

UK Biobank Project 15374

Documentation of environmental indicators attributed to participants based on home location grid references

Benedict Wheeler

European Centre for Environment and Human Health

University of Exeter Medical School

b.w.wheeler@exeter.ac.uk

Citations and data licences

The derived variables described here are derived from data subject to licences and citation requirements. Any re-use of these variables should cite accordingly as detailed in Annex 1.

Acknowledgement

The research was funded in part by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Environmental Change and Health at the London School of Hygiene and Tropical Medicine in partnership with Public Health England (PHE), and in collaboration with the University of Exeter, University College London, and the Met Office.

Data obtained from UK Biobank

Fields:

n_eid – Encoded anonymised participant ID

n_20033_0_0 – easting baseline wave 0

n_20034_0_0 – northing baseline wave 0

n_20033_1_0 – easting wave 1

n_20034_1_0 – northing wave 1

These high resolution (100m) UKB home location grid references were mapped in ArcGIS (v10, ESRI, Redlands, CA) to produce two new feature classes within a geodatabase – one for baseline (n=497679 of 502647, total due to missing location data), one for wave 1 follow-up locations (n=20330). Further, four withdrawn participants (notified Feb 2016) were removed from the dataset, making n=502643.

Mapping and Linkage: ‘Greenspace’

Measures of residential greenspace exposure were estimated for England residents using the 2005 Generalised Land Use Database for England (Department for Communities and Local Government, 2007). This provides data on land use distribution for 2001 Census Output Areas in England, and is consistent with previous related research (Alcock et al., 2014; White et al., 2013b). Data on natural environment were also linked using CEH 2007 Land Cover Map data (Morton et al., 2011), since these were available for the whole of GB, and hence the entire UKB sample, and also have been used in related research (Wheeler et al., 2015).

There is no consistent use of distance/proximity metrics in the literature (for example see (Gascon et al., 2016); however, based on existing evidence finding associations between greenspace density to health outcomes, and relevant public policy on greenspace accessibility, we applied proximity criteria of 300m and 1km (Agay-Shay et al., 2014; Maas et al., 2009; Natural England, 2010; Reklaitiene et al., 2014) to indicate nearby greenspace and wider-area greenspace relative to the participant’s home location.

Participant home locations (at wave 0 and wave 1) were therefore buffered at 300m and 1000m to create polygon feature classes in an ArcGIS Geodatabase. GLUD and CEH environmental data were allocated to these buffers using tools within the Geospatial Modelling Environment (GME, <http://www.spataleecology.com/gme/>) as described below.

Generalised Land Use Database (GLUD) 2005

1. GLUD 2005 data for 2001 Census Output Areas (COA) were obtained from Neighbourhood Statistics (<http://www.neighbourhood.statistics.gov.uk/>) – these data are available for England only. Data were joined to full resolution COA 2001 boundaries.
2. For each COA, the area percentage of the GLUD categories greenspace, domestic gardens and water were calculated, as a proportion of the total percentage of all land use types.
3. The GME command 'isectpolypoly' was used to intersect the COA feature class with the participant home location buffers, to allocate an area weighted mean of each land use percentage coverage to each participant based on the two buffer sizes.
4. Participants with a home location outside of England were excluded from this linkage due to the restricted availability of GLUD data.

Land Cover Map (LCM) 2007

To extend coverage to Wales and Scotland, Centre for Ecology and Hydrology (CEH) Land Cover Map 2007 data were used to derive a similar 'greenspace' density measure to that in the GLUD data. Raster LCM 2007 data for 25m grid cells were obtained (Morton et al., 2011), and the 23 classes were reclassified to a binary classification; classes 1-21 ('natural environment') and 22-23 ('built environment'). Note that the built environment classes here include gardens, and the minimum mapping unit for LCM is 0.5ha, so smaller patches of natural environment, especially within urban areas, will not be included. Exploratory analyses indicate this measure to be strongly correlated with the GLUD measure, although with some significant differences.

Based on this classification, approximately 94% of Britain's land cover is 'natural', with 6% built (including domestic gardens).

A similar procedure to that used for the GLUD data was used to calculate the percentage coverage of 'natural environment' for each of the two buffer sizes around participant residence locations (using the GME command 'isectpolyrst').

Mapping and Linkage: Coastal proximity

These indicators require the calculation of a proximity measure, indicating the distance from each UKB participant's residence location to the coast. As per previous research on this topic, there is no clear definition of the coast – i.e. where an estuary becomes coastline – so an arbitrary threshold is used, restricting the coast where an estuary narrows to approximately 1km (Wheeler et al., 2012). The geographical data implemented here are the coastline at mean high water, derived from dissolved statistical area boundaries (2004 Lower-layer Super Output Areas).

Previous work has been based upon crow-flies distance (Wheeler et al., 2012; White et al., 2013a; White et al., 2014). This has indicated associations between coastal proximity, coastal visit frequency and various health and wellbeing outcomes. Road network distance was also investigated.

Crow-flies distance

Typical linear measurement tools (e.g. the 'Near' function in ArcToolbox) to calculate the straight line distance between each participant's home location and the nearest point on the coastline boundary dataset take a very long time for these large datasets. An alternative method was used: a Euclidean distance raster from the coastline was calculated for a small grid cell size, then values from the grid allocated to UKB point locations.

Validation

A Euclidean distance raster to the coastline was calculated for 50m grid cells across GB. These raster values were allocated to Wave 1 point locations using Spatial Analyst>Extract Values to Points tool with interpolation. For a sample of ~2500 points, exact distance was calculated using the Near function, and interpolated grid distance as above. The resulting correlation coefficient for the two methods was ~1.00; mean difference = 12 metres; min diff -24m, max +34m, suggesting negligible error associated with the more efficient raster process.

Road network distance

Options for calculating road network distance to the coast were investigated, but given the time and analytical resources available to the project, the decision was made to make use of the linear/crow flies distance only. Future work with additional resources could investigate road network distances as an alternative proximity metric.

Indices of Deprivation (IoD)

To permit analyses comparable with previous research (Mitchell and Popham, 2007, 2008; Wheeler et al., 2015; Wheeler et al., 2012; White et al., 2014), area deprivation indices for relevant time periods were also linked to participants. Lower-layer Super Output Area (LSOA)/DataZones were allocated to participants' residence locations using ArcGIS, allowing allocation of deprivation indices. Selected index domain indicators were allocated (Income, Employment and Education, along with the combined Index of Multiple Deprivation) separately for residents of each country. The IoD for each country are calculated separately, and they are not directly comparable; this needs to be allowed for in any analyses.

Wales IoD 2008

Source: <https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/Archive/WIMD-2008>

Scotland IoD 2009

Source: <http://www.gov.scot/Topics/Statistics/SIMD/>

England IoD 2010

Source: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>

Data dictionary & citations

Annex 1 provides a data dictionary, including summary statistics for all variables and required citations for derived variables.

References

Agay-Shay, K., Peled, A., Crespo, A.V., Peretz, C., Amitai, Y., Linn, S., Friger, M., Nieuwenhuijsen, M.J., 2014. Green spaces and adverse pregnancy outcomes. *Occupational and Environmental Medicine* 71, 562-569.

Alcock, I., White, M., Wheeler, B., Fleming, L., Depledge, M., 2014. Longitudinal Effects on Mental Health of Moving to Greener and Less Green Urban Areas. *Environmental Science and Technology* 48, 1247-1255.

Department for Communities and Local Government, 2007. Generalised Land Use Database Statistics for England 2005 (Enhanced Basemap)
https://data.gov.uk/dataset/land_use_statistics_generalised_land_use_database. DCLG, London.

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Rojas-Rueda, D., Plasència, A., Nieuwenhuijsen, M.J., 2016. Residential green spaces and mortality: A systematic review. *Environment International* 86, 60-67.

Maas, J., Verheij, R.A., de, V.S., Spreeuwenberg, P., Schellevis, F.G., Groenewegen, P.P., 2009. Morbidity is related to a green living environment. *Journal of Epidemiology and Community Health* 63, 967-973.

Mitchell, R., Popham, F., 2007. Greenspace, urbanity and health: relationships in England. *Journal of Epidemiology and Community Health* 61, 681-683.

Mitchell, R., Popham, F., 2008. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet* 372, 1655-1660.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Simpson, I.C., 2011. Final report for LCM2007 – the new UK land cover map. NERC/Centre for Ecology & Hydrology.

Natural England, 2010. Nature nearby: Accessible Natural Greenspace Guidance. Natural England, Sheffield.

Reklaitiene, R., Grazuleviciene, R., Dedele, A., Virviciute, D., Vensloviene, J., Tamosiunas, A., Baceviciene, M., Luksiene, D., Sapranaviciute-Zabazlajeva, L., Radisauskas, R., Bernotiene, G., Bobak, M., Nieuwenhuijsen, M.J., 2014. The relationship of green space, depressive symptoms and perceived general health in urban population. *Scandinavian Journal of Public Health*.

Wheeler, B., Lovell, R., Higgins, S., White, M., Alcock, I., Osborne, N., Husk, K., Sabel, C., Depledge, M., 2015. Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. *International Journal of Health Geographics* 14, 17.

Wheeler, B.W., White, M., Stahl-Timmins, W., Depledge, M.H., 2012. Does living by the coast improve health and wellbeing? *Health and Place* 18, 1198-1201.

White, M.P., Alcock, I., Wheeler, B.W., Depledge, M.H., 2013a. Coastal proximity, health and well-being: Results from a longitudinal panel survey. *Health and Place* 23, 97-103.

White, M.P., Alcock, I., Wheeler, B.W., Depledge, M.H., 2013b. Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science* 24, 920-928.

White, M.P., Wheeler, B.W., Herbert, S., Alcock, I., Depledge, M.H., 2014. Coastal proximity and physical activity: Is the coast an under-appreciated public health resource? *Preventive Medicine* 69, 135-140.

Annex 1: Data dictionary

Data are provided in Stata format (v14) with appropriate variable labels.

Sources and citations

A: Proximity to coast:

Based on an amended version of coastline derived from:

Office for National Statistics, 2004. Lower-layer Super Output Areas. ONS, London.

<https://data.gov.uk/dataset/lower-layer-super-output-area-isoa-boundaries>

Amended coastline described and first used in the following publication:

Wheeler, B.W., White, M., Stahl-Timmins, W., Depledge, M.H., 2012. Does living by the coast improve health and wellbeing? *Health and Place* 18, 1198-1201. <http://dx.doi.org/10.1016/j.healthplace.2012.06.015>

Original boundary data licence: Open Government Licence: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

B. Generalised Land Use Database 2005:

Citation: Department for Communities and Local Government, 2007. Generalised Land Use Database Statistics for England 2005 (Enhanced Basemap)

<https://data.gov.uk/dataset/land-use-statistics-generalised-land-use-database>. DCLG, London.

Licence: Open Government Licence: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

C. CEH Land Cover Map 2007:

Citation & acknowledgement: Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Simpson, I.C. 2011. Final report for LCM2007 – the new UK land cover map. CS Technical Report No 11/07 NERC/Centre for Ecology & Hydrology 108pp. (CEH project number: C03259). Based upon LCM2007 © NERC (CEH) 2011. Contains Ordnance Survey data © Crown Copyright 2007. © third party licensors.

D. Indices of Deprivation

Wales IoD 2008

Source: <https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/Archive/WIMD-2008>

Scotland IoD 2009

Source: <http://www.gov.scot/Topics/Statistics/SIMD/>

England IoD 2010

Source: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>

Variable descriptions and summary statistics

Variable name	Description	Wave	Source and citation	n	Min	Max	Mean	SD
n_eid	UKB Participant ID							
coastkm_0	Coast distance euclidean, Wave 0, km	0	A	497675	0.00	107.98	41.64	27.71
gs1000_0	% GLUD Greenspace, Wave 0, Buffer 1000m	0	B	440990	4.42	99.19	44.99	21.62
grd1000_0	% GLUD Domestic garden, Wave 0, Buffer 1000m	0	B	440990	0.06	66.14	24.29	11.27
wa1000_0	% GLUD Water, Wave 0, Buffer 1000m	0	B	440990	0.00	59.44	1.26	2.47
gs300_0	% GLUD Greenspace, Wave 0, Buffer 300m	0	B	440990	0.23	99.20	35.27	23.22
grd300_0	% GLUD Domestic garden, Wave 0, Buffer 300m	0	B	440990	0.02	77.37	31.14	14.72
wa300_0	% GLUD Water, Wave 0, Buffer 300m	0	B	440990	0.00	97.88	0.89	2.94
ne1000_0	% LCM Natural Env, Wave 0, Buffer 1000m	0	C	497675	0.00	100.00	40.52	25.60
ne300_0	% LCM Natural Env, Wave 0, Buffer 300m	0	C	497675	0.00	100.00	26.10	25.10
coastkm_1	Coast distance euclidean, Wave 1, km	1	A	20330	0.01	93.65	49.22	17.56
gs1000_1	% GLUD Greenspace, Wave 1, Buffer 1000m	1	B	20327	8.03	98.29	48.24	20.61
grd1000_1	% GLUD Domestic garden, Wave 1, Buffer 1000m	1	B	20327	0.07	54.34	23.44	10.89
wa1000_1	% GLUD Water, Wave 1, Buffer 1000m	1	B	20327	0.00	52.95	1.23	2.06
gs300_1	% GLUD Greenspace, Wave 1, Buffer 300m	1	B	20327	0.87	98.34	38.43	23.32
grd300_1	% GLUD Domestic garden, Wave 1, Buffer 300m	1	B	20327	0.05	72.61	30.62	14.62
wa300_1	% GLUD Water, Wave 1, Buffer 300m	1	B	20327	0.00	73.88	0.94	2.76
ne1000_1	% LCM Natural Env, Wave 1, Buffer 1000m	1	C	20330	0.48	100.00	44.90	24.55
ne300_1	% LCM Natural Env, Wave 1, Buffer 300m	1	C	20330	0.00	100.00	30.49	26.27
wimd08_inc_s_0	Wales IoD 2008 income score	0	D	21283	0.00	99.10	14.59	18.99

Variable name	Description	Wave	Source and citation	n	Min	Max	Mean	SD
wimd08_inc_r_0	Wales loD 2008 income rank, 1=most deprived	0	D	21283	2.00	1894.00	1239.57	579.89
wimd08_emp_s_0	Wales loD 2008 employment score	0	D	21283	0.00	99.10	13.14	16.26
wimd08_emp_r_0	Wales loD 2008 employment rank, 1=most deprived	0	D	21283	2.00	1895.00	1256.46	539.16
wimd08_edu_s_0	Wales loD 2008 education score	0	D	21283	0.00	100.00	14.19	17.69
wimd08_edu_r_0	Wales loD 2008 education rank, 1=most deprived	0	D	21283	1.00	1896.00	1239.40	581.35
wimd08_s_0	Wales loD combined index 2008, score higher=more deprived	0	D	21283	2.20	77.20	15.26	13.29
wimd08_r_0	Wales loD combined index 2008, rank 1=most deprived	0	D	21283	4.00	1896.00	1272.06	578.02
simd09_inc_s_0	Scotland loD 2009 income score	0	D	35402	0.00	70.00	11.14	10.41
simd09_inc_r_0	Scotland loD 2009 income rank, 1=most deprived	0	D	35402	2.00	6504.00	4138.17	1902.55
simd09_emp_s_0	Scotland loD 2009 employment score	0	D	35402	0.00	59.00	8.67	7.50
simd09_emp_r_0	Scotland loD 2009 employment rank, 1=most deprived	0	D	35402	1.00	6503.00	4199.89	1877.99
simd09_edu_s_0	Scotland loD 2009 education score	0	D	35402	-2.89	2.73	-0.59	1.07
simd09_edu_r_0	Scotland loD 2009 education rank, 1=most deprived	0	D	35402	2.00	6505.00	4309.78	1933.69
simd09_s_0	Scotland loD combined index 2009, score higher=more deprived	0	D	35402	0.68	88.24	14.71	15.16
simd09_r_0	Scotland loD combined index 2009, rank 1=most deprived	0	D	35402	3.00	6504.00	4334.26	1951.20
eimd10_inc_s_0	England loD 2010 income score	0	D	440990	0.01	0.77	0.12	0.10
eimd10_inc_r_0	England loD 2010 income rank, 1=most deprived	0	D	440990	1.00	32478.00	19357.72	9365.80
eimd10_emp_s_0	England loD 2010 employment score	0	D	440990	0.00	0.75	0.09	0.06
eimd10_emp_r_0	England loD 2010 employment rank, 1=most deprived	0	D	440990	1.00	32478.00	18224.40	9323.93
eimd10_edu_s_0	England loD 2010 education score	0	D	440990	0.02	98.09	15.76	16.35
eimd10_edu_r_0	England loD 2010 education rank, 1=most deprived	0	D	440990	6.00	32480.00	19865.42	9377.88
eimd10_s_0	England loD combined index 2010, score higher=more deprived	0	D	440990	0.61	83.09	17.89	14.20
eimd10_r_0	England loD combined index 2010, rank 1=most deprived	0	D	440990	4.00	32481.00	18868.63	9314.73
wimd08_inc_s_1	Wales loD 2008 income score	1	D	1	12.20	12.20	12.20	
wimd08_inc_r_1	Wales loD 2008 income rank, 1=most deprived	1	D	1	1108.00	1108.00	1108.00	
wimd08_emp_s_1	Wales loD 2008 employment score	1	D	1	11.20	11.20	11.20	
wimd08_emp_r_1	Wales loD 2008 employment rank, 1=most deprived	1	D	1	1159.00	1159.00	1159.00	

Variable name	Description	Wave	Source and citation	n	Min	Max	Mean	SD
wimd08_edu_s_1	Wales loD 2008 education score	1	D	1	22.10	22.10	22.10	
wimd08_edu_r_1	Wales loD 2008 education rank, 1=most deprived	1	D	1	710.00	710.00	710.00	
wimd08_s_1	Wales loD combined index 2008, score higher=more deprived	1	D	1	18.40	18.40	18.40	
wimd08_r_1	Wales loD combined index 2008, rank 1=most deprived	1	D	1	919.00	919.00	919.00	
simd09_inc_s_1	Scotland loD 2009 income score	1	D	2	8.00	11.00	9.50	2.12
simd09_inc_r_1	Scotland loD 2009 income rank, 1=most deprived	1	D	2	3683.00	4565.00	4124.00	623.67
simd09_emp_s_1	Scotland loD 2009 employment score	1	D	2	10.00	11.00	10.50	0.71
simd09_emp_r_1	Scotland loD 2009 employment rank, 1=most deprived	1	D	2	3063.00	3114.00	3088.50	36.06
simd09_edu_s_1	Scotland loD 2009 education score	1	D	2	-0.35	0.04	-0.16	0.28
simd09_edu_r_1	Scotland loD 2009 education rank, 1=most deprived	1	D	2	3133.00	4021.00	3577.00	627.91
simd09_s_1	Scotland loD combined index 2009, score higher=more deprived	1	D	2	14.94	15.49	15.21	0.39
simd09_r_1	Scotland loD combined index 2009, rank 1=most deprived	1	D	2	3517.00	3640.00	3578.50	86.97
eimd10_inc_s_1	England loD 2010 income score	1	D	20327	0.01	0.77	0.10	0.09
eimd10_inc_r_1	England loD 2010 income rank, 1=most deprived	1	D	20327	1.00	32476.00	20573.48	8860.50
eimd10_emp_s_1	England loD 2010 employment score	1	D	20327	0.01	0.75	0.09	0.06
eimd10_emp_r_1	England loD 2010 employment rank, 1=most deprived	1	D	20327	1.00	32473.00	17704.52	8589.78
eimd10_edu_s_1	England loD 2010 education score	1	D	20327	0.16	96.19	13.59	14.60
eimd10_edu_r_1	England loD 2010 education rank, 1=most deprived	1	D	20327	15.00	32442.00	21146.78	9017.51
eimd10_s_1	England loD combined index 2010, score higher=more deprived	1	D	20327	1.12	81.59	16.96	13.40
eimd10_r_1	England loD combined index 2010, rank 1=most deprived	1	D	20327	9.00	32461.00	19410.46	8890.64