Supplementary materials

Associations of residential greenness with bone mineral density and osteoporosis:

The modifying effect of genetic susceptibility

Text S1 Assessment of PM_{2.5} and NO₂

Text S2 Construction of neuroticism score

Text S3 Methods of causal mediation analyses

Text S4 Methods of multiple imputation

Text S5 Assessment of land use greenspace indicators

Text S6 Assignment of time-varying greenness exposure.

Table S1 Demographic comparison of study population and UK Biobank full sample.

Table S2 Summary distribution of greenness exposures.

Table S3 Sensitivity analyses: Employing multiple imputation method to repair missing data on covariates.

Table S4 Sensitivity analyses: Excluding participants with osteopenia (T score less than -1) at baseline.

Table S5 Sensitivity analyses: Excluding participants with incident OP at first 2 years of follow-up.

Table S6 Sensitivity analyses: Restricting analyses to those who have lived in the current address for at least five years.

Table S7 Sensitivity analyses: Employing NDVI within different buffers as greenness indicators.

Table S8 Sensitivity analyses: Employing land use indicators as proxies for greenness exposure.

Table S9 Sensitivity analyses: Employing greenness exposure as time-varying variables in the models.

Table S10 Sensitivity analyses: Further adjusted for medical conditions.

Table S11 Sensitivity analyses: Further adjusted for vitamin/mineral supplements intake.

Table S12 Sensitivity analyses: Further adjusted for time spend outdoors in summer and winter.

Table S13 Sensitivity analyses: Only considering diagnosis with medical records (N=379,423, case=8,994).

Table S14 Cox regression for the associations between PRS and incident OP.

Table S15 Cox regression for the associations between residential greenness and incident OP, separately for PRS categories.

The annual average level of $PM_{2.5}$ and NO_2 were modelled using participants' residential address using Land Use Regression (LUR) models developed as part of the European Study of Cohorts for Air Pollution Effects (ESCAPE) (http://www.escapeproject.eu/), funded under the EU 7th Framework Program. Traffic variables were calculated within a geographic information system (GIS) during the course of LUR. The key predictive variables included traffic intensity, population, and land-use. Leave-one-out cross-validation indicated good model performance (cross-validation $R^2 =$ 77% and 87%, respectively for $PM_{2.5}$ and NO_2).

Text S2 Construction of neuroticism score

Neuroticism score was derived from the Eysenck Personality Questionnaire-Revised Short Form. 12 items reflecting neurotic behavior were used, comprising the following questions: (1) "Does your mood often go up and down?", (2) "Do you ever feel 'just miserable' for no reason?", (3) "Are you an irritable person?", (4) "Are your feelings easily hurt?", (5) "Do you often feel 'fed-up?" (6) "Would you call yourself a nervous person?" (7) "Are you a worrier?", (8) "Would you call yourself tense or 'highly strung'?", (9) "Do you worry too long after an embarrassing experience?", (10) "Do you suffer from 'nerves'?", (11) "Do you often feel lonely?", (12) "Are you often troubled by feelings of guilt?" (1 point for yes). The neuroticism score ranged from 0 to 12 for each participant.

Text S3 Methods of causal mediation analyses

The mediation analyses consist which consists of three steps to generate the result. First, IPF incidence was regressed by greenness exposures, potential mediators and confounders in cox models.

Afterwards, potential mediators were regressed by greenness variables in either logistic (for binary mediators) or multiple linear (for continuous mediators) models adjusting for confounders. The two models were then combined to compute the natural direct effect (NDE) and natural indirect effect (NIE). Mediation proportion was calculated as NIE/Total Effect (Valeri & Vanderweele, 2013, 2015). We assumed that age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, and alcohol consumption to be the common set of confounders for exposure-outcome, exposure-mediator, and mediator-outcome associations. To ensure that mediation analyses yield robust results, there must be no unmeasured confounding and the model specification must be correct. Therefore, the results should be viewed with caution.

Text S4 Methods of multiple imputation

The missing data pattern suggests that observations were missing at random. We imputed missing categorical variables with fully conditional specification (FCS) methods. We imputed five complete datasets in order to reduce the computational burden of analyzing UKB data. Discriminant function method was conducted to repair missing variables. The number of burn-in iterations was set to be 20 and the convergence was confirmed. Each of the 5 imputed datasets was analyzed independently, and then the results were pooled using the standard multiple imputation inference method (Rubin's rule) to account for the uncertainty due to missing values.

Text S5 Assessment of land use greenspace indicators

The land use data was obtained from the 2005 Generalized Land Use Database (GLUD) for England, provided by the Department for Communities and Local Government of the Government of the UK (https://www.gov.uk/government/statistics). The GLUD includes information on land use

distribution at the 2001 Census Output Areas level and was consistent with previous studies. Each home location polygon was allocated an areaweighted mean of the land use percentage coverage. Greenspace indicators were calculated as proportion of all land-use types at 300 m buffer intersecting each participant's residence.

Text S6 Assignment of time-varying greenness exposure.

When analyzing the Cox model with a time-varying variable, each observation was divided into multiple records by year (from wide data to long data). To construct the 1-year time-varying variable, the greenness exposure of each record was assigned based on the average concentration of that year (from the same data source as the main analyses). For example, participant No.X was followed up from June 14, 2006 to June 31, 2021. Table A1 presents the wide data of participant No.X, which is transformed to long data of 16 records by year, as shown in Table A2. The long data were used to calculate the time-varying Cox model.

Douticinents' ID	Participants' ID Start time	End time	Average concentration of air pollutants, $\mu g/m^3$			utants, $\mu g/m^3$
Participants ID		End time –	2006	2007		2021
No.X	2006-06-14	2021-06-31	C1	C2		C16

Table A1. Wide data of included participants (Example).

Note: C1, C2 and C3 denote the air pollutant concentrations for the corresponding years.

Table A2. Long data of included participants (Example	e).
---	-----

No. of records	Participants' ID	Start time	End time	concentration of air pollutants, $\mu g/m^3$
1	No.X	2006-04-10	2006-12-31	C1
2	No.X	2006-12-31	2007-12-31	C2
3	No.X	2007-12-31	2008-12-31	C3

16	No.X	2019-12-31	2020-12-31	C16
17	No.X	2020-12-31	2021-06-31	C17

	Study population	Full sample
Variables	N=391,298	N=502,482
Age (years), mean (SD)	56.2 (8.1)	56.5 (8.1)
Gender, n (%)		
Female	207299 (53.0%)	273366 (54.4%)
Ethnicity, n (%)		
White ethnicity	358626 (91.7%)	453651 (90.3%)
Mixed ethnicity	13732 (3.5%)	18862 (3.8%)
Asian ethnicity	13383 (3.4%)	17521 (3.5%)
Black ethnicity	1777 (0.5%)	2848 (0.6%)
Chinese ethnicity	1075 (0.3%)	1574 (0.3%)
Other ethnicity	2705 (0.7%)	4558 (0.9%)
Education level, n (%)		
College or university degree	138716 (35.5%)	161152 (32.1%)
A levels/AS levels	46215 (11.8%)	55320 (11.0%)
O levels/GCSEs	83923 (21.4%)	105192 (20.9%)
CSEs	20942 (5.4%)	26887 (5.4%)
NVQ or HND or HNC	25655 (6.6%)	32726 (6.5%)
Other qualifications	19981 (5.1%)	25802 (5.1%)
None of these above	138716 (35.5%)	85268 (17.0%)
Annual household income, n (%)		
< £18,000	85104 (21.7%)	97193 (19.3%)
£18,000 to 30,999	99321 (25.4%)	108173 (21.5%)
£31,000 to 51,999	103538 (26.5%)	110769 (22.0%)
£52,000 to 100,000	81652 (20.9%)	86258 (17.2%)
\geq £100,000	21683 (5.5%)	22928 (4.6%)
Employment status, n (%)		
Employed	239333 (61.2%)	287131 (57.1%)
Retired	123128 (31.5%)	166978 (33.2%)
Unemployed	28837 (7.4%)	42619 (8.5%)
Residential area, n (%)		
Urban	308469 (78.8%)	9481 (79.8%)
Smoking status, n (%)		
Never	213830 (54.6%)	273507 (54.4%)
Former	137326 (35.1%)	173050 (34.4%)
Current	40142 (10.3%)	52974 (10.5%)
Alcohol consumption, n (%)		
Never	14309 (3.7%)	22384 (4.5%)
Former	13068 (3.3%)	18102 (3.6%)
Current	363921 (93.0%)	460341 (91.6%)
Healthy diet score, n (%)		
0	3584 (0.9%)	4277 (0.9%)
1	17341 (4.4%)	20718 (4.1%)

T 11 91 D		C 1 1		
Table SI Demogra	phic comparison	of study population	ion and UK B	siobank full sample.

Variables	Study population	Full sample
variables	N=391,298	N=502,482
2	43237 (11.0%)	52018 (10.4%)
3	75402 (19.3%)	91134 (18.1%)
4	98950 (25.3%)	120141 (23.9%)
5	94034 (24.0%)	114737 (22.8%)
6	51893 (13.3%)	63187 (12.6%)
7	6857 (1.8%)	8269 (1.6%)
$PM_{2.5} (\mu g/m^3)$, mean (SD)	10.0 (1.1)	10.0 (1.1)
Physical activity, n (%)		
Not regular	101822 (26.9%)	126677 (25.2%)
Neuroticism score, mean (SD)	4.1 (3.2)	4.1(3.3)
NDVI _{300m} , mean (SD)	0.572 (0.108)	0.570 (0.108)
Heel eBMD (g/cm ²), mean (SD)	0.547 (0.139)	0.545 (0.140)
BMD T-score, mean (SD)	-0.295 (1.246)	-0.314 (1.263)

Abbreviations: OP, osteoporosis; SD, standard deviation; GCSE, general certificate of secondary education; CSE, certificate of secondary education; NVQ, national vocational qualification; HND, higher national diploma; HNC, higher national certificate. $PM_{2.5}$, particular matter with aerodynamic diameter ≤ 2.5 mm; NDVI_{300m}, normalized difference vegetation index within 300m buffer; eBMD, estimated bone mineral density.

Continues variables are displayed as means (SD), and categorical variables are displayed as numbers (percentages).

Variables	Median ± IQR	Minimum	P25	P75	Maximum
NDVI _{300m}	0.57 ± 0.13	0.01	0.51	0.64	0.86

T-1-1-62 6

Abbreviations: IQR, interquartile range; P25, P75 represent the 25th, and 75th percentiles; NDVI_{300m}, normalized difference vegetation index within 300m buffer.

2,329,308, ease 02,818).	
Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.95 (0.94, 0.96)
Quartile 1	Ref.
Quartile 2	0.93 (0.91, 0.95)
Quartile 3	0.92 (0.90, 0.95)
Quartile 4	0.91 (0.89, 0.93)
<i>P</i> for trend	<0.001

Table S3 Sensitivity analyses: Employing multiple imputation method to repair missing data on covariates (N=2,329,305, case=62,545).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

,527, Case=5,949).	
Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.95 (0.91, 0.99)
Quartile 1	Ref.
Quartile 2	0.97 (0.89, 1.06)
Quartile 3	0.92 (0.84, 1.00)
Quartile 4	0.91 (0.83, 1.00)
<i>P</i> for trend	0.02

Table S4 Sensitivity analyses: Excluding participants with osteopenia (T score less than -1) at baseline (N=276,527, case=3,949).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.95 (0.92, 0.98)
Quartile 1	Ref.
Quartile 2	0.93 (0.88, 0.99)
Quartile 3	0.92 (0.86, 0.97)
Quartile 4	0.91 (0.85, 0.97)
<i>P</i> for trend	<0.001

Table S5 Sensitivity analyses: Excluding participants with incident OP at first 2 years of follow-up (N=376,969, case=8,473).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

(1N-321,043, Case-8,093).	
Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.96 (0.93, 0.99)
Quartile 1	Ref.
Quartile 2	0.96 (0.90, 1.02)
Quartile 3	0.92 (0.86, 0.98)
Quartile 4	0.92 (0.87, 0.99)
<i>P</i> for trend	0.01

Table S6 Sensitivity analyses: Restricting analyses to those who have lived in the current address for at least five years (N=321,643, case=8,095).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Variable	HR (95%CI)
NDVI _{500m} , per IQR increment	0.95 (0.93, 0.98)
Quartile 1	Ref.
Quartile 2	0.95 (0.90, 1.01)
Quartile 3	0.92 (0.87, 0.98)
Quartile 4	0.92 (0.87, 0.98)
<i>P</i> for trend	<0.001
NDVI1000m, per IQR increment	0.95 (0.92, 0.97)
Quartile 1	Ref.
Quartile 2	0.93 (0.88, 0.99)
Quartile 3	0.89 (0.84, 0.95)
Quartile 4	0.92 (0.86, 0.97)
<i>P</i> for trend	<0.001
NDVI _{1500m} , per IQR increment	0.94 (0.92, 0.97)
Quartile 1	Ref.
Quartile 2	0.93 (0.88, 0.98)
Quartile 3	0.90 (0.85, 0.96)
Quartile 4	0.91 (0.86, 0.96)
<i>P</i> for trend	<0.001

Table S7 Sensitivity analyses: Employing NDVI within different buffers as greenness indicators (N=379,423, case=9,307).

Abbreviations: NDVI_{500m}, normalized difference vegetation index within 500m buffer; NDVI_{1000m}, normalized difference vegetation index within 1000m buffer; NDVI_{1500m}, normalized difference vegetation index within 1500m buffer HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, and healthy diet score.

Table S8 Sensitivity analyses: Employing land use indicators as proxies for greenness exposure (N=379,423, case=9,307).

Variable	HR (95%CI)
Greenspace percentage within 300m buffer, per IQR increment	0.95 (0.91, 0.98)
Domestic garden percentage within 300m buffer, per IQR increment	0.97 (0.93, 1.00)
Natural environment percentage within 300m buffer, per IQR increment	0.96 (0.93, 0.99)

Abbreviations: HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference. Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, and healthy diet score. Ranges for quartile NDVI_{300m}: quartile1, 0.01 to 0.51; quartile 2, 0.51 to 0.57; quartile 3, 0.57 to 0.64; quartile 4: 0.64 to 0.86.

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.96 (0.93, 0.98)
Quartile 1	Ref.
Quartile 2	0.94 (0.89, 1.00)
Quartile 3	0.92 (0.87, 0.98)
Quartile 4	0.92 (0.87, 0.98)
<i>P</i> for trend	<0.001

Table S9 Sensitivity analyses: Employing greenness exposure as time-varying variables in the models.

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.95 (0.93, 0.97)
Quartile 1	Ref.
Quartile 2	0.94 (0.89, 0.99)
Quartile 3	0.91 (0.86, 0.97)
Quartile 4	0.91 (0.86, 0.96)
<i>P</i> for trend	<0.001

Table S10 Sensitivity analyses: Further adjusted for medical conditions (N=377,868, case=9,251).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference; .

Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, vascular/heart problems (including heart attack, angina, stroke, hypertension), diabetes, and BMI. Ranges for quartile NDVI_{300m}: quartile1, 0.01 to 0.51; quartile 2, 0.51 to 0.57; quartile 3, 0.57 to 0.64; quartile 4: 0.64 to 0.86.

Table S11 Sensitivity analyses:	Further adjusted for vitan	nin/mineral supplements	intake (N=378,264,
case=9,274).			

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.96 (0.94, 0.99)
Quartile 1	Ref.
Quartile 2	0.95 (0.90, 1.01)
Quartile 3	0.93 (0.88, 0.98)
Quartile 4	0.93 (0.88, 0.99)
<i>P</i> for trend	0.01

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, vitamin supplements and mineral supplements.

Ranges for quartile NDVI_{300m}: quartile1, 0.01 to 0.51; quartile 2, 0.51 to 0.57; quartile 3, 0.57 to 0.64; quartile 4: 0.64 to 0.86.

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.96 (0.94, 0.99)
Quartile 1	Ref.
Quartile 2	0.96 (0.90, 1.02)
Quartile 3	0.92 (0.87, 0.98)
Quartile 4	0.93 (0.88, 0.99)
<i>P</i> for trend	0.02

Table S12 Sensitivity analyses: Further adjusted for time spend outdoors in summer and winter (N=362,187, case=8,715).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, time spend outdoors in summer and time spend outdoors in winter.

Ranges for quartile NDVI_{300m}: quartile1, 0.01 to 0.51; quartile 2, 0.51 to 0.57; quartile 3, 0.57 to 0.64; quartile 4: 0.64 to 0.86.

Variable	HR (95%CI)
NDVI _{300m} , per IQR increment	0.95 (0.93, 0.98)
Quartile 1	Ref.
Quartile 2	0.94 (0.89, 0.99)
Quartile 3	0.91 (0.86, 0.97)
Quartile 4	0.91 (0.86, 0.97)
<i>P</i> for trend	0.001

Table S13 Sensitivity analyses: Only considering diagnosis with medical records (N=379,423, case=8,994).

Abbreviations: NDVI_{300m}, normalized difference vegetation index within 300m buffer; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

P value for trend calculated treating the telomere length concentrations (quartile) as a continuous variable.

Models were adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, time spend outdoors in summer and time spend outdoors in winter.

Ranges for quartile NDVI_{300m}: quartile1, 0.01 to 0.51; quartile 2, 0.51 to 0.57; quartile 3, 0.57 to 0.64; quartile 4: 0.64 to 0.86.

Variable	Case (N	HR (95%CI)	
variable		Model 1	Model 2
PRS, continuous	8,344/340,119	1.42 (1.39, 1.45)	1.45 (1.42, 1.49)
Low genetic risk	1,835/113,372	Ref.	Ref.
Medium genetic risk	2,676/113,374	1.46 (1.38, 1.55)	1.48 (1.39, 1.57)
High genetic risk	3,833/113,373	2.11 (1.99, 2.23)	2.17 (2.05, 2.30)
<i>P</i> for trend		<0.001	<0.001

Table S14 Associations between PRS and incident OP.

Abbreviations: PRS, polygenic risk score; OP, osteoporosis; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Model 1: Unadjusted.

Model 2: Adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, , the first 10 genetic principal components, and genotyping batch.

Ranges for tertile PRS: low, -8.72 to -0.46; medium, -0.46 to 0.34; high, 0.34 to 4.72.

Variable	Case/N	HR (95%CI)
Low genetic risk		
NDVI _{300m} , per IQR increment	1,835/113,372	0.61 (0.39, 0.96)
Quartile 1	1,018 /28,596	Ref.
Quartile 2	940 /28,224	0.92 (0.81, 1.04)
Quartile 3	953 /28,321	0.84 (0.74, 0.96)
Quartile 4	922 /28,232	0.94 (0.82, 1.07)
Medium genetic risk		
NDVI _{300m} , per IQR increment	2,676/113,374	0.71 (0.49, 1.03)
Quartile 1	692 /28,278	Ref.
Quartile 2	688 /28,459	0.94 (0.85, 1.05)
Quartile 3	675 /28,409	0.95 (0.85, 1.06)
Quartile 4	621 /28,228	0.89 (0.79, 0.99)
High genetic risk		
NDVI _{300m} , per IQR increment	3,833/113,373	0.77 (0.57, 1.06)
Quartile 1	1,018 /28,596	Ref.
Quartile 2	940 /28,224	0.92 (0.85, 1.01)
Quartile 3	953 /28,321	0.94 (0.86, 1.02)
Quartile 4	922 /28,232	0.91 (0.83, 1.00)

Table S15 Associations between residential greenness and incident OP, separately for PRS
categories.

Abbreviations: PRS, polygenic risk score; OP, osteoporosis; HR, hazards ratio; CI, confidence interval; IQR, interquartile range; ref., reference.

Model 1: Unadjusted.

Model 2: Adjusted for age, gender, ethnicity, annual household income, education level, employment status, residential area, smoking status, alcohol consumption, healthy diet score, the first 10 genetic principal components, and genotyping batch.

Ranges for tertile PRS: low, -8.72 to -0.46; medium, -0.46 to 0.34; high, 0.34 to 4.72.

Ranges for quartile NDVI_{300m}: quartile 1, 0.01 to 0.51; quartile 2, 0.51 to 0.58; quartile 3, 0.58 to 0.65; quartile 4: 0.65 to 0.86.