility metrics for the Greater Manchester neighbourhoods. By contrast, greenspace accessibility only had a small correlation with the composite metric, and a weak negative correlation with the availability and visibility exposure metrics (Fig. 4A). The ‘heat map’ also illustrates a similar pattern, as availability, visibility and the composite metric are clustered in the same group, indicating high similarity among these metrics; whereas the accessibility metric was not included in this cluster with the other greenspace metrics (Fig. 4B). Such divergent patterns of greenspace accessibility measure were also observed in the spatial distribution greenspace exposure maps shown in Fig. 1.

The correlation analyses and the correlation ‘heat maps’ further indicated that the availability, visibility, and composite greenspace exposure measures have moderate to weak negative correlations with YPLL and are in contrasting cluster groups with it as well (Fig. 4B). Additionally, these metrics have negative correlations with deprivation scores and the air quality variable, but they have positive correlations with distance to GP (Fig. 4A). In contrast, the accessibility greenspace exposure metric has a positive correlation with YPLL, deprivation score, and air quality variables, all of which are clustered in the same group (Fig. 4B).

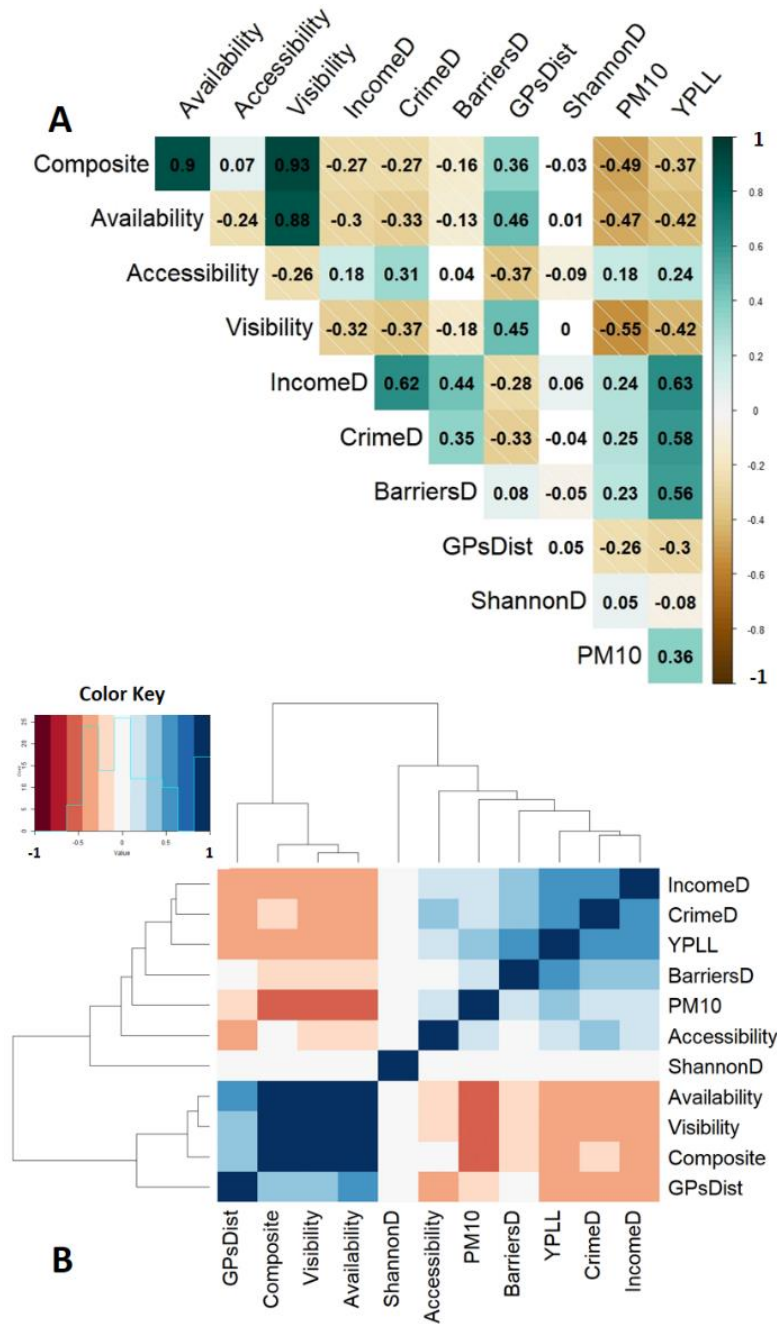


Fig. 4: (A) Correlations among greenspace exposure types, neighbourhood premature mortality (YPLL), socioeconomic variables, and the physical environment (i.e., land use mix, and air quality). Green boxes indicate positive correlation coefficients and brown boxes indicate negative correlation coefficients. Unshaded boxes indicate no significant correlations, shaded boxes indicate significance at $p < 0.05$. (B) correlation coefficient ‘heat map’ with clusters of variables. Blue boxes indicate similarity between clusters and observations, red boxes indicate dissimilarity of observations and dendrogram lines illustrate the hierarchical clusters among variables.

Regression analysis results are presented in Fig. 5 and Table 1. Similar to the correlation results, Fig. 5(A) indicates significant inverse associations of availability, visibility, and composite greenspace exposure with YPLL both for unadjusted and adjusted models. These associations are generally attenuated to some extent in the adjusted models. The results indicate increasing availability, visibility, and composite greenspace exposures were associated with lower YPLL values at the population level analysis. By contrast, for greenspace accessibility exposure, we found increasing accessibility was associated with higher YPLL values. In the unadjusted model, the effect is significant, but following the adjustment, the effect attenuates and becomes insignificant (Table 1). This result implies that, at the population level, accessibility exposures do not significantly relate to YPLL in the case study area. Overall, these findings meant that the association between greenspace exposure and YPLL, or premature mortality, differed across different types of greenspace exposure metrics. Therefore, measuring exposure using a single metric might not present a complete picture of the influence of greenspace exposure on health (e.g., premature mortality).

Table 1: Greenspace exposure and YPLL relationships using spatial regression models. The ‘Effect Estimate’ values are the numbers of years of potential life lost associated with unit greenspace exposure.

Greenspace exposure	Model Adjustment	Model outcomes			R ²
		Effect Estimate	95% CI (low; up)	p	
Availability	Unadjusted	-13.03***	-16.59; -9.48	6.661E-13	0.577
	Adjusted†	-9.8***	-15.90; -3.06	0.0037	0.702
Accessibility	Unadjusted	5.95***	2.47; 9.44	0.0008	0.565
	Adjusted	4.02	-3.94; 11.98	0.3228	0.70
Visibility	Unadjusted	-6.80***	-8.8; -4.80	2.64E-11	0.576
	Adjusted	-6.06***	-10.36; -1.75	0.0057	0.702
Composite	Unadjusted	-12.46***	-16.35; -8.56	3.621E-10	0.573
	Adjusted	-9.95***	-17.89; -2.00	0.014	0.702

CI: Confidence interval. *** indicate statistical significance at $p < 0.05$. R² is Nagelkerke pseudo-R-square. †model adjusted for: Income score, Crime score, Barriers to housing score, Shannon diversity

score, Distance to nearest general medical practice, and yearly average air pollution (PM₁₀). (Detailed model outputs can be found in Note-3, *Supplementary document*)

The adjusted CGEI exposure model indicates that a single unit increase (from CGEI 0 to 1, or 100%) in composite greenspace exposure is associated with a reduction of approximately ten years in the YPLL value. Thus, a 10% or 0.1 increase in the composite greenspace exposure value is associated with an approximately one year lower premature mortality value. It should be noted that both availability and the composite exposure models have nearly identical effect size and R² values (~ 0.702, Table 1), meaning that they both explain similar levels of variability in YPLL after controlling for confounders. This is partly a result of similarities between the availability and visibility metrics caused by the fact that 'visible' greenspace (a measure of line of sight) is also likely to be 'available', though the inverse is not necessarily true. The result of this is that the effect of the availability and visibility surfaces upon the composite are quite similar, meaning that (in the case of equal weighting of the three inputs in eq.4) the composite surface resembles both of them quite closely (Fig 4A). In addition, the exposure metrics and health indicator were aggregated at neighbourhood scale. The aggregated information may have reduced variance, which in turn affected the explanatory power of the models. Based on these results, we therefore cannot fully confirm which metric is the better predictor of neighbourhood level YPLL for our study area. However, despite having very similar R² values, the composite metric when compared with all the individual exposure metrics (Fig. 5A), has a slightly greater effect size (Table 1). In addition, conceptually the CGEI measures holistic greenspace exposure. Considering the preceding we used the CGEI metric for subgroup analysis in the case of each socioeconomic deprivation group.

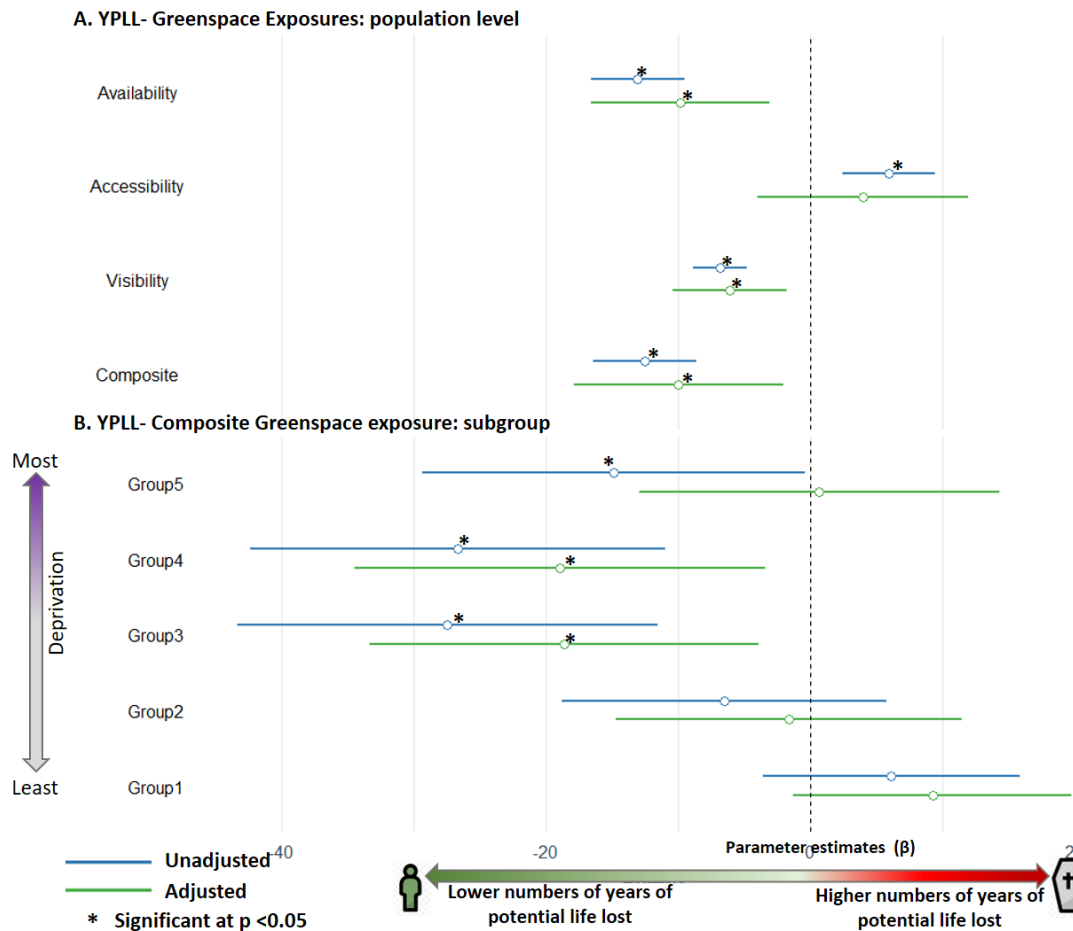


Fig. 5: The YPLL parameter estimates greenspace exposure relations for all neighbourhoods with a 95% confidence interval (A), and neighbourhoods stratified by different socioeconomic deprivation groups (B). The positive parameter estimate values indicate higher numbers of years of potential life lost, and negative values indicate lower numbers of years of potential life lost at the population and subgroup levels. (Detailed model outputs can be found in Note-3, *Supplementary document*)

Figure 5B shows variations in the effects of composite greenspace exposure on YPLL for people living in areas with different socioeconomic deprivation levels. Increasing greenspace exposure is significantly inversely correlated with premature mortality among people living in areas of greater deprivation (groups 3-4) compared to those in the least deprived areas (groups 1-2). It should be noted that there is no significant composite greenspace exposure effect on YPLL for people living in areas of relatively low or no deprivation (groups 2, 1). Additionally, for people living in the least deprived areas (group 1), the effect of composite greenspace exposure on YPLL is the opposite of the hypothesised

direction, implying that in the least deprived areas, increased greenspace may be associated with increased YPLL, though this effect is not statistically significant. For group 5 greenspace exposure is not significant in the adjusted model. It is possible that in the most deprived areas, the higher premature mortality rate may be attributed to other reasons that we did not consider in the models (e.g., unhealthy food habits, race/ethnicity) (Cecchini *et al.*, 2010). Nevertheless, overall we have observed a deprivation-related gradient in the effect of greenspace exposure on premature mortality.

4. Discussion

4.1 Multiple spatially explicit measures of greenspace exposures

To our knowledge, this is the first study that estimates greenspace exposure comprehensively by combining multiple spatially explicit greenspace exposure measures into a composite metric. Previously (Rugel *et al.*, 2017) developed a composite natural space index at a fixed spatial scale (i.e., postcode) combining multiple measures of greenness. The methodological approach that we have developed in this study has improved upon this fixed scale approach by standardising and harmonising different greenspace exposure measures in order to produce a composite metric that estimates aggregate greenspace exposure at hierarchical spatial scales and fine spatial resolutions. We named this the composite greenspace exposure index (CGEI). We argue that the greenspace exposure modelling and mapping methods we developed in this study offer the opportunity to provide comprehensive estimates at the population level and provide the basis to infer individual-level aggregate greenspace exposures at meaningful temporal and spatial scales.

CGEI differs from traditional single exposure metrics in that it can capture objective greenspace exposure holistically and combine multiple inter-twined greenspace exposure types. Such a holistic exposure measure allows for a better understanding of the overall impact of greenspace exposure upon health outcomes such as mortality, morbidity, and non-communicable diseases that may be linked with multiple greenspace exposures. However, it should be noted that the composite is an aggregated expression of the single greenspace indices. Although it provides generalised exposure information it obviously cannot determine the specific cause-effect relations of each exposure type to specific health

outcomes. For example, visibility may be more important in understanding the impact on a specific mental health outcome such as stress recovery, (Ulrich, 1984; Kaplan, 1995), than total greenspace exposure as represented in the composite measure. We therefore argue that the use of the CGEI does not diminish the importance of the use of individual exposure maps, as they provide the opportunity to understand specific types of exposure and, hence, which pathways correspond to which health indicators. As a result, these metrics are complementary to one another when evaluating the health impact of greenspace.

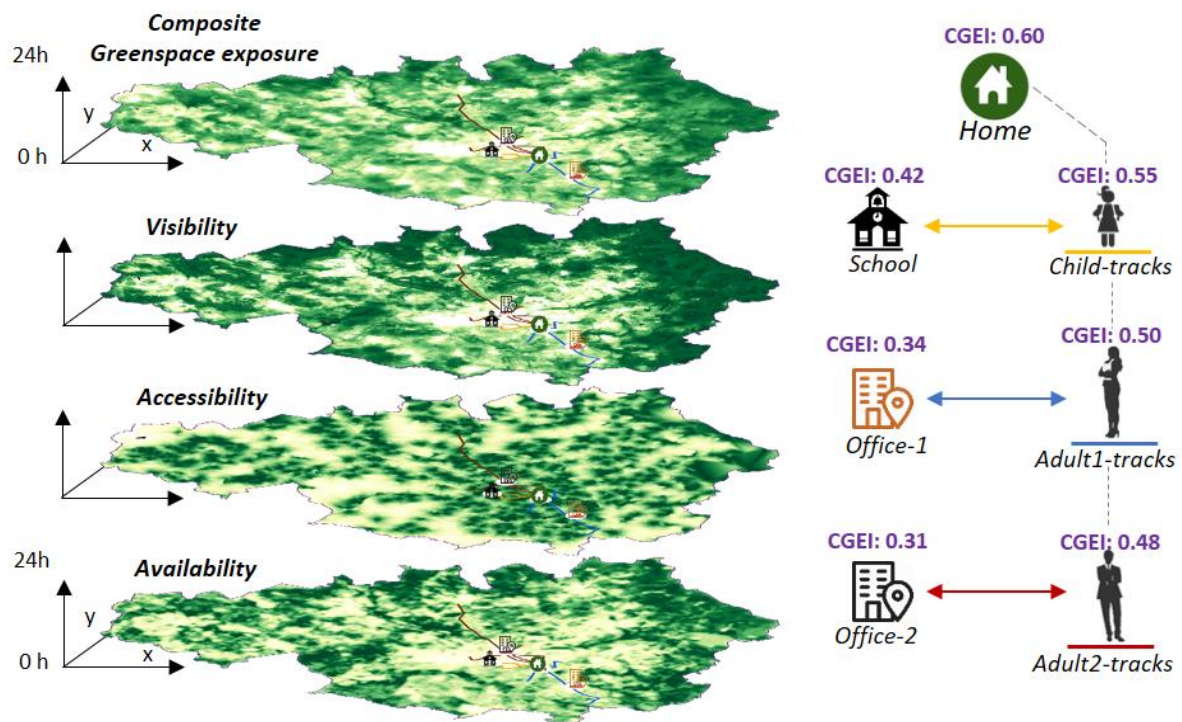


Fig. 6: A schematic illustration of space-time mobility-based greenspace exposures for three individuals living in the same household, and the varying exposure profiles each experience depending on the characteristics of their daily movement patterns.

Greenspace exposure maps (both composite and single metrics) can be used to assess greenspace exposures for places where people live (e.g., home), work (e.g., office), or go about their daily activities (e.g., school, shops) (Fig. 6). In addition to these fixed locations, the results allow the estimation of individual exposure profiles by taking account of people’s everyday movement (e.g., using GPS tracks) and the amount of time they spend in different locations (*details in Note-4,*

Supplementary document). In Fig. 6, we provide the example of a hypothetical household of three people (two adults, one child) with a CGEI value of 0.60 for the home. However, each household member spends differing amounts of time in various locations, and each member has different travel routes. The child travels to school (CGEI 0.42) along streets exhibiting moderate levels of green cover. The child has an overall CGEI of 0.55 based on time spent at home, school, and during periods of movement. Similarly, the first adult (adult1) travels to the office (office-1) (CGEI 0.34), taking a route that exhibits relatively low greenspace exposures for an overall CGEI of 0.501. Finally, the second adult (adult2) travels to a different office (office-2) (CGEI 0.31) along a route with low greenspace exposure, providing an overall CGEI of 0.48.

In the illustrative example (Fig. 6), we demonstrate how the mapped outputs could be readily used to generate aggregate greenspace exposure estimates for individuals in households with different exposure profiles. In turn these individualised exposure profiles could be linked with individualised health data. While the composite provides an estimation of aggregate greenspace exposures, the single exposure maps can nevertheless provide crucial insights into the different types of greenspace exposure that an individual may experience at different spatiotemporal scales. This type of space-time exposure profiling has been utilised in other exposure studies relying on activity-based exposure measurements (Briggs, 2005; Zhang *et al.*, 2018), but multiple exposure pathways have not been considered. Our new approach reduces existing methodological challenges in objective greenspace exposure assessments that rely on single exposure metrics, including measurement of exposure at fixed spatial scales, the uncertain geographic context problem, and issues with heterogeneous methods of exposure assessments across studies (Kwan, 2012; Turner *et al.*, 2017; Helbich, 2018; Labib, Lindley and Huck, 2020b).

4.2 Impact of greenspace exposure types on premature mortality and deprivation

We tested the effects of greenspace exposure in relation to neighbourhood-level socioeconomic deprivation in its relation to years of potential life lost (YPLL), a measure of premature mortality. We found that areas with higher availability and visibility exposures are usually the least deprived, and the most deprived areas have comparatively low greenspace exposures according to the same metrics. When we considered accessibility exposure. The results demonstrated the opposite trend, indicating

that the most deprived areas have greater accessibility to public greenspaces (Dennis et al., 2020), than the least deprived due to the frequent proximity of small or medium size (e.g., 2-20 ha) public greenspaces. We speculate that such patterns could be attributed to the nature of the development of post-industrial towns in England, where parks were historically built by mill owners amongst dense terraced housing in order to support the factory workers, following the Open Spaces Act of 1877 and Public Health Act 1875 (Clark, 1973; Jordan, 1994). In such towns and cities, such dense terraced housing areas typically present with higher deprivation values. Our accessibility result is therefore consistent with previous work suggesting that greenspace accessibility is greater in areas of greater socioeconomic deprivation in the UK (Barbosa *et al.*, 2007; Jones, Hillsdon and Coombes, 2009).

When comparing the mean composite greenspace exposure values for neighbourhoods at differing socioeconomic levels, we observed a deprivation-related gradient in greenspace exposure, in which the least deprived areas exhibited higher composite greenspace exposures than the most deprived areas. These results align with previous research showing that greener neighbourhoods are also less deprived communities (Mitchell and Popham, 2008; Rigolon *et al.*, 2021). However, we show that the composite measure presented here can estimate and differentiate greenspace exposures to a greater extent than the individual single exposure metrics which are typically used to evaluate the relationships between inequality and greenspace exposure. The results of the present study also suggest that the composite index has the potential to reveal new or refined health associations in previously reported studies which have so far only used traditional (e.g., percentage of neighbourhood greenspace) or individual metrics (e.g., accessibility). As we have demonstrated, single exposure metrics measuring differing factors such as accessibility or availability may provide results that appear different or contradictory (Fig. 3). As a result, single metric-based results can potentially misrepresent the true level of differences in greenspace exposure experienced by neighbourhoods at different levels of socioeconomic deprivation.

When we evaluated the associations between greenspace exposure and premature mortality (i.e., YPLL), the results showed that individual greenspace exposure metrics (i.e., availability, accessibility, and visibility) provided considerable variances in effect sizes (e.g., effect size for

availability: -9.80, visibility: -6.06) and directions regarding the associations between exposures and outcomes (Table 1). Both the availability and visibility greenspace exposure metrics indicate significant negative associations and imply that higher levels of greenspace availability or visibility are associated with lower YPLL (reduced premature mortality). These findings are consistent with previous studies indicating greater availability or visibility of greenspace can have positive health benefits, including reductions in mortality (Fong, Hart and James, 2018; Kondo *et al.*, 2018; Rojas-Rueda *et al.*, 2019).

Our results, which show no significant associations between greenspace accessibility and YPLL, are consistent with several previous studies reporting associations between greenspace exposure and health outcomes (Richardson *et al.*, 2010; Klompaker *et al.*, 2018; Jarvis *et al.*, 2020; Labib *et al.*, 2020) but not with others (Coutts, Horner and Chapin, 2010; Wilker *et al.*, 2014; Dennis *et al.*, 2020). These latter studies reported positive health effects related to increased greenspace accessibility. Such inconsistent results might be linked to different approaches to conceptualising and measuring accessibility (e.g., shortest distance vs. fixed distance) between the present study and previous studies (Jarvis *et al.*, 2020; Labib, Lindley and Huck, 2020b). It must also be noted that several other studies have argued that the positive effect of greenspace accessibility may be context-specific (Jones, Hillsdon and Coombes, 2009; Richardson *et al.*, 2010). For example, proximity to greenspaces does not always ensure greater use in many communities if areas have poor walkability (Zuniga-Teran *et al.*, 2019) or where the poor perception of greenspace may deter its use (Jones, Hillsdon and Coombes, 2009). Therefore, even if greenspace accessibility is high, if that greenspace attracts few visits then the associated proactive health benefits might not be significant.

Despite these inconsistent associations when using single metrics, in line with our hypothesis, we found increasing overall community greenspace (i.e., measured as composite greenspace exposure) was associated with a lower level of premature mortality. Our overall findings are consistent with previous reports of the positive impact of greenspace exposure associated with reductions in different types of mortality among people living in greener communities (Mitchell and Popham, 2008; Coutts and Horner, 2016; James *et al.*, 2016; Rojas-Rueda *et al.*, 2019).

Our research suggested that variations in the effects of different greenspace exposure measurements might link with the underlying mechanistic pathways associated with each exposure metric (Marselle *et al.*, 2021; Zhang *et al.*, 2021). For example, greenspace availability (i.e., the overall amount of vegetation in an area) might reduce the effect of harmful environmental stressors such as air pollution or heat (Markevych *et al.*, 2017; Lindley *et al.*, 2019) and reduce premature death. Greenspace accessibility (i.e., the relative proximity of public green spaces) may be associated with building capacities by encouraging physical activity and social cohesion (Hartig *et al.*, 2014; Markevych *et al.*, 2017). However, in our case study area at the population level, such mechanisms might not be as effective as other pathways. This conclusion was supported by the insignificant association between accessibility and YPLL in the adjusted model. Visibility of greenspace may be linked with restoring capacities related to psychological mechanisms such as stress recovery that reduce morbidity (Ulrich, 1984; Frumkin *et al.*, 2017; Markevych *et al.*, 2017). In the present study, the effect size of visibility was even smaller than that for availability, indicating the positive effect of visible greenspace might have an even lower relative impact than availability in lowering premature mortality. Overall, the composite greenspace exposure showed a slightly larger effect size than any individual metric in the adjusted model (Table 1). As the composite greenspace exposure accounts for all three well-known important mechanistic pathways through which nature influences health (Markevych *et al.*, 2017), it is reasonable to argue that composite greenspace exposure might have a stronger relationship with premature mortality than has previously been considered when based only on individual metrics (e.g., visibility, accessibility). This finding is critical because it implies that many health-related indicators such as mortality, morbidity, and other non-communicable diseases and health syndromes might be influenced by the aggregate impact of multiple greenspace exposure types due to the multiple associated exposure benefit pathways.

We also found significant differences in the relationship between composite greenspace exposure and YPLL among groups at different socioeconomic deprivation levels. Our results indicate that increasing greenspace exposure in areas of moderate to extreme deprivation corresponds to a lower YPLL (reduced premature death) than in the least deprived areas. These findings are consistent with

published studies that suggest that the health benefits of increased greenspace exposure are more pronounced among low-income and socioeconomically deprived populations (Marmot, 2020; Rigolon *et al.*, 2021; Wu and Kim, 2021).

4.3 Methodological approach in multiple spatially explicit exposure modelling

In this study, we developed a new methodological solution for measuring aggregate greenspace exposures at hierarchical scales and fine resolutions. The complete measurement of a set of environmental exposures can be linked with the broader framework of the ‘*Exposome*,’ which considers “the totality of human environmental exposures from conception onward” (Wild, 2012). The exposome concept includes all external and internal exposures throughout a human lifespan (Wild, 2012; Vrijheid *et al.*, 2020). Analysis of a subset of external greenspace exposures might provide a more useful framework for studying the associations between aggregate greenspace exposure and human health, as opposed to using a single type of greenspace exposure. The methodological process we introduce in this study illustrates an approach that allows a measurement of the summation of the multiple spatially explicit exposures and allows capturing high spatial-temporal exposure variability at multiple scales. Therefore, a spatially explicit composite measurement could be considered a step toward a “*spatial exposome*” and results from such measurements can be utilised to support the broader exposome framework. A spatial exposome allows for spatial hierarchies in aggregate exposure assessments for any given set of external environmental exposures, which can then, in turn, allow the inference of comprehensive exposure profiles for individuals as well as for populations of varying sizes. While our composite exposure metric provides an introductory example of the spatial exposome concept, it will require further development in terms of other sets of external exposure variables (e.g., air, noise pollution) and time points.

4.4 Limitations and future developments

Although we have evaluated the robustness of our methodology and elucidated the implications for understanding premature mortality and socioeconomic inequality, our study does have some limitations. First, our exposure metrics are spatially explicit objective measurements of greenspace exposure based on the different vegetated areas alone (i.e., they do not include blue spaces).

Furthermore, the measurements primarily represent the quantity of greenspace rather than its perceived quality (which we have used as a confounder). As a result, consideration should be given to integrating objective (e.g., vegetation diversity) or subjective (e.g., attractiveness) measurements of quality with the composite greenspace exposure index in future studies. Such subjective perceptions of greenspace exposure could be collected using individual surveys and integrated with our objective assessment to provide an overall understanding of the effect of both objective and perceived greenspace exposure on health.

Second, to evaluate our new metric, we investigated the direct association between greenspace exposure and premature mortality. We did not conduct an explicit mediation analysis to understand the mediating effect of variables such as air or noise pollution or heat. Although some previous studies have indicated mixed results for, and relatively lower effects from, these variables in terms of their mediation of the associations between greenspace exposure and health (James *et al.*, 2016; Vienneau *et al.*, 2017). In future studies, these variables should be critically evaluated for their potential effects on results generated using our composite greenspace measurement approach applying more explicit mediation and moderation approaches, such as structural equation modelling (Dzhambov *et al.*, 2020).

Third, we used equal weights for each of the single greenspace exposure metrics in the composite greenspace calculations. No pre-existing evidence existed to indicate the precise level of health benefits associated with each type of greenspace exposure based on their differing effect pathways, the individualistic nature of the impacts, and the spatiotemporal variability. However, it would be possible to construct a weighted average for the composite exposure estimation if such information is available. In *supplementary Note-1*, we provided a worked example of weighted composite greenspace exposure index estimation based on a range of hypothetical weights. To generate actual weights for different input layers of the composite in future applications, a meta-analysis or analytical hierarchy approach could be adopted to identify the relative importance of availability, accessibility, and visibility exposures considering different health outcomes.

Finally, our population-level inference of greenspace exposure and premature mortality was made using cross-sectional data, and we had no means of knowing the extent to which individuals

experienced variances in their levels of greenspace exposure throughout their lives. Therefore, we cannot confirm that the relations we observed can be inferred as a causal effect of greenspace exposure on premature mortality. Additionally, the population and exposure information were aggregated at the neighbourhood administrative boundary scale (LSOA). As a result, the correlations observed in aggregate cannot be transferred to individuals or to different spatial scales of analysis, due to the potential presence of zoning effects associated with the modifiable areal unit problem and ecological fallacy (Openshaw, 1984; Goodchild, 2011). Although (Annerstedt Van Den Bosch *et al.*, 2016) noted that spatially aggregated population data is an acceptable alternative to using individual data, we recognise that to fully understand the effect of comprehensive greenspace exposure on individual health, longitudinal and life-course assessments of greenspace exposure need to be accounted for (Turner *et al.*, 2017; Jia, 2019).

5. Conclusion

In this study, we provided a new methodology for estimating multiple greenspace exposure metrics. Our approach has the potential to reduce the heterogeneity inherent in current approaches to greenspace measurements, as it provides an analytical technique that allows multiple quantitative greenspace exposure measurement practices to be harmonised and standardised. This enables a shift from the traditional approach of measuring single greenspace exposure towards a more holistic approach to greenspace exposure estimation at multiple spatial scales. Our analytical approach also allows more spatially explicit aggregate exposure assessment and can be partially linked with the exposome framework. When applying the new index to population-level analyses, we found significant negative associations between community greenspace exposure and premature mortality. The association between increased greenspace exposure and lower premature mortality was particularly marked in areas with higher aggregate greenspace exposure levels. However, such associations differed significantly based on any given community's level of socioeconomic inequality. Overall, our results suggest that composite greenspace exposure measurements can reveal new insights into the public health benefits of exposure to the natural environment. Furthermore, the variations in the level of benefit

derived by individuals from given levels of greenspace exposure may depend on the population of which they are members and their socioeconomic status.

Declaration of interest

We declare that we have no conflicts of interest.

Acknowledgement

We would like to thank the reviewers and editor of this paper for their constructive comments and suggestions. S.M. Labib is funded by the School of Environment, Education and Development Postgraduate Research Scholarship at the University of Manchester. This study was funded by the Natural Environment Research Council, the Arts and Humanities Research Council and the Economic and Social Research Council under the Valuing Nature Programme. NERC grant reference number NE/N013530/1.

Author contributions

S.M. Labib: Conceptualisation, Methodology, Data curation, Formal analysis, Investigation, Software, Writing - original draft, Visualisation, Writing - Review & Editing. **Sarah Lindley:** Conceptualisation, Methodology, Supervision, Resources, Methodology, Writing - review & editing, Funding acquisition. **Jonny J. Huck:** Conceptualisation, Methodology, Software, Resources, Supervision, Writing - review & editing.

References

- Annerstedt Van Den Bosch, M. *et al.* (2016) 'Development of an urban green space indicator and the public health rationale', *Scandinavian Journal of Public Health*. SAGE PublicationsSage UK: London, England, 44(2), pp. 159–167. doi: 10.1177/1403494815615444.
- Anselin, L. and Rey, S. J. (2014) *Modern Spatial Econometrics in Practice: A Guide to Geoda, Geodaspace and Pysal, GeoDaSpace and PySAL-GeoDa Press, Chicago, Estados Unidos de*

América.

Barbosa, O. *et al.* (2007) 'Who benefits from access to green space? A case study from Sheffield, UK', *Landscape and Urban Planning*. doi: 10.1016/j.landurbplan.2007.04.004.

Bivand, R. S., Pebesma, E. and Gómez-Rubio, V. (2013) *Applied Spatial Data Analysis with R: Second Edition*, *Applied Spatial Data Analysis with R: Second Edition*. doi: 10.1007/978-1-4614-7618-4.

Bratman, G. N. *et al.* (2019) 'Nature and mental health: An ecosystem service perspective', *Science Advances*, 5(7). doi: 10.1126/sciadv.aax0903.

Briggs, D. (2005) 'The role of GIS: Coping with space (and time) in air pollution exposure assessment', *Journal of Toxicology and Environmental Health - Part A*. doi: 10.1080/15287390590936094.

Cecchini, M. *et al.* (2010) 'Tackling of unhealthy diets, physical inactivity, and obesity: Health effects and cost-effectiveness', *The Lancet*. doi: 10.1016/S0140-6736(10)61514-0.

Clark, F. (1973) 'Nineteenth-Century Public Parks from 1830', *Garden History*. doi: 10.2307/1586332.

Comber, A., Brunsdon, C. and Green, E. (2008) 'Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups', *Landscape and Urban Planning*. Elsevier, 86(1), pp. 103–114. doi: 10.1016/j.landurbplan.2008.01.002.

Coutts, C., Horner, M. and Chapin, T. (2010) 'Using geographical information system to model the effects of green space accessibility on mortality in Florida', *Geocarto International*. doi: 10.1080/10106049.2010.505302.

Coutts, C. J. and Horner, M. W. (2016) 'Nature and death: An individual level analysis of the relationship between biophilic environments and premature mortality in Florida', in *Spatial Analysis in Health Geography*. doi: 10.4324/9781315610252.

Dadvand, P. *et al.* (2015) 'Green spaces and cognitive development in primary schoolchildren',

Proceedings of the National Academy of Sciences of the United States of America, 112(26), pp. 7937–7942. doi: 10.1073/pnas.1503402112.

Dadvand, P. and Nieuwenhuijsen, M. (2019) ‘Green space and health’, *Integrating Human Health into Urban and Transport Planning: A Framework*, pp. 409–423. doi: 10.1007/978-3-319-74983-9_20.

Daras, K. *et al.* (2019) ‘Open data on health-related neighbourhood features in Great Britain’, *Scientific Data*, 6(1). doi: 10.1038/s41597-019-0114-6.

Davis, Z. *et al.* (2021) ‘The association between natural environments and childhood mental health and development: A systematic review and assessment of different exposure measurements’, *International Journal of Hygiene and Environmental Health*. Urban & Fischer, 235, p. 113767. doi: 10.1016/j.ijheh.2021.113767.

DCLG (2015) *English indices of deprivation 2015*. Available at: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015%0A> (Accessed: 2 January 2019).

Dennis, M. *et al.* (2018) ‘Mapping Urban Green Infrastructure: A Novel Landscape-Based Approach to Incorporating Land Use and Land Cover in the Mapping of Human-Dominated Systems’, *Land*, 7(1), p. 17. doi: 10.3390/land7010017.

Dennis, M. *et al.* (2020) ‘Relationships between health outcomes in older populations and urban green infrastructure size, quality and proximity’, *BMC Public Health*, 20(1), pp. 1–15. doi: 10.1186/s12889-020-08762-x.

Díaz, S. *et al.* (2018) ‘Assessing nature’s contributions to people’, *Science*, 359(6373), pp. 270–272. doi: 10.1126/science.aap8826.

Dzhambov, A. M. *et al.* (2020) ‘Analytical approaches to testing pathways linking greenspace to health: A scoping review of the empirical literature’, *Environmental Research*. Elsevier Inc., 186(March), p. 109613. doi: 10.1016/j.envres.2020.109613.

- Ekkel, E. D. and de Vries, S. (2017) 'Nearby green space and human health: Evaluating accessibility metrics', *Landscape and Urban Planning*. Elsevier B.V., 157, pp. 214–220. doi: 10.1016/j.landurbplan.2016.06.008.
- Engemann, K. *et al.* (2019) 'Residential green space in childhood is associated with lower risk of psychiatric disorders from adolescence into adulthood', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1807504116.
- Ensor, T. and Cooper, S. (2004) 'Overcoming barriers to health service access: Influencing the demand side', *Health Policy and Planning*, 19(2), pp. 69–79. doi: 10.1093/heapol/czh009.
- Environmental Agency (2016) *Lidar composite digital surface model England 2m resolution [ASC geospatial data], scale 1:8000, tile: SJ89*.
- Fong, K. C., Hart, J. E. and James, P. (2018) 'A Review of Epidemiologic Studies on Greenness and Health: Updated Literature Through 2017', *Current environmental health reports*, 5(1), pp. 77–87. doi: 10.1007/s40572-018-0179-y.
- Frumkin, H. *et al.* (2017) 'Nature contact and human health: A research agenda', *Environmental Health Perspectives*, 125(7), pp. 1–18. doi: 10.1289/EHP1663.
- GMCA (2019) *Greater Manchester's Plan For Homes, Jobs And The Environment- GREATER MANCHESTER SPATIAL FRAMEWORK*. Available at: https://www.greatermanchester-ca.gov.uk/media/1710/gm_plan_for_homes_jobs_and_the_environment_1101-web.pdf.
- Goodchild, M. F. (2011) 'Scale in GIS: An overview', *Geomorphology*. doi: 10.1016/j.geomorph.2010.10.004.
- Hartig, T. *et al.* (2014) 'Nature and Health', *Annual Review of Public Health*, 35(1), pp. 207–228. doi: 10.1146/annurev-publhealth-032013-182443.
- Helbich, M. (2018) 'Toward dynamic urban environmental exposure assessments in mental health research', *Environmental Research*. doi: 10.1016/j.envres.2017.11.006.
- Holland, I. *et al.* (2021) 'Measuring Nature Contact: A Narrative Review', *International Journal of*

- Environmental Research and Public Health*, 18(8), p. 4092. doi: 10.3390/ijerph18084092.
- James, P. *et al.* (2015) 'A Review of the Health Benefits of Greenness', *Current Epidemiology Reports*, 2(2), pp. 131–142. doi: 10.1007/s40471-015-0043-7.
- James, P. *et al.* (2016) 'Exposure to greenness and mortality in a nationwide prospective cohort study of women', *Environmental Health Perspectives*, 124(9), pp. 1344–1352. doi: 10.1289/ehp.1510363.
- Jarvis, I. *et al.* (2020) 'Greenspace access does not correspond to nature exposure: Measures of urban natural space with implications for health research', *Landscape and Urban Planning*. Elsevier B.V., 194, p. 103686. doi: 10.1016/j.landurbplan.2019.103686.
- Jeanjean, A. P. R., Monks, P. S. and Leigh, R. J. (2016) 'Modelling the effectiveness of urban trees and grass on PM_{2.5} reduction via dispersion and deposition at a city scale', *Atmospheric Environment*. doi: 10.1016/j.atmosenv.2016.09.033.
- Jia, P. (2019) 'Spatial lifecourse epidemiology', *The Lancet Planetary Health*. doi: 10.1016/S2542-5196(18)30245-6.
- Jimenez, M. P. *et al.* (2021) 'Associations between Nature Exposure and Health: A Review of the Evidence', *Int. J. Environ. Res. Public Health*, 18, p. 4790. doi: 10.3390/ijerph18094790.
- Jones, A., Hillsdon, M. and Coombes, E. (2009) 'Greenspace access, use, and physical activity: Understanding the effects of area deprivation', *Preventive Medicine*. doi: 10.1016/j.ypmed.2009.10.012.
- Jordan, H. (1994) 'Public Parks, 1885-1914', *Garden History*. doi: 10.2307/1587004.
- Kaplan, R. and Kaplan, S. (1989) 'The experience of nature: a psychological perspective', *The experience of nature: a psychological perspective*. doi: 10.1097/00005053-199111000-00012.
- Kaplan, S. (1995) 'The restorative benefits of nature: Toward an integrative framework', *Journal of Environmental Psychology*. doi: 10.1016/0272-4944(95)90001-2.
- Kelejian, H. H. and Prucha, I. R. (2010) 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics*. doi:

10.1016/j.jeconom.2009.10.025.

Khomenko, S. *et al.* (2021) 'Premature mortality due to air pollution in European cities: a health impact assessment', *The Lancet Planetary Health*, 5(3), pp. e121–e134. doi: 10.1016/S2542-5196(20)30272-2.

Klompaker, J. O. *et al.* (2018) 'Green space definition affects associations of green space with overweight and physical activity', *Environmental Research*. Academic Press Inc., 160, pp. 531–540. doi: 10.1016/j.envres.2017.10.027.

Kondo, M. C. *et al.* (2018) 'Urban green space and its impact on human health', *International Journal of Environmental Research and Public Health*, 15(3). doi: 10.3390/ijerph15030445.

Kwan, M. P. (2012) 'The Uncertain Geographic Context Problem', *Annals of the Association of American Geographers*. doi: 10.1080/00045608.2012.687349.

Labib, S. M. *et al.* (2020) 'Noncommunicable Diseases, Park Prescriptions, and Urban Green Space Use Patterns in a Global South Context: The Case of Dhaka, Bangladesh', *International journal of environmental research and public health*. NLM (Medline), 17(11). doi: 10.3390/ijerph17113900.

Labib, S. M. and Harris, A. (2018) 'The potentials of Sentinel-2 and Landsat-8 data in green infrastructure extraction, using object based image analysis (OBIA) method', *European Journal of Remote Sensing*. Taylor and Francis Ltd., 51(1), pp. 231–240. doi: 10.1080/22797254.2017.1419441.

Labib, S. M., Huck, J. J. and Lindley, S. (2021) 'Modelling and mapping eye-level greenness visibility exposure using multi-source data at high spatial resolutions', *Science of the Total Environment*. Elsevier BV, 755, p. 143050. doi: 10.1016/j.scitotenv.2020.143050.

Labib, S. M., Lindley, S. and Huck, J. J. (2020a) 'Scale effects in remotely sensed greenspace metrics and how to mitigate them for environmental health exposure assessment', *Computers, Environment and Urban Systems*. Elsevier, 82(May), p. 101501. doi: 10.1016/j.compenvurbsys.2020.101501.

Labib, S. M., Lindley, S. and Huck, J. J. (2020b) 'Spatial dimensions of the influence of urban green-blue spaces on human health: A systematic review', *Environmental Research*, 180, p. 108869. doi:

10.1016/j.envres.2019.108869.

Lindley, S. J. *et al.* (2019) 'Biodiversity, Physical Health and Climate Change: A Synthesis of Recent Evidence', in *Biodiversity and Health in the Face of Climate Change*, pp. 17–46. doi: 10.1007/978-3-030-02318-8_2.

Markevych, I. *et al.* (2017) 'Exploring pathways linking greenspace to health: Theoretical and methodological guidance', *Environmental Research*. Elsevier Inc., 158(June), pp. 301–317. doi: 10.1016/j.envres.2017.06.028.

Marmot, M. (2020) 'Health equity in England: The Marmot review 10 years on', *The BMJ*. doi: 10.1136/bmj.m693.

Marselle, M. R. *et al.* (2021) 'Pathways linking biodiversity to human health: A conceptual framework', *Environment International*. Elsevier Ltd, p. 106420. doi: 10.1016/j.envint.2021.106420.

Medina, S. *et al.* (2004) 'Aphis: Public health impact of PM10 in 19 European cities', *Journal of Epidemiology and Community Health*, 58(10), pp. 831–836. doi: 10.1136/jech.2003.016386.

Mitchell, R. and Popham, F. (2008) 'Effect of exposure to natural environment on health inequalities: an observational population study', *The Lancet*. Elsevier Ltd, 372(9650), pp. 1655–1660. doi: 10.1016/S0140-6736(08)61689-X.

Moran, P. A. (1950) 'Notes on continuous stochastic phenomena.', *Biometrika*. doi: 10.1093/biomet/37.1-2.17.

Nieuwenhuijsen, M. J. *et al.* (2017) 'Fifty shades of green', *Epidemiology*, 28(1), pp. 63–71. doi: 10.1097/EDE.0000000000000549.

Openshaw, S. (1984) 'Ecological fallacies and the analysis of areal census data (UK, Italy).', *Environment & Planning A*. doi: 10.1068/a160017.

Pauleit, S. *et al.* (2003) 'Promoting the natural greenstructure of towns and cities: English nature's accessible natural greenspace standards model', *Built Environment*. doi: 10.2148/benv.29.2.157.54469.

- Pearce, J. R. (2018) 'Complexity and Uncertainty in Geography of Health Research: Incorporating Life-Course Perspectives', *Annals of the American Association of Geographers*. Taylor and Francis Ltd., 108(6), pp. 1491–1498. doi: 10.1080/24694452.2017.1416280.
- Picavet, H. S. J. *et al.* (2016) 'Greener living environment healthier people? Exploring green space, physical activity and health in the Doetinchem Cohort Study', *Preventive Medicine*, 89, pp. 7–14. doi: 10.1016/j.ypmed.2016.04.021.
- R Team, C., others and R Core Team (2013) 'R: A Language and Environment for Statistical Computing'. Available at: <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing> (Accessed: 10 September 2020).
- Richardson, E. *et al.* (2010) 'The association between green space and cause-specific mortality in urban New Zealand: An ecological analysis of green space utility', *BMC Public Health*. doi: 10.1186/1471-2458-10-240.
- Richardson, E. A. and Mitchell, R. (2010) 'Gender differences in relationships between urban green space and health in the United Kingdom', *Social Science and Medicine*, 71(3), pp. 568–575. doi: 10.1016/j.socscimed.2010.04.015.
- Rigolon, A. *et al.* (2021) 'Green Space and Health Equity: A Systematic Review on the Potential of Green Space to Reduce Health Disparities', *J. Environ. Res. Public Health*, 18, p. 2563. doi: 10.3390/ijerph18052563.
- Rojas-Rueda, D. *et al.* (2019) 'Green spaces and mortality: a systematic review and meta-analysis of cohort studies', *The Lancet Planetary Health*. World Health Organization, 3(11), pp. e469–e477. doi: 10.1016/S2542-5196(19)30215-3.
- Rugel, E. J. *et al.* (2017) 'Beyond the Normalized Difference Vegetation Index (NDVI): Developing a Natural Space Index for population-level health research', *Environmental Research*. Academic Press Inc., 159, pp. 474–483. doi: 10.1016/j.envres.2017.08.033.
- Silva, R. A., Rogers, K. and Buckley, T. J. (2018) 'Advancing Environmental Epidemiology to Assess the Beneficial Influence of the Natural Environment on Human Health and Well-Being',

Environmental Science and Technology. American Chemical Society, 52(17), pp. 9545–9555. doi: 10.1021/acs.est.8b01781.

Stas, M. *et al.* (2021) ‘Exposure to green space and pollen allergy symptom severity: A case-crossover study in Belgium’, *Science of The Total Environment*, 781, p. 146682. doi: 10.1016/j.scitotenv.2021.146682.

Turner, M. C. *et al.* (2017) ‘Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations’, *Annual Review of Public Health*. doi: 10.1146/annurev-publhealth-082516-012802.

Ulrich, R. S. (1984) ‘View through a window may influence recovery from surgery’, *Science*. doi: 10.1126/science.6143402.

Vienneau, D. *et al.* (2017) ‘More than clean air and tranquillity: Residential green is independently associated with decreasing mortality’, *Environment International*. doi: 10.1016/j.envint.2017.08.012.

Vrijheid, M. *et al.* (2020) ‘Early-life environmental exposures and childhood obesity: An exposome-wide approach’, *Environmental Health Perspectives*. doi: 10.1289/EHP5975.

Wheeler, B. W. *et al.* (2015) ‘Beyond greenspace: An ecological study of population general health and indicators of natural environment type and quality’, *International Journal of Health Geographics*. doi: 10.1186/s12942-015-0009-5.

Wild, C. P. (2012) ‘The exposome: From concept to utility’, *International Journal of Epidemiology*. doi: 10.1093/ije/dyr236.

Wilker, E. H. *et al.* (2014) ‘Green space and mortality following ischemic stroke’, *Environmental Research*. doi: 10.1016/j.envres.2014.05.005.

Wu, L. and Kim, S. K. (2021) ‘Health outcomes of urban green space in China: Evidence from Beijing’, *Sustainable Cities and Society*, 65. doi: 10.1016/j.scs.2020.102604.

Zhang, J. *et al.* (2021) ‘Deciphering the changes in residential exposure to green spaces: The case of a rapidly urbanizing metropolitan region’, *Building and Environment*, 188, p. 107508. doi:

10.1016/j.buildenv.2020.107508.

Zhang, L. *et al.* (2018) 'Impacts of individual daily greenspace exposure on health based on individual activity space and structural equation modeling', *International Journal of Environmental Research and Public Health*. doi: 10.3390/ijerph15102323.

Zhang, Z. *et al.* (2017) 'Hierarchical cluster analysis in clinical research with heterogeneous study population: Highlighting its visualization with R', *Annals of Translational Medicine*. AME Publishing Company, 5(4), pp. 9–9. doi: 10.21037/atm.2017.02.05.

Zuniga-Teran, A. A. *et al.* (2019) 'Exploring the influence of neighborhood walkability on the frequency of use of greenspace', *Landscape and Urban Planning*. doi:
10.1016/j.landurbplan.2019.103609.

Supplementary Information

Cite As: Labib, S. M., Lindley, S., & Huck, J. J. (2021). Estimating multiple greenspace exposure types and their associations with neighborhood premature mortality: a socioecological study. *Science of the Total Environment*. DOI:

Contents

Supplementary Notes (1 to 4) ; Pages (1-13)

Figures S1 to S2 ; Pages (14-15)

Tables S1 to S4 ; Pages (16- 24)

SI References ; Page (25)

Note 1: Experiment with different weights for composite

In the main text, we estimated the composite greenspace exposure index (CGEI) as a mean of the GAVI, GACI, and VGVI. We took the average of the single exposure metric providing equal weights to each exposure type. As noted in the main text, there is no pre-existing evidence available that identified the relative weights of these exposure metrics in relation to premature mortality; therefore, we assumed each greenspace exposure might have an equal effect on premature mortality rate. However, it can be argued that certain exposure type may have relatively higher weights than others, depending on the health outcomes considered. For example, access to greenspace may be more important for physical activity-related outcome variables. Considering the potentials of different weights for different greenspace exposure types, we also developed a weighted approach for the CGEI estimation.

We developed the following equation to estimate the weighted CGEI value for varying weighting of the single exposure indices.

$$CGEI_{Wxj} = \frac{(GAVI_j * W_{xGAVI}) + (GACI_j * W_{xGACI}) + (VGVI_j * W_{xVGVI})}{W_{xGAVI} + W_{xGACI} + W_{xVGVI}}$$

Here, $CGEI_{Wxj}$ is the composite greenspace exposure index value for cell j for weight Wx ($x = 1$ to n); $GAVI_j$, $GACI_j$, and $VGVI_j$ represent the respective availability, accessibility, and visibility index values for cell j obtained from single exposure maps; W_{xGAVI} , W_{xGACI} , W_{xVGVI} , indicate weight value for the corresponding exposure layer.

To test the sensitivity of the weighted CGEI, we generated three hypothetical weighting scenarios for single exposure layers (Table NS1-1). The weighting system considers one metric to have twice more important than the other two metrics. We used these three weighting scenarios and produced

three weighted CGEI maps (Figure NS1-1). Finally, we compared the weighted CGEI maps with the non-weighted (average) CGEI map.

Table NS1-1: Hypothetical weights for different exposure layers

Weights _x	GAVI	GACI	VGVI
W1	0.5	0.25	0.25
W2	0.25	0.5	0.25
W3	0.25	0.25	0.5

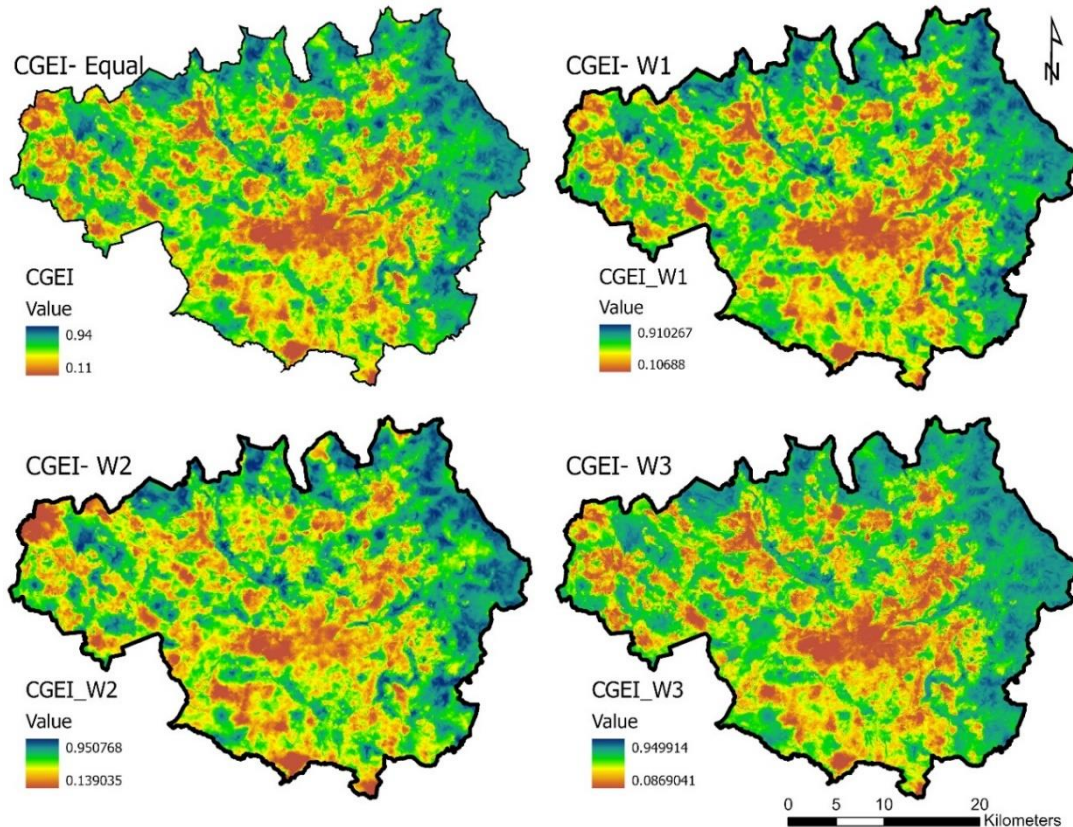


Figure NS1-1: CGEI maps with varying weighing for the input exposure metrics. CGEI-equal where the input exposure maps have equal weights; CGEI-W1, CGEI-W2, and CGEI-W3 reflect the different weights for GAVI, GACI and VGVI metrics.

As illustrated in Figure NS1-1, the weighted CGEI maps showed a similar spatial pattern in exposure to the equal-weighted CGEI map. However, the CGEI-W2 map has a slightly different spatial pattern in exposure gradient. It has more weights for the accessibility metric (i.e., GACI), which has a varying exposure pattern compared to the availability visibility metrics. As the maps look very similar, we further explored the correlations among these maps. The correlations result presented in table NS1-2. It is clear that the weighted maps have very high correlations (coefficients are near 0.9) with the equal-weighted map. Such relations are reasonable as they have similar spatial patterns. Furthermore, the weights are not considerably different from the equal weights (e.g., GEVI weight is 0.5 in CEGIW1, in CGEI-equal GAVI weight was 0.333).

Table NS1-2: Correlations among different CGEI maps with varying weights and YPLL

	CGEI (equal)	CGEIW1	CGEIW2	CGEIW3	YPLL
CGEI (equal)	1	0.98797	0.95209	0.98654	-0.3681
CGEIW1		1	0.91032	0.97622	-0.3893
CGEIW2			1	0.89899	-0.2676
CGEIW3				1	-0.3976

All correlations are significant at $p < 0.05$

We recognise that if specific input layers have very high weights (e.g., 0.9 for one layer and 0.5 for the other two layers), the weighted CGEI map may have a considerable correlation with the equal-weighted CGEI map. To further evaluate the sensitivity, additional weighting combinations need to be checked with other evaluation approaches such as simulation (e.g., Monte Carlo) and the analytical hierarchy process, which are beyond the scope of this current study.

Nonetheless, to explore the sensitivity of different weighted CGEI maps with our outcome variable (YPLL), we tested their correlations at a neighbourhood scale (Table NS1-2). The results showed that higher weights for availability and visibility in the CGEI have slightly stronger correlations with YPLL than the equal weight CGEI. In contrast, the higher weighted accessibility CGIE (W2) has relatively lower correlations with YPLL than the equal-weighted CGIE. These examples indicate that varying weights of the composite input layers might identify stronger or weaker correlations with the health outcome variable. Such differences in associations need to be considered in modelling the effects of CGEI on health outcomes; as for different health outcomes, certain weighting may be more relevant than equal-weighted CGEI.

Note 2: Non-spatial modelling and outputs

Non-spatial models do not consider the spatial patterns (e.g., clustering) and dependence observed among variables; therefore they do not account for spatial autocorrelation, lags, and covariance among, and between, dependent and independent variables in the models. However, if autocorrelation, and spatial lag are ignored in regression models those models become vulnerable to the violation of the assumption of independence of residuals and Type-I error estimation (Waller and Gotway, 2004). These are issues that can affect non-spatial statistical models such that, by consequence, they produce more apparently significant results than are justified by the data. Such issues can also inflate parameter estimates (Labib et al., 2020a; Waller and Gotway, 2004). Nonetheless, in existing studies of health and greenspace, non-spatial models are the most commonly applied to the data. Considering this, we applied the same variables and model structure used in adjusted spatial models to such non-spatial models.

Generalised linear models (GLM)

In the present study, we used non-spatial generalised linear models (GLM) using a gamma distribution and an identity link function to account for positive values of YPLL that could skew distributions. GLM is a robust modelling approach that allows variables to have non-linear distribution and it does not require assumptions usually associated with linear regression models to be fulfilled. The model results for greenspace exposure as related to YPLL are presented in Table NS2-1.

Table NS2-1: Greenspace exposure and YPLL associations using GLM

Exposure	Intercept/ variables	Model outcomes			R ²	AIC	Moran's I
		Beta	S.E.	p			
Availability	(Intercept)	39.17***	9.33	2.83e-05	0.612	12718	0.347***
	Availability	-21.14***	2.97	1.79e-12			
	Income Sco◆	49.35***	3.36	< 2e-16			
	Crime Scor	4.61***	0.508	< 2e-16			
	Barriers_H	0.66***	0.04	< 2e-16			
	SHDiv	-14.32***	2.55	4.24e-06			
	GP_Dist	-1.77***	0.47	0.000226			
	pm10_mean	3.31***	0.79	3.19e-05			
Accessibility	(Intercept)	1.19	8.72	0.89	0.59	12770	0.358***
	Accessibility	4.66	2.72	0.08			
	Income Sco◆	51.03***	3.41	< 2e-16			
	Crime Scor	4.73***	0.52	< 2e-16			
	Barriers_H	0.66***	0.04	< 2e-16			
	SHDiv	-11.42***	2.61	1.27e-05			
	GP_Dist	-2.46***	0.49	4.98e-07			
	pm10_mean	5.669***	0.74	6.60e-14			
Visibility	(Intercept)	33.59***	9.67	0.00052	0.605	12745	0.356***
	Visibility	-8.91***	1.72	2.48e-07			
	Income Sco◆	50.12***	3.38	< 2e-16			
	Crime Scor	4.58***	0.51	< 2e-16			
	Barriers_H	0.65***	0.04	< 2e-16			
	SHDiv	-11.85	2.57	4.33e-06			
	GP_Dist	-2.05	0.48	2.06e-05			
	pm10_mean	3.46	0.83	3.53e-05			
Composite	(Intercept)	35.68***	9.99	0.000367	0.605	13085	0.354***
	Composite	-15.97***	3.2	6.56e-07			
	Income Sco◆	49.71***	3.39	< 2e-16			
	Crime Scor	4.78***	0.51	< 2e-16			
	Barriers_H	0.66***	0.04	< 2e-16			
	SHDiv	-12.16***	2.56	2.38e-06			
	GP_Dist	-2.43***	0.46	2.36e-07			
	pm10_mean	3.78***	0.81	3.55e-06			

*** indicate statistical significance at $p < 0.05$. ◆ income deprivation score, higher values indicates lower income level, and positive association implies lower income areas have higher YPLL rate.

Table NS2-1 indicates the coefficient, standard error, and p-value for each exposure and YPLL model. Each of the exposure variables demonstrated a significant association with YPLL. Availability, visibility, and composite exposure indicated significant negative associations with YPLL, implying

higher values of greenspace exposure are associated with significantly lower values of YPLL. Among these three exposure types, composite greenspace exposure demonstrated a slightly higher coefficient value than any given individual exposure metric. However, accessibility exposure showed a significant positive association with YPLL, indicating that increasing accessibility is associated with higher YPLL values. Nonetheless, it should be noted that this trend is specific to areas such as the Greater Manchester area and similar post-industrial UK cities as discussed in Section 3.3, and Section 4 in the main manuscript.

The pseudo R^2 (i.e., Nagelkerke) and AIC values are also presented in Table NS2-1. All four models demonstrated a similar trend for pseudo R^2 , and AIC values, with Availability models showing slightly higher values for R^2 and lower values for AIC indicating that this model possessed slightly better explanatory power in relation to variations in the model. By contrast, the Accessibility model demonstrated lower values for R^2 and higher values for AIC, implying this model has lesser explanatory and predictive power. While the R^2 and AIC values address the validity of the models under discussion, it must be noted that the most important part of Table NS2-1 is the part presenting Moran's I values. We estimated the Moran's I value by applying a Monte Carlo simulation (a permutation bootstrap test) for the regression residuals of each model. The Moran's I value for each model was positive and statistically significant ($p < 0.001$). This suggests that the model residuals are clustered, and that there is high spatial autocorrelation among the variables used in the models (Waller and Gotway, 2004; Bivand et al., 2008).

Multiple Linear regression models

While the GLM models are more flexible with model assumptions and allowing varying distribution, these models are not useful in comparing with spatial model. Therefore to provide easier comparisons with spatial models, we also run multiple linear regression for the same model structure. The results of the multiple linear regression presented in Table NS2-2.

Table NS2-2: Greenspace exposure and YPLL associations using linear regression model

Greenspace exposure	Model Adjustment	Model outcomes			R^2	AIC	Moran's I
		Beta	S.E.	p			
Availability	Unadjusted	-67.02***	3.58	<2e-16	0.17	14165	0.52***
	Adjusted†	-23.00 ***	3.12	2.82e-13	0.598	12970	0.32***
Accessibility	Unadjusted	42.84***	4.302	<2e-16	0.056	14387	0.60***
	Adjusted	6.03	3.166	0.0569	0.586	13020	0.33***
Visibility	Unadjusted	-37.91***	1.97	<2e-16	0.18	14150	0.53***
	Adjusted	-9.697***	1.84	1.65e-07	0.59	12996	0.33***
Composite	Unadjusted	-67.42***	- 4.16	<2e-16	0.135	14240	0.56***
	Adjusted	-18.06 ***	- 3.44	1.84e-07	0.59	12997	0.32***

*** indicate statistical significance at $p < 0.05$. R^2 is Nagelkerke pseudo-R-square. †model adjusted for: Income score, Crime score, Barriers to housing score, Shannon diversity score, Distance to nearest general medical practice, and yearly average air pollution (PM_{10}).

The linear regression models also indicate similar associations as the GLM models, however the effect sizes are slightly larger for all the exposure variables. The AIC values are relatively higher compared to the GLM models, and the presence of spatial autocorrelations in the residuals are confirmed for these linear regression models. These results indicate while the linear regression models show similar pattern in associations between greenspace exposure and YPLL, there estimates are inflated due to both violation of model assumptions and spatial autocorrelations.

As the presence of spatial autocorrelation in the models has been validated, it can be argued that the model estimations cannot be considered accurate, and that the models might inflate the parameter estimates (Waller and Gotway, 2004; Bivand et al., 2008; Griffith, 2005). It is noteworthy that for the same exposure variables when comparing spatial regression model parameter estimates (Note 2) to non-spatial model parameter estimates that the latter result in considerably higher values. This further indicates that the use of non-spatial models with spatially-explicit exposure and health data is likely to result in estimation of greater effect sizes, which may be attributed to spatially autocorrelated observations and spatial dependence (Veloz, 2009; Waller and Gotway, 2004). Overall experiments with non-spatial models have indicated that greenspace exposure is significantly associated with health outcomes (i.e., YPLL). However, the model residuals are spatially autocorrelated implying the parameter estimates may not be accurate and, hence, the statistical significance may also be vulnerable to Type-I errors. Considering the previous, we adopted spatial regression as the primary modelling approach used in this study.

Note 3: Spatial modelling and outputs

Spatial Regression modelling

Non-spatial modelling confirmed the presence of significant spatial autocorrelation in the regression residuals (Note-2). Therefore, the values for the coefficients and p-values of the non-spatial models may not provide adequate confidence values such that inferential reasoning in regard to their relationships can be supported (Bidvan et al., 2008; Waller and Gotway, 2004). In this regard, spatial regression models such as spatial autoregressive models or geographically weighted regressions (GWR) are the alternative solutions. For the purposes of the present study we selected spatial autoregressive models to test the relationships between greenspace exposure and YPLL, given their robustness in explaining the overall pattern of relationships between the dependent and independent variables that account for spatial dependence, and autocorrelation. While the GWR provides estimations of coefficients using local regression, GWR models are unable to make inferences about overall relationships between dependent and independent variables (Fotheringham et al., 2003). As a result, GWR was not used in the present study.

Before running the spatial models, we applied the lagrange multiplier (LM) test to investigate for spatial dependence and spatial heterogeneity among the model variables and errors (details in Anselin, 1988a). The LM test was used to verify whether conducting spatial modelling could improve model reliability and resolve issues with spatial autocorrelation, and dependence (Anselin, 1988a,b). In our case, we ran the LM test for each non-spatial model (Note-S1). The LM test results indicated that both the spatial error model, and the spatial auto-regressive moving-average (SARMA) model could significantly improve reliability of the modelled relationships (Table NS3-1). Therefore, use of spatial models was indicated versus use of non-spatial models.

Table NS3-1: Lagrange multiplier (LM) test for the non-spatial models

Spatial dependence	Availability model		Accessibility model		Visibility model		Composite model	
	Stat	p	Stat	p	Stat	p	Stat	p
LMerr	1.01E+03***	<2e-16	1.1295e+03***	<2e-16	1.0567e+03***	<2e-16	1.0371e+03***	<2e-16
LMlag	-5.13E-04	1	-7.0171e-04	1	-5.1775e-04	1	-5.5281e-04	1
SARMA	1.01E+03***	<2e-16	1.1295e+03***	<2e-16	1.0567e+03***	<2e-16	1.0371e+03***	<2e-16

***indicate statistical significance at $p < 0.05$.

As illustrated in Table SN3-1, the LMerr, and SARMA dependence tests indicated the presence of statistical significance implying that the spatial error and moving average models could account for spatial dependence in the relationship between greenspace exposure and YPLL. In this case, the values for the lag tests did not reach statistical significance, implying the independent spatial lag variable may not have significant impact on the models. However, SARAR tests did provide statistically significant values, which indicated that the dependent variable of the model was affected by spatial lag, and that the error terms of the models demonstrated spatial autocorrelation (Anselin, 2013). The previous was also true in our case, in which the YPLL (dependent variable) showed spatial lag, and the error terms of the models were spatially autocorrelated. As a result, a combined nested model would be needed to account for both a spatially lagged dependent variable and a moving average for error terms. On the basis of these observations, in the present study, we used a spatial autoregressive model with autoregressive disturbances (SARAR), also known as the “Spatial simultaneous autoregressive” (SAC) model (Kelejian and Prucha, 2010).

As the name suggests the SAC/SARAR models combine two types of spatial models, one to deal with spatial dependence in the lagged variables in the model and another model to account for spatial autocorrelation in the error distribution. Simply put, SAC models allow higher-order spatial dependence in the dependent variable, independent variables, and spatial errors (Kelejian and Prucha, 2010; LeSage, 2008). For n spatial units, the model can be expressed with the following equations (details of the derivation in : Kelejian and Prucha, 2010; 1998);

$$y_n = X_n\beta_n + \lambda_n W_n y_n + u_n \text{ (eq1)}$$

$$u_n = \rho_n M_n u_n + \varepsilon_n \text{ (eq2)}$$

Here, y_n denotes the dependent variable (in our case YPLL), X_n denotes the independent variables (greenspace exposure and, W_n , and M_n are the spatial weight matrix, u_n denotes the regression disturbances, and ε_n indicates the non-spatial component of the errors. The λ_n (lag simultaneous autoregressive lag coefficient), and ρ_n (error simultaneous autoregressive error coefficient) are the spatial autoregressive parameters. Lambda accounts for spatial lag in the dependent variable and rho accounts for spatial autocorrelations among the regression errors. The model allows the spatial matrixes to be the same with $W_n = M_n$, which will frequently be the case when the model is in use (Kelejian and Prucha, 2010). The spatial weight matrix represents the spatial structure of the data and defines the spatial relations among locations and, therefore, determines the spatial autocorrelation statistics (Getis, 2009; Zhou and Lin, 2007).

There are several approaches that can be applied in developing a spatial weight matrix, such as use of a distance based matrix, or weights based on boundaries between spatial units (Tsai et al., 2009). Commonly, however, spatial regression models are used to compute spatial weights following boundary-based approaches such as contiguity-based weight metrics (e.g., queen criterion of contiguity) due to their robust and objective estimations of spatial weights among spatial units that share a common border of non-zero length (Tsai et al., 2009; Bidvan et al., 2019). In our case, we selected a first order queen contiguity-based weight metric for the estimation of spatial weights for the regression models. Queen contiguity combines the spatial contiguity of both the Rook and Bishop relationships into a single measure which provides a better characterisation of spatial weights among features, especially when health-related data are used in spatial models (Tsai et al., 2009; Li et al., 2017). Based on the model structure described in the main manuscript, the results of both unadjusted and adjusted spatial regressions of exposure variables and YPLL are presented in Table NS3-2.

Table NS3-2: Greenspace exposure and YPLL relationships using spatial regression models. The ‘Estimate’ values are the numbers of years of potential life lost associated with each variable.

Model	Adj	Intercept/ variables	Model outcomes			R ²	AIC	LR	Moran's I
			Estimate	95% CI (low ; up)	p				
Availability	U	(Intercept)	11.81***	8.85 ; 14.78	6.00E-15	0.577	13047	1123 ***	-
		Availability	-13.03***	-16.59 ; -9.48	6.661E-13				
		Rho	0.91***	0.88 ; 0.93	< 2.22e-16				
		Lambda	-0.62***	-0.73 ; -0.50	< 2.22e-16				
Availability	A	(Intercept)	102.45***	38.02; 166.87	0.001827	0.702	12475	500. 21** *	-
		Availability	-9.802***	-16.56; -3.03	0.0045040				
		Income Sco♦	58.411***	52.44; 64.38	< 2.2e-16				
		Crime Scor	1.182***	0.186; 2.17	0.0199093				
		Barriers_H	0.189***	0.099; 0.279	3.64e-05				
		SHDiv	-4.204	-8.783; 0.37	0.0718512				
		GP_Dist	-2.238***	-3.494; -0.98	0.0004754				

		pm10_mean	0.544	-5.24; 6.32	0.8537027				
		Rho	-0.502***	-0.62 ; -0.422	< 2.22e-16				
		Lambda	0.913***	0.90 ; 0.94	< 2.22e-16				
Accessibility	U	(Intercept)	0.02	-3.08 ; 3.13	0.98727	0.565	13094	1297	-
		Accessibility	5.95***	2.47 ; 9.44	0.00080			***	0.007
		Rho	0.93***	0.91 ; 0.95	< 2.22e-16				
		Lambda	-0.63***	-0.74 ; -0.52	< 2.22e-16				
	A	(Intercept)	90.51***	25.97; 155.06	0.0059852	0.70	12481	543*	-0.03
		Accessibility	4.017	-3.94; 11.98	0.3228795			**	
		Income Sco♦	59.571***	53.65; 65.49	< 2.2e-16				
		Crime Scor	0.969	-0.02; 1.96	0.0553086				
		Barriers_H	0.171	0.082; 0.26	0.0001672				
		SHDiv	-4.168	-8.75; 0.415	0.0746727				
	GP_Dist	-2.468***	-3.71; -1.21	0.0001104					
	pm10_mean	1.051	-4.74; 6.85	0.7223157					
	Rho	-0.501***	-0.60; -0.39	< 2.22e-16					
	Lambda	0.916***	0.98; 0.94	< 2.22e-16					
Visibility	U	(Intercept)	9.75***	7.13 ; 12.38	3.34E-13	0.576	13051	1103	-
		Visibility	-6.80***	-8.8 ; -4.80	2.64E-11			***	0.005
		Rho	0.91***	0.88 ; 0.93	< 2.22e-16				
		Lambda	-0.61***	-0.72 ; -0.50	< 2.22e-16				
	A	(Intercept)	104.82***	40.08; 169.55	0.0015047	0.701	12475	525	-
		Visibility	-6.06***	-10.36; -1.75	0.0057690			***	0.031
		Income Sco♦	58.57***	52.61; 64.53	< 2.2e-16				
		Crime Scor	1.101***	0.11; 2.09	0.0292400				
		Barriers_H	0.184***	0.095; 0.274	5.284e-05				
		SHDiv	-3.97	-8.56; 0.601	0.0886282				
	GP_Dist	-2.23***	-3.487; -0.97	0.0005111					
	pm10_mean	0.238	-5.57; 6.054	0.9360075					
	Rho	-0.504***	-0.60; -0.40	< 2.22e-16					
	Lambda	0.914***	0.887; 0.94	< 2.22e-16					
Composite	U	(Intercept)	13.07***	9.56 ; 16.58	2.713E-13	0.573	13061	1183	-
		Composite	-12.46***	-16.35 ; -8.56	3.621E-10			***	0.006
		Rho	0.918***	0.89 ; 0.94	< 2.22e-16				
		Lambda	-0.624***	-0.73 ; -0.51	< 2.22e-17				
	A	(Intercept)	105.36***	40.32; 170.39	0.0014972	0.702	12477	524*	-
		Composite	-9.947***	-17.8; -2.001	0.0141391			**	0.033
		Income Sco♦	58.646***	52.68; 64.60	< 2.2e-16				
		Crime Scor	1.142***	0.148; 2.13	0.0242880				
		Barriers_H	0.181***	0.092; 0.271	7.093e-05				
		SHDiv	-4.13	0.092; 0.271	0.0770259				
	GP_Dist	-2.35***	-3.60; -1.10	0.0002116					
	pm10_mean	0.478	-5.32; 6.28	0.8716693					
	Rho	-0.503***	-0.60; -0.40	< 2.22e-16					
	Lambda	0.914***	0.88; 0.94	< 2.22e-16					

U = Unadjusted, A = Adjusted model. CI: Confidence interval. *** indicate statistical significance at $p < 0.05$. Dependent variable is YPLL. ♦ income deprivation score, higher values indicate lower income levels, and positive association implies lower income areas have higher YPLL rates.

For the availability, visibility, and composite greenspace exposure values there are significant negative associations between exposure and YPLL in both the unadjusted and adjusted models. The 'estimate' values indicate the numbers of years of YPLL changed with one unit increase in greenspace exposure value (from 0 to 1). However, the effect sizes are attenuated in the adjusted models compared

to the unadjusted models. Note, that the accessibility models indicate that there is a positive association between greenspace accessibility and YPLL. However, in the adjusted model the relationship between greenspace accessibility and YPLL becomes insignificant. The composite greenspace exposure variable demonstrated a slightly greater effect size than the other exposure variables. This indicates that the composite metric demonstrates stronger associations with YPLL, even after controlling for confounders.

The spatial regression models also produced estimates for rho, lambda, and likelihood ratio (LR) test results. As presented in Table NS3-2, the rho, and lambda parameters in all the models are statistically significant. This implies that the simultaneous autoregressive parameters accounting for spatial lag and errors have influence on the modelled variables and results. Adding rho and lambda as parameters has allowed us to adjust spatial lag and autocorrelation and thus provided a robust estimation of the observed associations. This also implied that, compared to the non-spatial models, the results of the spatial models provide greater confidence for making statistical inferences about the modelled relationships. In this regard, the LR test results for all the models are also significant indicating the SAC models have better goodness of fit compared to non-spatial linear models.

Additionally, Table NS3-2 provides the Moran's I value for each model and the Moran's I statistical values are near zero and insignificant in all the models. Insignificant Moran's I values imply that there are no significant spatial autocorrelations in the model residuals. In short, these results indicated spatial lag, and errors are critical components in modelling the relationships between greenspace exposure and YPLL. Spatial models account for both spatial lag and autocorrelation and therefore provide a better estimation of the associations between greenspace exposure and YPLL.

Table NS3-2 indicates the R^2 values for different exposure models are nearly same, in particular the availability and composite model have exact same R^2 value. This is crucial as they imply each of the exposure variable may have similar level of expandability in determining the variance in YPLL for our observed population.

The comparisons of estimated coefficients of spatial (Table NS3-2) and non-spatial models indicate that the parameters estimates provided by the spatial models are considerably smaller than the parameter estimates provided by the non-spatial models (Table NS2-1, NS2-2). This observation provides an interesting insight, the larger parameter estimates in the non-spatial models may have been attributed to the autocorrelated errors and spatial lags, which are not considered by non-spatial models (Veloz, 2009; Waller and Gotway, 2004). Therefore, in non-spatial models the parameter estimates might have been inflated by these correlated observations (Veloz, 2009), which in turn exhibited apparently larger effects for greenspace exposure on YPLL, than actually exist. These are critical observations because many existing studies do not consider spatial lag and autocorrelations when modelling the relationship between greenspace exposure and health outcomes, and, as a result, these

models might have inflated the effect size of the estimates. Therefore, the inconsistent results provided by previous studies that used non-spatial models might be due to varying spatial lag and autocorrelation values existing in their models. Future studies should consider these effects carefully to provide, better, more accurate, and robust estimations of the associations between greenspace and health.

Note 4: Dynamic exposure profiling for hypothetical household members

Greenspace exposure can vary for each individual depending on their spatial and temporal variations in daily activities (Helbich, 2018; Zhang et al., 2018). Existing studies usually ignore such spatial and temporal variations in the exposure of individuals and measure exposure only in the context of a fixed geographic location (Helbich, 2018). However, such an approach to exposure estimation is insufficient to provide inferences as to level of individual exposure because they do not consider the temporal dimensions and movement of people engaged in differing activities. Both of these dimensions can have an impact on individual levels of exposure. In the present study we developed greenspace an exposure index map, that can be used to measure individual greenspace exposures while taking account both of spatiotemporal variations and individual’s daily movements. Using GPS tracks, geocoded addresses, and duration of exposure at different activity locations or along paths of movement, individual daily greenspace exposure can be measured with greater accuracy than heretofore. In this study we provided an example of such exposure estimation using a hypothetical household with three members (i.e., child, mother, and father). Each individual in the household had both different activity locations and different movement patterns for each typical workday in the week. The method applied to individual daily exposure estimation followed the steps listed below:

Step-1: Geocode the address of their activities. For simplification of this hypothetical example, we listed, home, office, school, and park. Other places such as shops, gym, restaurants can also be considered.

Step-2: Record their GPS tracks of paths travelled from place to place. We used hypothetical GPS track records created by the lead author.

Step-3: From GPS tracks or travel diary entries estimate the amount of time each individual spent at each location as well as travelling to and from different locations. Time of travel depended on the mode used to travel. In this regard we simplified matters by using standard travel speeds for different modes of travel (Table NS4-1).

Table NS4-1: Assumed travel speed for different modes used by the individuals in this example.

Mode	Speed	<i>F</i> (hr / km)
Cycle	16 km hr	0.0625
Walk	3 km per hr	0.333
Car	48 km per hr	0.02

Step-4: Extract greenspace exposure value for each activity location and mean exposure value along the paths of travel. We used 5 m intervals along the GPS tracks to extract composite greenspace exposure values, and then took an average of all exposure values to estimate mean greenspace exposure along the each path of travel.

Step-5: Estimate the time-weighted greenspace exposure value for different locations, and travel paths. We used the following equation to estimate the 24 hr greenspace exposure for each individual.

$$IDGE_x = \left(CGEI_{L1} \times \frac{t_{L1}}{24} + CGEI_{L2} \times \frac{t_{L2}}{24} + \dots CGEI_{Ln} \times \frac{t_{Ln}}{24} \right) + \left(CGEI_{R1} \times \frac{t_{R1}}{24} + CGEI_{R2} \times \frac{t_{R2}}{24} \dots CGEI_{Rn} \times \frac{t_{Rn}}{24} \right) \text{ (eq. 1)}$$

Where, $IDGE_x$ is Individual daily greenspace exposure for individual x . $CGEI$ is composite greenspace exposure value for different locations (L_1, L_2, \dots, L_n) or along the route (R_1, R_2, \dots, R_n). Time spent at different locations expressed as t_L , and travel time along different routes expressed as t_R . The sum of all t_L and t_R resulted in the 24 hour value.

Equation 1 was applied to the activities of each member of the household to estimate individual daily greenspace exposure as presented in table NS4-2. Table NS4-2 also presented the data for each of the different activities, locations, and routes that applied to the individuals in the household. Additionally, temporal profiles were provided that indicated the amount of time each individual spent at different locations or along travel routes.

The results indicated that the child spent more time at home than the parents and that home had a higher CGEI value than most of the other locations (e.g., school) that were considered. Additionally the child also visited the park. The child used cycling and walking as modes of travel, which resulted in longer durations spent on travel compared to motorised modes (e.g., car) and the paths of travel used by the child had higher CGEI values compared to paths of travel applicable to the parents. In aggregate the child's daily activities resulted in an IDGE value of 0.557, a higher value than for either parent. The mother spent considerable time at the office (with a low CGEI), the routes of her travels also had lower exposure values and included the use of a car. Thus, the mother had a lower IDGE (0.501) than her child (Table NS4-2). Finally, the father had an even lower IDGE (0.489) than the mother, due to the low CGEI values obtained at the father's office and the combination of modes of travel and long travel routes the father experienced.

Table NS4-2: Exposure profiling for each individual in a household

Person	Exposure locations/tracks	Mode	Distance (km)	Time (hr)	CGEI	Time weighted CGEI
	I	II	III	IV = (III * F)	V	VI = (IV * V)/ 24
Child	Route1	Cycle	5.580	0.348	0.446	0.006

	Route2	Cycle	6.260	0.391	0.480	0.008
	Route3	Walk	2.440	0.813	0.584	0.020
	Home	NA	NA	16.448	0.600	0.411
	School	NA	NA	5.000	0.420	0.088
	Park	NA	NA	1.000	0.580	0.024
	<i>Sum of 24 hr exposure (IDGEchild)</i>					0.557
Mom	Route1	Car	5.400	0.112	0.471	0.002
	Route2	Car	7.330	0.152	0.466	0.003
	Home	NA	NA	14.736	0.600	0.368
	Office-1	NA	NA	9.000	0.340	0.128
	<i>Sum of 24 hr exposure (IDGEmother)</i>					0.501
Dad	Route1	Car	5.890	0.123	0.404	0.002
	Route2	Car	5.760	0.120	0.397	0.002
	Home	NA	NA	14.757	0.600	0.369
	Office-2	NA	NA	9.000	0.310	0.116
	<i>Sum of 24 hr exposure (IDGEfather)</i>					0.489

F = Travel time per kilometre depending on mode of travel (listed in Table NS4-1)

This simple but typical hypothetical example of the daily lives of British urban families demonstrates that individual daily greenspace exposure varies depending on individual variations in spatial and temporal experience. This example also illustrates why home-only or single location-based exposure estimations are vulnerable to misrepresentation of the daily exposure each individual experiences dependent on their activity and travel patterns. Considering the example provided above, we argue that our greenspace exposure mapping technique has the potential to improve individual daily greenspace exposure estimations by integrating additional data sets such as GPS tracks. Furthermore, such detailed exposure assessments would allow the creation of complete greenspace exposure profiles for any given individual and that would provide the opportunity to estimate greenspace exposure over long periods of time, and at multiple locations. Such processes could be used to support the wider framework of the exposome by allowing better characterisations of how greenspace exposure is associated to human health on an ongoing basis.

SI Figures

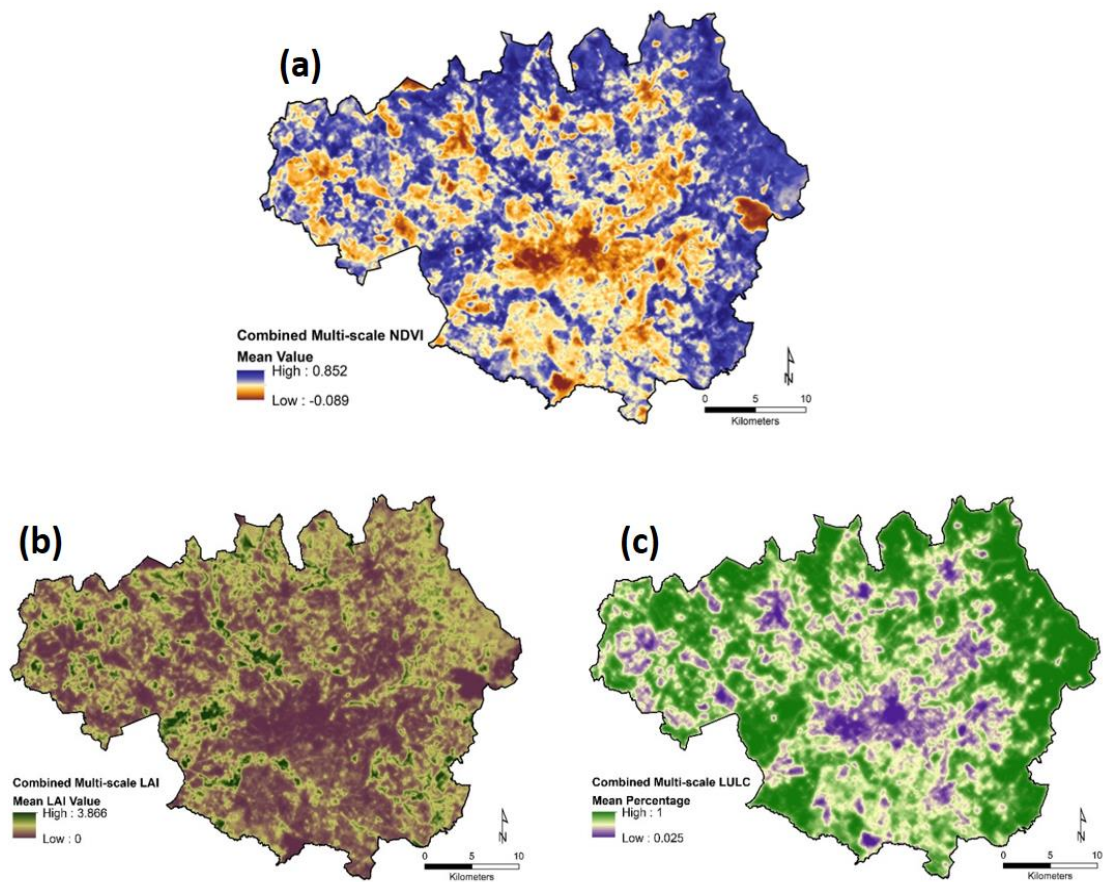


Fig. S1. Combined multiscale NDVI (a), LAI (b), and LULC (c) exposure values.

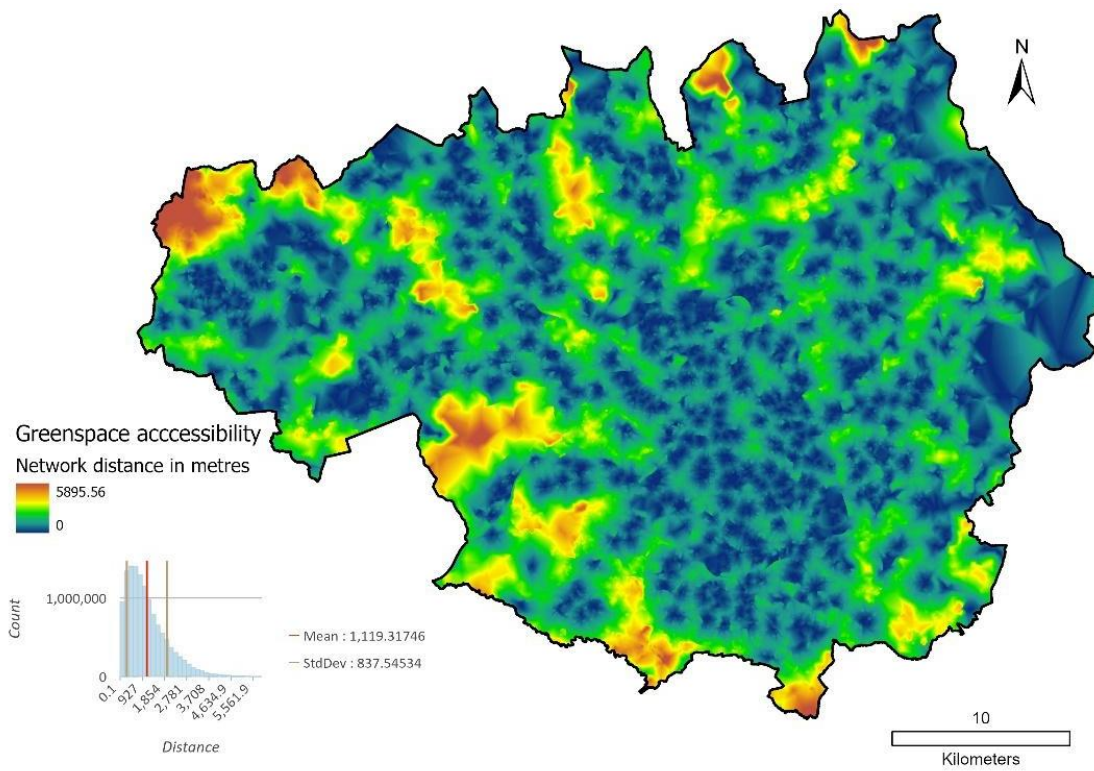


Fig. S2. Network distance to greenspace access points from each raster cell of the map. Maximum distance: 5895.6 m; Minimum distance: 0 m; Mean: 1,119.3 m; Standard Deviation: 837.5 m.

SI Tables

Table S1. Greenspaces considered in accessibility exposure measurements and rationale for selecting these greenspaces.

Name	Description	Rational
Public parks and open gardens	Areas of land designed, constructed, managed and maintained as a public park or garden. These usually have a defined perimeter and generally sit within or close to urban areas.	Free public access and usually the most common open areas in urban areas. Also commonly used in other studies as urban greenspace for accessibility assessment.
National parks	Areas designated as 'National Park' by the National Parks and Access to the Countryside Act 1949. Usually natural, semi-natural areas of conservation due to importance to the national heritage and maintenance of nature reserves.	National Parks and Access to the Countryside Act 1949 ensure the public has access for open-air recreation to open country. A major national park located at the edge of the study area.
Country parks	Areas designated public green spaces often at the edge of urban areas for public visits in a countryside environment.	Large countryside greenspaces with free public access. Several country parks located in the case study area.
Local nature reserves	Natural reserves in the UK, considered as nature reserves and conservation areas.	National Parks and Access to the Countryside Act 1949 ensure the public has access to these greenspaces for recreational purposes.
Countryside and Rights of Way Act Natural areas	Uncultivated areas such as moor, mountain, heath and down, as well as common land.	Large among of such areas are located at the edge of the study. Countryside and Rights of Way Act 2000, enables public to roam these area freely.

Table S2. IMD quintiles

Quintile group	IMD score range
1	≤ 8.49 (Least deprived)
2	8.5 - 13.79
3	13.8 - 21.35
4	21.36 - 34.17
5	≥ 34.18 (Most deprived)

Sources: <https://tools.npeu.ox.ac.uk/imd/> and

<http://imd-by-postcode.opendatacommunities.org/imd/2015>

Table S3. Summary statistics for different greenspace exposure in relation to IDM deciles at neighbourhood scale

IMD Decile*	Availability		Accessibility		Visibility		Composite	
	mean	std	mean	std	mean	std	mean	std
1	0.336	0.107	0.867	0.078	0.347	0.188	0.517	0.093
2	0.344	0.104	0.858	0.089	0.358	0.182	0.520	0.089
3	0.356	0.116	0.857	0.086	0.386	0.211	0.533	0.106
4	0.371	0.111	0.859	0.086	0.401	0.192	0.543	0.096
5	0.382	0.115	0.846	0.087	0.438	0.199	0.555	0.101
6	0.410	0.106	0.850	0.099	0.483	0.197	0.581	0.098
7	0.415	0.093	0.836	0.090	0.505	0.175	0.585	0.082
8	0.454	0.098	0.808	0.131	0.563	0.172	0.609	0.088
9	0.452	0.087	0.792	0.134	0.562	0.176	0.602	0.092
10	0.464	0.081	0.789	0.125	0.592	0.150	0.615	0.083

Table S4. ANOVA and Post-Hoc analyses for IMD between group comparisons. Decile 1 indicates most deprivation, and 10 least deprivation.

<i>ANOVA</i>							
Exposure		Df	Sum Sq	Mean Sq	F value	p	sig
Availability	IMD Decile	1	3.429	3.429	314.7	< 2e-16	***
	Residuals	1671	18.209	0.011			
Accessibility	IMD Decile	1	1	0.9997	103.5	< 2e-16	***
	Residuals	1671	16.14	0.0097			
Visibility	IMD Decile	1	12.26	12.255	351.3	< 2e-16	***
	Residuals	1671	58.29	0.035			
Composite	IMD Decile	1	2.101	2.1012	239.8	< 2e-16	***
	Residuals	1671	14.64	0.0088			
<i>Tukey multiple comparisons (Post-Hoc)</i>							
	IMD Deciles	diff	lower	upper	p-adj	sig	
Availability	2-1	0.008619	-0.01938	0.036617	0.9935918	null	
	3-1	0.019934	-0.00995	0.049821	0.5183904	null	
	4-1	0.034847	0.003958	0.065736	0.0132842	*	
	5-1	0.046053	0.011763	0.080343	0.0009197	***	
	6-1	0.07396	0.03967	0.10825	0	***	
	7-1	0.079409	0.042025	0.116792	0	***	
	8-1	0.118408	0.085634	0.151181	0	***	
	9-1	0.116334	0.08243	0.150238	0	***	
	10-1	0.128305	0.091064	0.165546	0	***	
	3-2	0.011315	-0.02106	0.043694	0.9840681	null	
	4-2	0.026228	-0.00708	0.059534	0.271857	null	
	5-2	0.037434	0.000952	0.073916	0.0388404	*	
	6-2	0.065341	0.028859	0.101823	0.0000007	***	
	7-2	0.07079	0.031386	0.110194	0.0000007	***	
	8-2	0.109789	0.074728	0.144849	0	***	

9-2	0.107715	0.071595	0.143835	0	***
10-2	0.119686	0.080417	0.158955	0	***
4-3	0.014913	-0.02	0.049821	0.9407698	null
5-3	0.026119	-0.01183	0.064069	0.4708217	null
6-3	0.054026	0.016075	0.091976	0.0002963	***
7-3	0.059475	0.018707	0.100242	0.0001766	***
8-3	0.098473	0.061887	0.13506	0	***
9-3	0.0964	0.058797	0.134002	0	***
10-3	0.108371	0.067734	0.149008	0	***
5-4	0.011206	-0.02754	0.049951	0.9959796	null
6-4	0.039113	0.000368	0.077858	0.0456584	*
7-4	0.044562	0.003054	0.08607	0.0239725	*
8-4	0.083561	0.046151	0.12097	0	***
9-4	0.081487	0.043083	0.119891	0	***
10-4	0.093458	0.052078	0.134838	0	***
6-5	0.027907	-0.0136	0.069414	0.5063395	null
7-5	0.033356	-0.01074	0.077453	0.3285345	null
8-5	0.072355	0.032091	0.112618	0.0000007	***
9-5	0.070281	0.029092	0.11147	0.0000033	***
10-5	0.082252	0.038275	0.126229	0.0000002	***
7-6	0.005449	-0.03865	0.049546	0.9999965	null
8-6	0.044448	0.004184	0.084711	0.0173417	*
9-6	0.042374	0.001185	0.083563	0.0378296	*
10-6	0.054345	0.010368	0.098322	0.0037274	**
8-7	0.038999	-0.00393	0.081928	0.1125507	null
9-7	0.036925	-0.00687	0.080723	0.1864315	null
10-7	0.048896	0.002466	0.095326	0.0296168	*
9-8	-0.00207	-0.04201	0.037862	1	null
10-8	0.009897	-0.03291	0.052702	0.9993032	null

	10-9	0.011971	-0.03171	0.055648	0.9973331	null
Accessibility	2-1	-0.00928	-0.03558	0.017014	0.9829145	null
	3-1	-0.01032	-0.03839	0.017755	0.9775089	null
	4-1	-0.00757	-0.03658	0.021443	0.9981801	null
	5-1	-0.0214	-0.0536	0.010811	0.5244355	null
	6-1	-0.01676	-0.04896	0.01545	0.8239775	null
	7-1	-0.03133	-0.06644	0.003787	0.1284158	null
	8-1	-0.05864	-0.08943	-0.02786	0.0000001	***
	9-1	-0.07453	-0.10637	-0.04268	0	***
	10-1	-0.07807	-0.11305	-0.04309	0	***
	3-2	-0.00103	-0.03144	0.029379	1	null
	4-2	0.001713	-0.02957	0.032996	1	null
	5-2	-0.01211	-0.04638	0.022154	0.9827569	null
	6-2	-0.00747	-0.04174	0.026792	0.9995661	null
	7-2	-0.02204	-0.05905	0.014968	0.6781815	null
	8-2	-0.04936	-0.08229	-0.01643	0.0000966	***
	9-2	-0.06524	-0.09917	-0.03132	0.0000001	***
	10-2	-0.06878	-0.10567	-0.0319	0.0000002	***
	4-3	0.002746	-0.03004	0.035534	0.9999999	null
	5-3	-0.01108	-0.04673	0.024565	0.9931163	null
	6-3	-0.00644	-0.04209	0.029204	0.9999089	null
	7-3	-0.02101	-0.0593	0.017281	0.7734333	null
	8-3	-0.04833	-0.08269	-0.01396	0.0003797	***
	9-3	-0.06421	-0.09953	-0.02889	0.0000004	***
	10-3	-0.06775	-0.10592	-0.02958	0.000001	***
	5-4	-0.01383	-0.05022	0.022566	0.971899	null
	6-4	-0.00919	-0.04558	0.027205	0.9985956	null
	7-4	-0.02376	-0.06274	0.015231	0.6481981	null
	8-4	-0.05107	-0.08621	-0.01594	0.0001911	***

	9-4	-0.06696	-0.10303	-0.03089	0.0000002	***
	10-4	-0.0705	-0.10936	-0.03163	0.0000005	***
	6-5	0.004639	-0.03435	0.043624	0.9999975	null
	7-5	-0.00993	-0.05135	0.031489	0.9990681	null
	8-5	-0.03725	-0.07507	0.00057	0.0576052	null
	9-5	-0.05313	-0.09182	-0.01444	0.0006054	***
	10-5	-0.05667	-0.09798	-0.01537	0.0006169	***
	7-6	-0.01457	-0.05599	0.02685	0.9833357	null
	8-6	-0.04189	-0.0797	-0.00407	0.0166625	*
	9-6	-0.05777	-0.09646	-0.01908	0.0001053	***
	10-6	-0.06131	-0.10262	-0.02	0.0001206	***
	8-7	-0.02732	-0.06764	0.013003	0.4947413	null
	9-7	-0.0432	-0.08434	-0.00206	0.0305251	*
	10-7	-0.04674	-0.09035	-0.00313	0.0244134	*
	9-8	-0.01588	-0.05339	0.021626	0.9438687	null
	10-8	-0.01942	-0.05963	0.020781	0.879949	null
	10-9	-0.00354	-0.04456	0.037483	0.9999999	null
Visibility	2-1	0.010831	-0.03926	0.060922	0.9995957	null
	3-1	0.038926	-0.01454	0.092395	0.385275	null
	4-1	0.053469	-0.00179	0.108732	0.0675764	null
	5-1	0.090805	0.029458	0.152152	0.0001283	***
	6-1	0.135664	0.074317	0.197011	0	***
	7-1	0.157771	0.090889	0.224654	0	***
	8-1	0.216215	0.157581	0.274849	0	***
	9-1	0.215172	0.154515	0.275829	0	***
	10-1	0.245016	0.178389	0.311644	0	***
	3-2	0.028094	-0.02983	0.086022	0.8774249	null
	4-2	0.042638	-0.01695	0.102225	0.4113988	null
	5-2	0.079974	0.014705	0.145243	0.0042432	**

6-2	0.124833	0.059563	0.190102	0.0000001	***
7-2	0.14694	0.076443	0.217437	0	***
8-2	0.205384	0.142657	0.26811	0	***
9-2	0.204341	0.13972	0.268962	0	***
10-2	0.234185	0.163929	0.304441	0	***
4-3	0.014543	-0.04791	0.076997	0.9992624	null
5-3	0.05188	-0.01602	0.119776	0.3139578	null
6-3	0.096738	0.028842	0.164635	0.0002912	***
7-3	0.118846	0.045909	0.191782	0.0000121	***
8-3	0.177289	0.111834	0.242745	0	***
9-3	0.176247	0.108973	0.24352	0	***
10-3	0.206091	0.133388	0.278794	0	***
5-4	0.037337	-0.03198	0.106655	0.7920515	null
6-4	0.082195	0.012877	0.151513	0.0068424	**
7-4	0.104303	0.030041	0.178564	0.0003897	***
8-4	0.162746	0.095817	0.229675	0	***
9-4	0.161703	0.092995	0.230411	0	***
10-4	0.191548	0.117516	0.26558	0	***
6-5	0.044858	-0.0294	0.119118	0.6597152	null
7-5	0.066966	-0.01193	0.14586	0.1788104	null
8-5	0.125409	0.053375	0.197444	0.0000018	***
9-5	0.124367	0.050677	0.198057	0.0000045	***
10-5	0.154211	0.075533	0.232889	0	***
7-6	0.022108	-0.05679	0.101001	0.9968391	null
8-6	0.080551	0.008517	0.152585	0.0148267	*
9-6	0.079508	0.005818	0.153199	0.022653	*
10-6	0.109353	0.030675	0.188031	0.00048	***
8-7	0.058443	-0.01836	0.135246	0.3198844	null
9-7	0.057401	-0.02096	0.135759	0.3759234	null

	10-7	0.087245	0.004179	0.170311	0.0304795	*
	9-8	-0.00104	-0.07249	0.070405	1	null
	10-8	0.028802	-0.04778	0.105383	0.9737262	null
	10-9	0.029844	-0.0483	0.107985	0.9709001	null
Composite	2-1	0.003275	-0.02183	0.028384	0.9999944	null
	3-1	0.016051	-0.01075	0.042854	0.6710854	null
	4-1	0.026719	-0.00098	0.05442	0.0694474	null
	5-1	0.038415	0.007663	0.069166	0.0031505	**
	6-1	0.064257	0.033506	0.095009	0	***
	7-1	0.068538	0.035012	0.102064	0	***
	8-1	0.091965	0.062573	0.121356	0	***
	9-1	0.085391	0.054985	0.115796	0	***
	10-1	0.098265	0.064866	0.131663	0	***
	3-2	0.012776	-0.01626	0.041814	0.9293002	null
	4-2	0.023444	-0.00643	0.053313	0.2763722	null
	5-2	0.03514	0.002422	0.067858	0.0238547	*
	6-2	0.060983	0.028265	0.0937	0.0000002	***
	7-2	0.065263	0.029925	0.100601	0.0000003	***
	8-2	0.08869	0.057247	0.120133	0	***
	9-2	0.082116	0.049723	0.114509	0	***
	10-2	0.09499	0.059773	0.130207	0	***
	4-3	0.010668	-0.02064	0.041974	0.9866522	null
	5-3	0.022364	-0.01167	0.056399	0.5408	null
	6-3	0.048206	0.014172	0.082241	0.0003288	***
	7-3	0.052487	0.015926	0.089048	0.0002488	***
	8-3	0.075914	0.043103	0.108725	0	***
	9-3	0.06934	0.035617	0.103062	0	***
	10-3	0.082214	0.04577	0.118658	0	***
	5-4	0.011696	-0.02305	0.046443	0.9877596	null

6-4	0.037539	0.002791	0.072286	0.0223256	*
7-4	0.041819	0.004594	0.079044	0.0140011	*
8-4	0.065246	0.031696	0.098795	0	***
9-4	0.058672	0.02423	0.093113	0.0000035	***
10-4	0.071546	0.034436	0.108657	0.0000001	***
6-5	0.025843	-0.01138	0.063067	0.4575414	null
7-5	0.030123	-0.00942	0.06967	0.3184728	null
8-5	0.05355	0.017441	0.089659	0.000123	***
9-5	0.046976	0.010037	0.083915	0.0023658	**
10-5	0.05985	0.020411	0.099289	0.0000726	***
7-6	0.004281	-0.03527	0.043828	0.9999989	null
8-6	0.027707	-0.0084	0.063816	0.3079408	null
9-6	0.021133	-0.01581	0.058072	0.7274028	null
10-6	0.034008	-0.00543	0.073447	0.1618199	null
8-7	0.023427	-0.01507	0.061926	0.6500226	null
9-7	0.016853	-0.02243	0.056132	0.939191	null
10-7	0.029727	-0.01191	0.071366	0.414854	null
9-8	-0.00657	-0.04239	0.029241	0.9998961	null
10-8	0.0063	-0.03209	0.044689	0.9999597	null
10-9	0.012874	-0.0263	0.052044	0.9896733	null

*p-value significance level: '***' 0.001; '**' 0.01; '*' 0.05; >0.05 is null/insignificant*

SI References

- Anselin, L., 1988a. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, 20(1), pp.1-17.
- Anselin, L., 1988b. A test for spatial autocorrelation in seemingly unrelated regressions. *Economics Letters*, 28(4), pp.335-341.
- Anselin, L., 2013. *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.
- Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V. and Pebesma, E.J., 2008. *Applied spatial data analysis with R* (Vol. 747248717, pp. 237-268). New York: Springer.
- Fotheringham, A.S., Brunson, C. and Charlton, M., 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Getis, A., 2009. Spatial weights matrices. *Geographical Analysis*, 41(4), pp.404-410.
- Griffith, D.A., 2005. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95(4), pp.740-760.
- Helbich, M., 2018. Toward dynamic urban environmental exposure assessments in mental health research. *Environmental Research*, 161, pp.129-135.
- Kelejian, H.H. and Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of econometrics*, 157(1), pp.53-67.
- Kelejian, H.H. and Prucha, I.R., 1998. A generalised spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), pp.99-121.
- LeSage, J.P., 2008. An introduction to spatial econometrics. *Revue d'économie industrielle*, (123), pp.19-44.
- Li, R., Cheng, S., Luo, C., Rutherford, S., Cao, J., Xu, Q., Liu, X., Liu, Y., Xue, F., Xu, Q. and Li, X., 2017. Epidemiological characteristics and spatial-temporal clusters of mumps in Shandong Province, China, 2005–2014. *Scientific Reports*, 7, p.46328.
- Su, S., Zhang, Q., Pi, J., Wan, C. and Weng, M., 2016. Public health in linkage to land use: Theoretical framework, empirical evidence, and critical implications for reconnecting health promotion to land use policy. *Land Use Policy*, 57, pp.605-618.
- Tsai, P.J., Lin, M.L., Chu, C.M. and Perng, C.H., 2009. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*, 9(1), p.464.
- Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12), pp.2290-2299.
- Waller, L.A. and Gotway, C.A., 2004. *Applied spatial statistics for public health data* (Vol. 368). John Wiley & Sons.
- Zhou, X. and Lin, H., 2007. Spatial Weights Matrix. In Shekhar, S. and Xiong, H. eds. *Encyclopaedia of GIS*. Springer Science & Business Media.
- Zhang, L., Zhou, S., Kwan, M.P., Chen, F. and Lin, R., 2018. Impacts of individual daily greenspace exposure on health based on individual activity space and structural equation modeling. *International journal of environmental research and public health*, 15(10), p.2323.