

The Future is Green.

Integrating Green and Blue Space Data from European Urban Atlas into UK

Biobank.

Mariya Geneshka^{1*}, Colin J McClean², Simon Gilbody^{1,3}, Joana Cruz⁴, and Peter Coventry^{1,5*}

¹ Department of Health Sciences, University of York, York YO10 4DD, UK

² Department of Environment and Geography, University of York, York, YO10 5NG, UK

³ Hull York Medical School, University of York, York, YO10 5DD, UK

⁴ Population, Policy and Practice Department, UCL Great Ormond St. Institute of Child Health, London WC1N 1EH, UK

⁵ York Environmental Sustainability Institute, University of York, York YO10 4DD, UK

*Corresponding author(s). Department of Health Sciences, University of York, York YO10 4DD, UK. Email: mmg529@york.ac.uk , peter.coventry@york.ac.uk

Word Count: 3043

Abstract

Background

Green and blue spaces can promote good physical and mental health and prevent the development of long-term conditions. Evidence suggests that not all green spaces affect health equally, and that certain types and properties of green spaces are stronger predictors of health than others. However, research into the causal mechanisms is limited in large cohorts due to lack of objective and comparable data on green space type, accessibility, and usage.

Methods

We used data from Urban Atlas to compute measures of urban park accessibility, street trees availability, and total green and blue space availability for 300,000 UK Biobank participants. Exposure metrics were computed using circular buffers with radii of 100 m to 3000 m. Pearson correlation coefficients and other descriptive statistical parameters were used to test agreement between variables and explore the utility of indicators in capturing different types of green spaces.

Results

Strong positive correlations were observed between variables of the same indicator with different buffer sizes. The presence of park and proportion of street tree canopy variables were negatively correlated with amount of total green space variables. This signifies distinct differences in type of green spaces captured by these variables.

Conclusions

Overall, five distinct indicators of park accessibility, street trees availability, and total green and blue space availability have been integrated into a large sample of the UK Biobank. Our method is replicable to settings across Europe and facilitates evidence-based research on the roles of different green and blue spaces in health promotion and ill-health prevention.

Key Words: environmental epidemiology, methods, data linkage, data manipulation, green space, blue space, large cohort, health, chronic health, evidence-based research

Key Messages

- Different types of green spaces and their position in the neighbourhood can promote and protect health by mitigating pollution and increasing physical activity and socialisation.
- We present the methods of constructing and linking data on urban green spaces, street trees and natural vegetation into a large health cohort, the UK Biobank.
- The ability to distinguish between types of green spaces and their intended use can help inform public health interventions, influence urban policy, and aid urban planning in building sustainable and healthy cities.
- Our methods are transferable and will allow others to explore the links between environment and health in UK Biobank and other health cohorts.

Introduction

Green and blue spaces are umbrella terms used to describe the presence of vegetation (green) and water (blue) bodies in the surrounding environment.[1] Green spaces can affect health through several bio-physiological pathways, including promoting health-related behaviours like physical activity; increasing socialisation and connectedness with people and nature; and reducing the presence of air pollutants and noise.[2, 3] Exposure to green space can potentially reduce the risk of several non-communicable diseases (NCDs). In a longitudinal study, for example, the risk of developing cardio-vascular disease (CVD), myocardial infarction, and stroke was respectively 15%, 23% and 13% lower in those who had more green spaces in their neighbourhood compared to those who had fewer green spaces.[4] Moreover, meta-analyses found 28% reduction in the risk of type II diabetes and 23% reduction in the risk of stroke mortality in those with greater amount of green space in their neighbourhood.[5, 6]

Green spaces also impact mental health and wellbeing. Observational, epidemiological studies found higher amount of green space was associated with lower odds of stress, higher self-perceived general health, and higher quality of life.[7-9] Furthermore, longitudinal research showed that those living in areas with lowest amount of residential greenery had 24% to 52% higher risk of developing schizophrenia compared with those living in areas with most greenery.[10, 11] The relationship between green space and depression is less well understood, but cross-sectional studies showed a moderate decrease in the odds of depression with greater amount of green space and higher accessibility to parks of large areas.[12, 13]

Growing evidence suggests that, in addition to amount of green space, different types of green and blue space environments affect health differently.[14] Green areas, such as grasslands, serene environments, higher number of forests and higher number of urban green spaces reduced the risk of poor mental health.[15, 16] However, no protective relationships with mental health were observed for exposures to saltwater bodies, wetlands, rangeland, and agricultural land.[15, 16] In urban settings, higher availability of street trees, but not higher availability of grass or total green space, was associated with lower odds of diabetes and CVD events.[17, 18] Furthermore, a comparison between higher availability of public parks and total green space area showed that public parks reduced blood pressure while larger areas of green space in the neighbourhood showed no significant relationship with blood pressure change.[19]

Type, position, and duration of exposure to green/ blue space affect health at different rates and potentially through different causal pathways. In systematic reviews, higher amount of street trees, good accessibility to parks and some types of land use classes showed stronger relationships with health than higher availability of grass or total green space.[20-22] However, the mechanisms and direction of causation behind these relationships are still not fully understood. This is partly due to lack of high quality comparable observational research. Despite the World Health Organisation's (WHO) [23] call to include more objective and comparable measures of green space accessibility and usage in evidence-based research, cohort studies still mainly use single measures of greenery, such as the Normalized Difference Vegetation Index (NDVI), or proportion of green space.[13, 14, 24] Reasons for this include lack of availability of objective environmental data in health cohorts, and limited characterisation of different green spaces present in the surrounding environment.[24]

The UK Biobank is a comprehensive resource for studying the complex interactions between the environment and health. However, studies on the relationship of exposure to green and blue spaces and health have mainly used NDVI and data from older land use datasets to capture proportion of greenery on limited spatial scales.[25-28] We describe here the methods to link high resolution data on street trees, parks, and amount of green and blue space from 2006 and 2012 Urban Atlas dataset into the UK Biobank. Based on prior research, we hypothesise that it is certain types of green spaces, such as parks and street trees, and their properties (e.g., type, position, size, or intended use) that influence health through different bio-physiological pathways (fig. 1).[21, 22] Therefore, the aim of this study is to construct multiple distinct indicators of green and blue space availability and accessibility that can be used in evidence-based health research. To achieve this, we first describe the process and methods of computing and integrating variables of urban park accessibility, street trees availability, and total green and blue space availability into UK Biobank participants using geographical information systems programming. Second, we calculate the statistical parameters of these variables to descriptively characterise each exposure metric and its spatial scale. Third, we use correlation coefficients to test the agreement between variables and explore the ability of indicators in capturing different types of green spaces in the surrounding neighbourhood. This linked dataset will be used to model the cross-sectional associations between green space availability and accessibility and the prevalence of multimorbidity disease clusters. It will also become available for anyone seeking to conduct research using UK Biobank data.

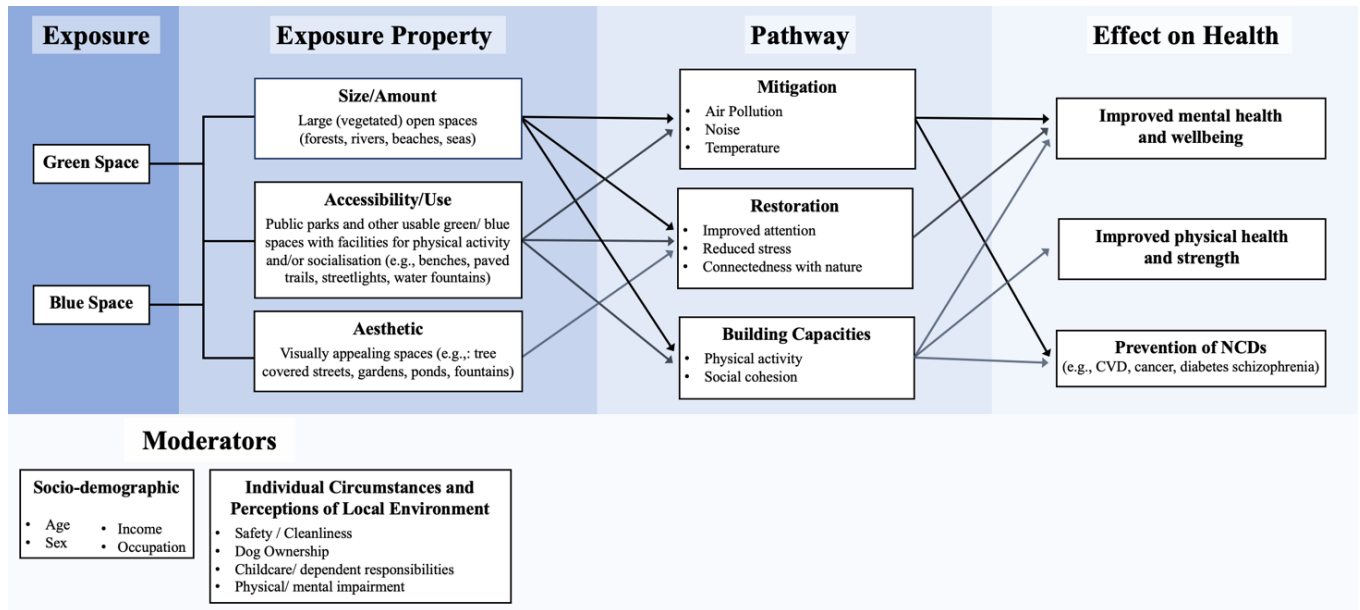


Figure 1: Conceptual framework for the relationship between green and blue space properties and health

Methods

Description of Data Sources

UK Biobank

The UK Biobank is a large, population-based prospective cohort of 502,650 men and women aged 40-69 years.[29] Recruitment for voluntary participation started in 2006 and finished in 2010. Participants were selected if they resided within 25 miles of one of the 22 UK Biobank assessment centres around the UK and were registered with a National Health Service (NHS) General Practitioner (GP). Data about participants' socio-demographic characteristics, living arrangements, lifestyle factors, built-environment exposures, and medical history were collected through touchscreen questionnaires and nurse-led interviews. Follow-up of the cohort is ongoing and disease status and mortality are tracked through electronic health records and cancer registries.[29]

Urban Atlas

The European Urban Atlas (UA) is a land use dataset covering over 200 European Functional Urban Areas (FUA) with a population of at least 50,000 people.[30, 31] FUAs are defined by the European audit as the area of a metropolitan city and its surrounding commuting zone. The commuting zone is the area around the city where at least 15% of the employed residents are commuters into the city.[32] The UA dataset contains nomenclature of 20 land use classes, 17 of which are built-environment classes with a minimum mapping unit (MMU) of 0.25 ha and 3 are natural classes with a MMU of 1 ha. The overall minimum accuracy for the

data is 80% and the minimum mapping width is 10 m. The UA datasets are collated from topographic maps, 2.5 m Earth Observation spatial resolution multispectral data, SPOT 5 satellite imagery and other Very High Resolution (VHR) imagery for the years 2006, 2012 and 2018.[31] The data is open access and available from the European Environment Agency and the Copernicus Land Monitoring website (<https://land.copernicus.eu>).

The validity of UA in capturing amount of green space has been previously tested against the CORINE, UK Land Cover Map and NDVI datasets. It was established that UA produces comparable results to all three of these indicators.[33, 34] The *Urban green areas* layer from the UA has also been endorsed by WHO [23] as a suitable indicator of urban green space accessibility and is a preferred indicator for capturing usable green spaces in urban areas due to its high resolution and ability to measure green space change over time.[35]

Data Selection and Processing

UK Biobank participants' residential address location coordinates (rounded to 100 m accuracy) at baseline were used as proxies for permanent residence. UK Biobank participants predominantly live in urban areas and UA data is available for about 300,000 UK Biobank participants who reside within the boundaries of the following FUAs: London, Bristol, Cardiff, Stoke-on-Trent, Nottingham, Sheffield, Leeds, Manchester, Liverpool, Newcastle upon Tyne, Edinburgh, and Glasgow.

Exposure Metrics

Several natural and artificial land use classes were used to compute exposure metrics of accessibility to park; availability of street tree canopy; availability of total green space; availability of total blue space; and availability of total green and blue space at different spatial scales (see table 1 for further information).[36]

Table 1: Descriptions of Computed Exposure Metrics

Exposure Indicator	Exposure Metric	UA Nomenclature used (year of data collection) [36]	UA Nomenclature Description [40]	Buffer Sizes
Accessibility to Usable Urban Park	Distance to Park	14100 Green urban areas (2006)	Green areas accessible to the public for predominantly recreational use, such as public parks, gardens, and zoos.	N/A
	Presence of Park			300m, 1500m
Availability of Street Trees	Amount of Street Tree Canopy	Street Tree Layer (2012)	Contiguous patches of trees (≥ 5 m height) covering an area of at least 500m ² over Level 1 <i>Artificial surfaces</i> nomenclature layers	300m, 1500m
Availability of Total Green Space	Amount of Total Green Space	14100 Green urban areas (2006) 20000 Agricultural + seminatural + wetland areas (2006) 31000 Forests (2006)	<i>Agricultural + seminatural + wetland areas</i> layer: range of natural open spaces including arable land, pastures, grasslands, arable crops, moors, beaches, bare rocks, snow and ice and wetlands. <i>Forests</i> layer: areas with tree canopy coverage of over 30% and tree height of over 5 m.	100m, 300m, 1500m, 3000m

Availability of Blue Space	Amount of Total Blue Space	5000 Water Bodies	Visible water surface area of rivers, lakes, ponds and canals	100m, 300m, 1500m, 3000m
Availability of Total Green and Blue Space	Amount of Total Green & Blue Space	14100 Green urban areas (2006) 20000 Agricultural + seminatural + wetland areas (2006) 31000 Forests (2006) 5000 Water Bodies	See above	100m, 300m, 1500m, 3000m

Accessibility to urban park was measured as the straight-line (Euclidean) distance from a participant's residential address to their nearest public park using data from the UA 2006 nomenclature layer: *Green urban areas*.

Availability of street trees in the neighbourhood was measured as amount (proportion) of street tree canopy in straight-line (Euclidean) radial distance buffers around the residential address. Availability of total green space, blue space and total green and blue space was measured in amount (proportion) within a straight-line (Euclidean) radial distance buffers around the residential address. The UA data layers used to construct each indicator are described in Table 1.

Data Processing and Analysis

All data were processed in ArcGIS Pro, a 64-bit Desktop application facilitated by Esri ArcGIS Platform. The 2006 UA vector datasets were rasterized to 50 m resolution and the 2012 *Street Tree Layer* dataset was rasterised to 10 m resolution. The six-digit residential location coordinates of UK Biobank participants were overlaid with the rasterized UA data. Proportion green/blue space within distance buffers and the Euclidean distance to park were then computed for each UK Biobank participant (see table 1 for further details on buffer size). Participants whose residential address or buffer area fell on or outside the boundary of the UA data were excluded.

Data Analysis

The derived environment data was transferred from ArcGIS Pro to RStudio for analysis. Descriptive statistics, such as central tendency (i.e., means, medians), frequencies and dispersion measures (i.e., standard deviation and inter-quartile ranges) were calculated. Bivariate data analysis, chi-squared tests, Pearson tetrachoric, point biserial and product moment correlations were also applied according to data type.

Results

Table 2 describes the statistical parameters of the computed green and blue space variables.

Just under two-thirds (62.94%) of UK Biobank participants had a park within 300 m of their residential address and almost all (98%) had a park within 1500 m of the residential address.

The mean straight-line distance to a park from the residential address was 291.48 m (table 2).

The Chi-squared test suggests there's strong correlation between presence of park in 300 m and presence of park in 1500 m variables.

Table 2: Statistical parameters of computed green and blue space variables

Metrics	Var. Name	n (observed cases)	Mean (%)	Standard Deviation	Median (%)	Min (%)	Max (%)
Total green space in 100m	total_100_green	311,451	10.26	20.32	0.00	0.00	100.00
Total green space in 300m	total_300_green	308,979	15.40	18.32	8.84	0.00	99.03
Total green space in 1500m	total_1500_green	239,747	27.78	20.57	21.19	0.04	99.74
Total green space in 3000m	total_3000_green	192,094	32.31	21.10	25.94	1.47	98.85
Total blue space in 100m	blue_100	311,451	0.24	2.68	0.00	0.00	95.49

Total blue space in 300m	blue_300	308,979	0.46	2.66	0.00	0.00	87.54
Total blue space in 1500m	blue_1500	271,118	1.12	2.55	0.21	0.00	50.51
Total blue space in 3000m	blue_3000	219,462	1.20	1.70	0.56	0.00	47.92
Tree canopy in 300m	tree_300	138,831	20.03	17.13	15.30	0.04	99.60
Tree canopy in 1500m	tree_1500	171,734	19.14	13.47	16.70	0.00	85.06
Total green and blue space in 100m	green_blue_100	311,451	10.48	20.51	0.00	0.00	100.00
Total green and blue space in 300m	green_blue_300	308,979	15.86	18.52	8.84	0.00	100.00
Total green and blue space in 1500m	green_blue_1500	239,719	28.93	20.50	22.85	0.29	99.74
Total green and blue space in 3000m	green_blue_3000	180,572	34.04	21.19	27.70	2.10	99.02
Distance to park	distance_park	312,284	291.48*	378.90*	201.15*	0.00*	15843.71*
	Var. Name	Yes (%)			No (%)		
Presence of park in 300m	park_300	194,481 (62.94%)			114,498 (37.06%)		
Presence of park in 1500m	park_1500	273,172 (98.14%)			5,190 (1.86%)		
X-squared = 9902.8, df = 1, p-value < 2.2e-16							

* Unit of measurement is meters (not percentage)

Median and interquartile range values for all other environment variables are shown in box and whisker plots in figure 2. Overall, the data are skewed towards the null. The median amount of green/blue space increases with buffer size. The median amount of total green space in a 100 m buffer is around 0%. This increases to 26% in a 3000 m buffer. Median amount of blue space follows a similar pattern, but values tend to stay around 0% for all buffer sizes. Median amount of street tree canopy is 15% in a 300 m buffer and 17% in a 1500 m buffer.

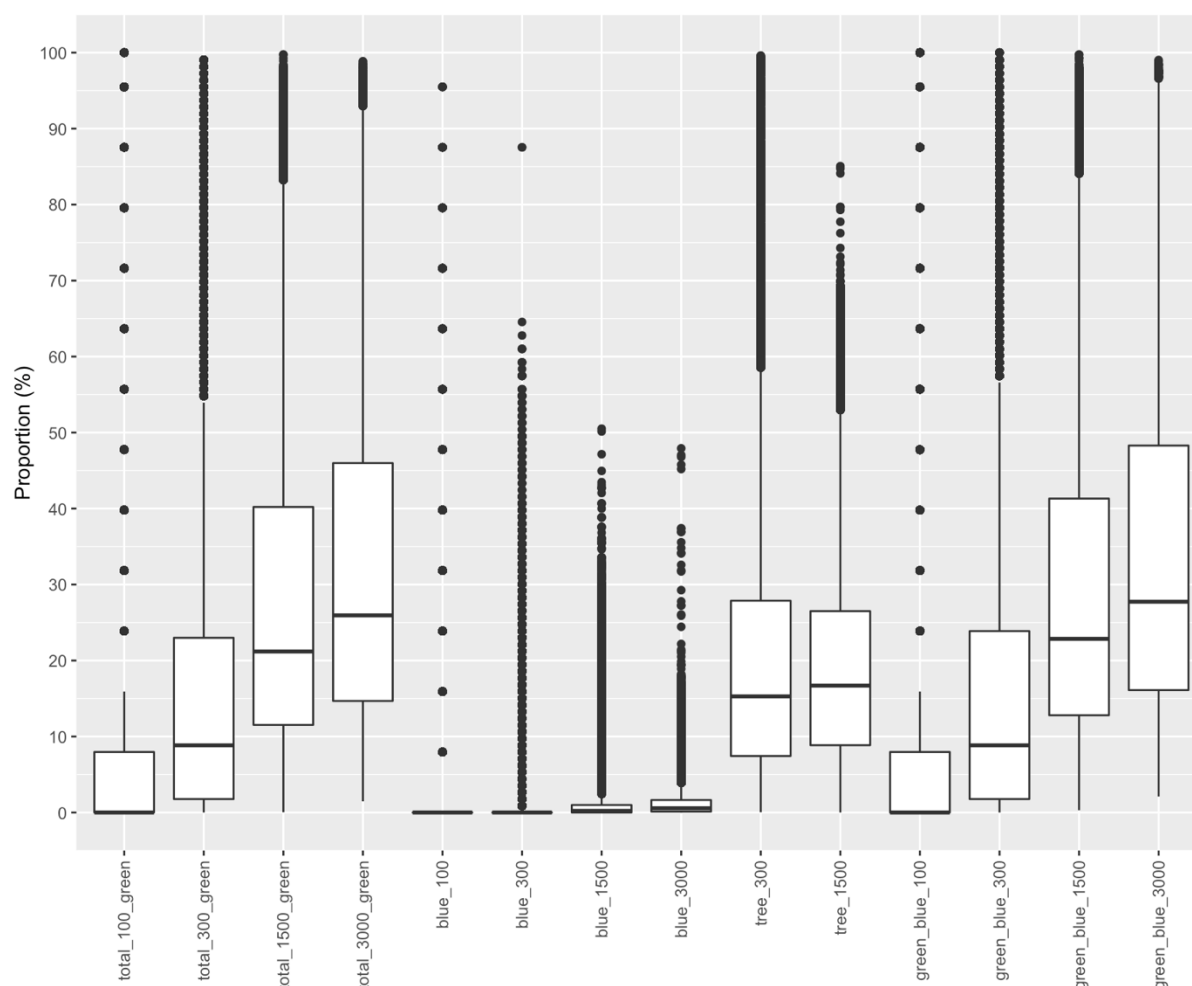


Figure 2: Box and whisker plots of variables

Figure 3 shows a Pearson correlation matrix of the computed variables. Overall, variables of the same exposure indicator with different buffer sizes tend to have strong positive correlations with each other. The correlation coefficients are weaker between variables that have larger buffer size differences. There's strong positive correlation between street tree canopy in 300 m buffer and street tree canopy in 1500 m buffer ($r = 0.78$).

The correlation matrix also shows negative correlations between the presence of park variables, street tree canopy variables and amount of total green space variables. Weak positive correlations were observed for presence of park in 300 m buffer variable (park_300) with amount of total green space and amount of total green and blue space variables. However, weak to moderate negative correlations were observed for presence of park in a 1500 m buffer variable (park_1500) with amount of total green space and amount of total green and blue space variables (fig. 2). There are also weak negative correlations between the amount of street tree canopy in 1500 m buffer variable (tree_1500) and the amount of total green and blue space variables.

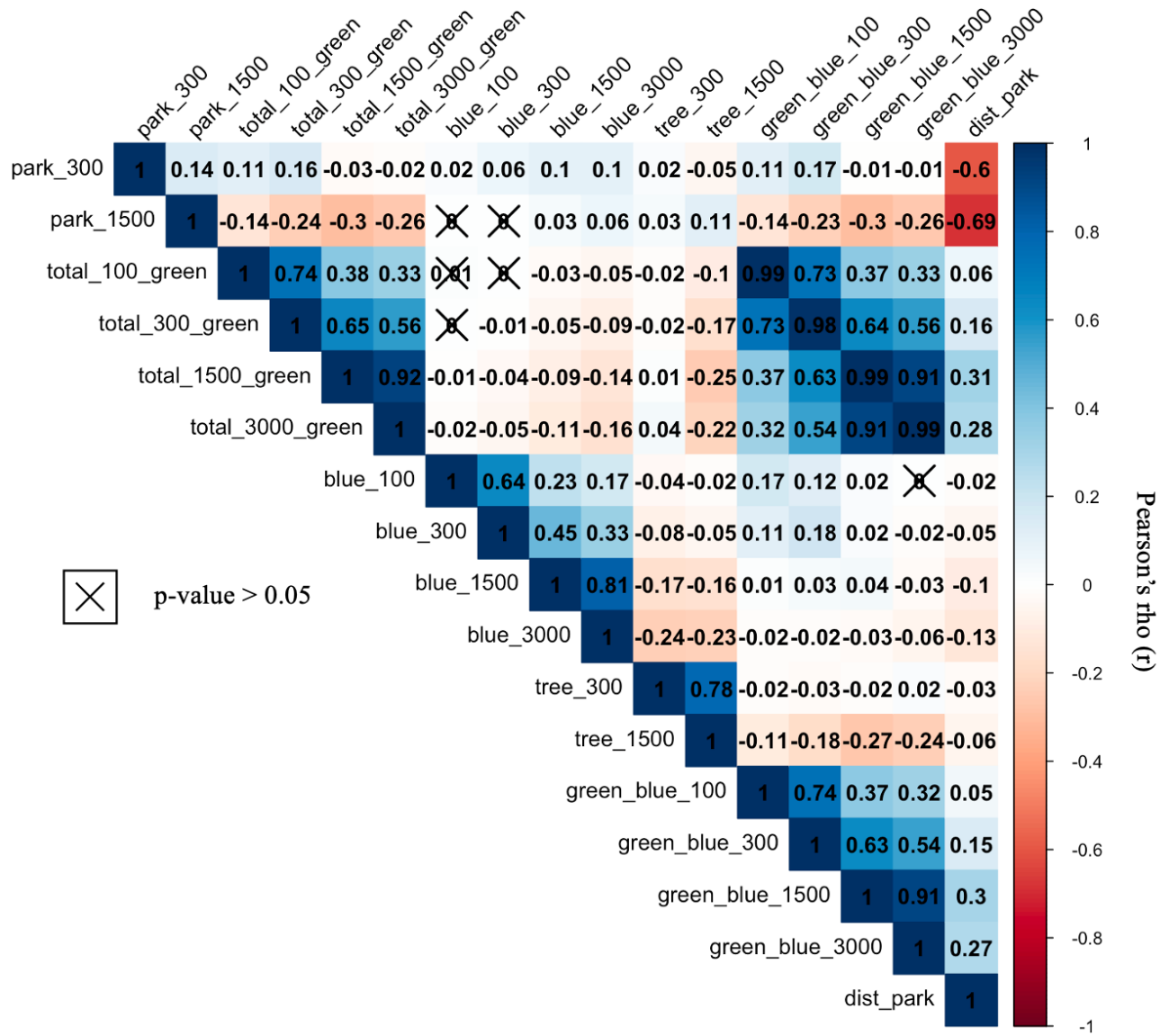


Figure 3: Pearson correlation matrix of variables

Discussion

Interpretation of Computed Green Space Indicators

In this data linkage study, we present the process and results of linking green and blue space exposure data from Urban Atlas with the UK Biobank. Five distinct indicators were constructed: accessibility to park; availability of street trees; availability of total amount of green space; availability of total amount of blue space; and availability of total amount of green and blue space. Accessibility to park is measured by the presence of a park in 300 m buffer, presence of park in 1500 m buffer, and the Euclidean distance to nearest park. Having a park within 1500 m of the residential address was negatively correlated with having higher amount of total green space. The amount of total green space variables were computed using two natural land use layers: *Forests* and *Agricultural + seminatural + wetland areas*, and one artificial surfaces layer: *Green urban areas*. The *Green urban areas* layer was used to compute the presence of park variables and while there's overlap in data, the negative correlations between the two indicators suggest that participants who have high proportion of total green space in 1500 m buffer around their residential address likely live on the rural-urban fringe and have higher availability of forests and agricultural vegetation rather than urban green areas. The amount of total green space variables, therefore, are indicators of availability of naturally growing outdoor vegetation, while the presence of park variables are indicators of accessibility to usable urban green spaces.

With the emerging evidence on the health-promoting role of street trees we also aimed to compute an indicator of street trees availability.[20] This is captured by the amount of street tree canopy variables. Higher proportion of street trees showed to be negatively correlated with both higher proportion of total green space and presence of parks. This is expected

because UA nomenclature, *Street Tree Layer*, is a non-overlapping dataset of patches of street tree canopy over artificial surfaces only.[36] These results show that, together, the five indicators are distinct but complementary to each other. They each capture different types of green spaces in the surrounding neighbourhood. The street trees and distance to park variables are indicators of two types of accessible green spaces: public parks and trees lining roads and paths. The amount of total green space variables, on the other hand, are measures of natural vegetation mainly found on the fringe of rural and urban areas.

Implications for Health Research and Policy

The integration of multiple, distinct green space indicators into UK Biobank addresses an important data gap in health research. Lack of high-quality, comparable data on green space accessibility, availability and use in epidemiological studies and in the UK Biobank has resulted in little research into the causal mechanisms behind the relationship of exposure to green and blue space and health.[24, 37, 38] The ability to distinguish between types of green spaces, their position in the neighbourhood and their intended use can strengthen evidence-based research and inform public health practice. Outdoor green and blue spaces can be utilised to deliver targeted interventions to promote healthy lifestyles and prevent ill-health. This research can also help wider urban planning and policy in creating healthier urban environments where green and blue spaces are integral parts of healthy daily living.[23]

Strengths and Limitations

This data linkage study has several strengths. To the best of our knowledge, it is the first study to integrate multiple comparable indicators of urban park accessibility and street trees availability into a large, population-based UK cohort. Previously, data linkages have been limited to single area case studies or focused on built-environment indicators.[24] The large sample size of the UK Biobank improves precision and power of future research, avoids bias, and opens opportunities to explore the still limited but strong evidence for the health-promoting benefits of street trees, parks and natural vegetation.[39-41] Furthermore, our method and exposure metrics are replicable to settings across Europe. The UA dataset covers urban areas across Eastern and Western Europe, allowing objective exposure assessment for different populations and settings.[36]

We also computed exposures at an individual level, which means every participant has a unique measure of greenery in their neighbourhood. This improves accuracy in exposure measurement and prevents ecological bias (and potential fallacy) that commonly occurs when aggregate measures are used in individual analyses.[42] The computation of buffer sizes ranging from 100 m to 3000 m, on the other hand, allows for comparative research at different spatial scales. Three-hundred meters is recommended as the maximum distance anyone should live from an accessible green space; however, this figure is arbitrary, and research has shown that both small (500 m) and large (up to 2000 m) buffer sizes can be strong predictors of health. [23, 43, 44]

Our data linkage study has multiple strengths, but it is not without limitations. Our analyses are primarily based on measuring proportion of green and blue space within straight-line

distance buffers. These buffers capture any greenery that falls within the buffer's area, irrespective of the green space's accessibility or ownership. One advance in geo-computation has been the ability to capture greenery only along publicly accessible roads and paths.[45] This approach might be useful for some exposure metrics, such as street trees, but it's also computationally intensive and doesn't necessarily improve accuracy in exposure measurement because circular and road-network buffers tend to capture similar amount of greenery.[46]

Other limitations include the geo-processing resolution of the data. The 2006 UA layers were converted to raster datasets with a resolution of 50 meters. Rasterising a vector feature may cause potential data loss of areas smaller than the size of the raster grid cell. We chose 50 meters as a grid cell size because it facilitated geoprocessing for large sample sizes. Smaller grid cell sizes, while more accurate, could not allow large scale data analysis due to large memory storage.

Conclusions

In conclusion, we have described a unique set of methods and results about how to link high quality, open-access land use data from Urban Atlas with the UK Biobank health cohort. We produced five novel indicators of urban park accessibility; street tree availability; and total green/ blue space availability. This broadens opportunities for comparable epidemiological research in a large health cohort. Given the availability of UA data for UK and Europe, we believe our methods are also replicable to settings, populations and curated cohorts outside of the UK Biobank.

References

- 1 White MP, Elliott LR, Grellier J, et al. Associations between green/blue spaces and mental health across 18 countries. *Scientific Reports* 2021;11(1):1-2.
- 2 Markevych I, Schoierer J, Hartig T, et al. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environ Res* 2017;158:301-17.
- 3 Nieuwenhuijsen MJ, Khreis H, Triguero-Mas M, et al. Fifty shades of green. *Epidemiol* 2017;28(1):63-71.
- 4 Seo S, Choi S, Kim K, et al. Association between urban green space and the risk of cardiovascular disease: A longitudinal study in seven Korean metropolitan areas. *Environ Int* 2019;125:51-7.
- 5 Twohig-Bennett C, Jones A. The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environ Res* 2018;166:628-37.
- 6 Yuan Y, Huang F, Lin F, et al. Green space exposure on mortality and cardiovascular outcomes in older adults: a systematic review and meta-analysis of observational studies. *Aging Clin Exp Res* 2021;33(7):1783-97.
- 7 Stigsdotter UK, Ekholm O, Schipperijn J, et al. Health promoting outdoor environments- Associations between green space, and health, health-related quality of life and stress based on a Danish national representative survey. *Scand J Public Health* 2010;38(4):411-7.
- 8 de Vries S, Verheij RA, Groenewegen P, et al. Natural environments—healthy environments? An exploratory analysis of the relationship between greenspace and health. *Environ Plan A* 2003;35(10):1717-31.
- 9 Maas J, Verheij RA, et al. Green space, urbanity, and health: how strong is the relation?. *J Epidemiol Community Health* 2006;60(7):587-92.
- 10 Engemann K, Pedersen CB, Arge L, et al. Childhood exposure to green space—a novel risk-decreasing mechanism for schizophrenia?. *Schizop Res* 2018;199:142-8.
- 11 Rotenberg M, Tuck A, Anderson KK, et al. Green Space and the Incidence of Schizophrenia in Toronto, Canada. *Canadian J Psychiatry* 2022;67(3):238-40.
- 12 Min KB, Kim HJ, Kim HJ, et al. Parks and green areas and the risk for depression and suicidal indicators. *Int J Public Health* 2017;62(6):647-56.

- 13 Gascon M, Triguero-Mas M, Martínez D, et al. Mental health benefits of long-term exposure to residential green and blue spaces: a systematic review. *Int J Environ Res Public Health* 2015;12(4):4354-79.
- 14 Lachowycz K, Jones AP. Greenspace and obesity: a systematic review of the evidence. *Obes Rev* 2011;12(5):e183-9
- 15 Alcock I, White MP, Lovell R, et al. What accounts for 'England's green and pleasant land'? A panel data analysis of mental health and land cover types in rural England. *Landsc Urban Plan* 2015;142:38-46.
- 16 Akpinar A, Barbosa-Leiker C, Brooks KR. Does green space matter? Exploring relationships between green space type and health indicators. *Urban For Urban Green* 2016;20:407-18.
- 17 Astell-Burt T, Feng X. Association of urban green space with mental health and general health among adults in Australia. *JAMA Netw Open* 2019;2(7):e198209-.
- 18 Astell-Burt T, Feng X. Urban green space, tree canopy and prevention of cardiometabolic diseases: a multilevel longitudinal study of 46 786 Australians. *Int J Epidemiol* 2020;49(3):926-33.
- 19 Jimenez MP, Wellenius GA, James P, et al. Associations of types of green space across the life-course with blood pressure and body mass index. *Environ Res* 2020;185:109411.
- 20 Wolf KL, Lam ST, McKeen JK, et al. Urban trees and human health: A scoping review. *Int J Environ Res Public Health* 2020;17(12):4371.
- 21 Nguyen PY, Astell-Burt T, Rahimi-Ardabili H, et al. Green Space Quality and Health: A Systematic Review. *Int J Environ Res Public Health* 2021;18(21):11028.
- 22 Rigolon A, Browning MH, McAnirlin O, et al. Green space and health equity: a systematic review on the potential of green space to reduce health disparities. *Int J Environ Res Public Health* 2021;18(5):2563.
- 23 World Health Organization. Urban green spaces and health. World Health Organization. Regional Office for Europe; 2016.
- 24 Geneshka M, Coventry P, Cruz J, et al. Relationship between Green and Blue Spaces with Mental and Physical Health: A Systematic Review of Longitudinal Observational Studies. *Int J Environ Res Public Health* 2021;18(17):9010.

- 25 Zhang Z, Chen L, Qian ZM, et al. Residential green and blue space associated with lower risk of adult-onset inflammatory bowel disease: Findings from a large prospective cohort study. *Environ Intl* 2022;160:107084.
- 26 Sarkar C. Residential greenness and adiposity: Findings from the UK Biobank. *Environ Intl* 2017;106:1-0.
- 27 Sarkar C, Webster C, Gallacher J. Residential greenness and prevalence of major depressive disorders: a cross-sectional, observational, associational study of 94 879 adult UK Biobank participants. *The Lancet Planet Health* 2018; 2(4):e162-73.
- 28 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP)—a nationwide resource for evidence-based healthy city planning and public health interventions. *Annals of GIS* 2015;21(2):135-48.
- 29 Allen N, Sudlow C, Downey P, et al. UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology* 2012;1(3):123-6.
- 30 [dataset]* European Environment Agency. (2006), Data from: European Urban Atlas, Copernicus Land Monitoring, Dataset, <https://land.copernicus.eu/local/urban-atlas/urban-atlas-2006?tab=download>
- 31 Montero E, Wolvelaer JV, Garzón A. The European urban atlas. In: Land use and land cover mapping in Europe 2014, Springer, Dordrecht: 2014:115-124.
- 32 European Commission and Statistical Office of the European Union. Applying the degree of urbanisation—a methodological manual to define cities, towns and rural areas for international comparisons—2021 edition. Publications Office of the European Union. 2021.
- 33 Barboza EP, Cirach M, Khomenko S, et al. Green space and mortality in European cities: a health impact assessment study. *Lancet Planet Health* 2021;5(10):e718-30.
- 34 Zhou Y, von Lengerke T, Dreier M. Comparing different data sources by examining the associations between surrounding greenspace and children's weight status. *Int J Health Geogr* 2021;20(1):1-3.
- 35 van den Bosch M, Mudu P, Uscila V, et al. Development of an urban green space indicator and the public health rationale. *Scand J Public Health* 2016;44(2):159-67.
- 36 Seifert FM. Improving Urban Monitoring toward a European Urban Atlas. Global Mapping of Human Settlement: Experiences, Datasets, and Prospects 2009 Jun 23:231.

- 37 Thompson DA, Geary RS, Rowney FM, et al. Cohort Profile: The Green and Blue Spaces (GBS) and mental health in Wales e-cohort. *Int J Epidemiol* 2022;0:1-10
- 38 Gonzales-Inca C, Pentti J, Stenholm S, et al. Residential greenness and risks of depression: Longitudinal associations with different greenness indicators and spatial scales in a Finnish population cohort. *Health Place* 2022;74:102760.
- 39 Faber J, Fonseca LM. How sample size influences research outcomes. *Dental Press J Orthodontics* 2014;19:27-9.
- 40 Pearce J, Shortt N, Rind E, et al. Life course, green space and health: Incorporating place into life course epidemiology. *Int J Environ Res Public Health* 2016;13(3):331.
- 41 Noordzij JM, Beenackers MA, Groeniger JO, et al. Effect of changes in green spaces on mental health in older adults: a fixed effects analysis. *J Epidemiol Community Health* 2020;74(1):48-56.
- 42 Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989;18(1):269-74.
- 43 Mazumdar S, Chong S, Astell-Burt T, et al. Which Green Space Metric Best Predicts a Lowered Odds of Type 2 Diabetes?. *Int J Environ Res Public Health* 2021;18(8):4088.
- 44 Browning M, Lee K. Within what distance does “greenness” best predict physical health? A systematic review of articles with GIS buffer analyses across the lifespan. *Int J Environ Res Public Health* 2017;14(7):675.
- 45 Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *Int J Health Geogr* 2007;6(1):1-1.
- 46 Bodicoat DH, O'Donovan G, Dalton AM, et al. The association between neighbourhood greenspace and type 2 diabetes in a large cross-sectional study. *BMJ Open* 2014;4(12):e006076.

Ethics approval

Researchers do not require a separate ethics approval to use UK Biobank data and are covered under the UK Biobank ethics approval by the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank. This project was approved by the UK Biobank's Access Management System (AMS), application no. 73700, and grants access to restricted UK Biobank Fields 22701 and 22703 (*home location – east coordinate and home location – north coordinate*).

Author contributions

MG designed and led the development of the study. MG produced the data fields, conducted the data linkage and drafted the paper. PC, CM, JC and SG made critical intellectual contributions to the paper. CM advised and assisted on the geoprocessing of the data. JC sourced the UA data, formulated the theoretical background underpinning the green and blue space indicators construction and the metric spatial scales. PC formulated the theoretical background underpinning the study's aims and objectives, and aided the interpretation of results in the discussion section. SG advanced the development of the introduction and discussion sections and helped obtain UK Biobank data. PC and SG obtained funding for the UK Biobank data. PC, CM and SG reviewed and edited the paper.

Data availability

Data fields derived from this study and all other data underlying this article are owned by UK Biobank and can be accessed with an approved project application from the UK Biobank's Access Management System (AMS). Raw environment data from Urban Atlas is open access and available from European Environment Agency.

Supplementary data

None

Funding

The research was funded by the NIHR Applied Research Collaboration Yorkshire and Humber <https://www.arc-yh.nihr.ac.uk/>. The views expressed are those of the author(s), and not necessarily those of the NIHR or the Department of Health and Social Care. PC and SG are part funded by the UK Research and Innovation Closing the Gap Network+ [ES/S004459/1]. UKRI does not necessarily endorse the view expressed by the authors.

Acknowledgements

We'd like to thank Prof. Piran White from Department of Environment and Geography at University of York, Dr. Rachel Pateman from Stockholm Environment Institute, and Dr.

Andre Bedendo De Souza from Health Sciences Department at University of York for their help and intellectual contribution to this paper. Their expertise in environmental data analysis and statistics has helped shape this project.

We'd also like to acknowledge the UK Biobank for providing the participant location coordinate data and storing and maintaining the derived data fields in this study. Finally, we'd also like to acknowledge the European Environment Agency for funding and facilitating the development of the Urban Atlas dataset. The Urban Atlas is an open access resource available from Copernicus Land Monitoring (<https://land.copernicus.eu>).

Conflict of interest

None declared