**Supplemental Materials**

**Association between residential greenness and metabolic dysfunction-associated fatty liver disease: evidence from a large population-based epidemiological study.**

**Table of Contents**

**Supplemental Methods:** Causal mediation methods

**Table S1.** The sensitivity parameter for combination of exposure and mediator

**Table S2.** Distributions for NDVI, EVI with different buffers and exposure window.

Table S3. Associations between per IQR increase in one-year and two -year average value of greenness metrics within different buffers and MAFLD prevalence.

Table S4. Association between greenness metrics (per IQR increase) within different buffers and MAFLD after additional adjusting history of diseases.

Table S5. Association between quartile greenness and MAFLD

**Causal mediation methods**

What we used in mediation analysis is regression-based causal mediation[1], instead of traditional analysis like product method or difference method. And compare to traditional analysis, causal mediation analysis has following advantages[2]: 1) causal mediation proposed unambiguous assumptions to identify the mediating effects. 2) The results of causal mediation analysis are independent of types of outcome variable. When outcome is binary variable and is a common event, the results of difference method and product method are differing. 3) The causal mediation analysis allows for the model including the interaction between exposure and mediator, nonlinearity.

In this study, first, we construct mediation model and outcome model as follows:

mediation model:

outcome model:

In these equations, X is observed exposure (i.e., residential greenness), E(M) represents the average level of observed mediator, Ck denotes set of observed confounding variables and Y is a binary variable, Y = 1 indicate individual have been observed MALFD. The mediating effect was calculated by following steps: (a)simulate model parameters from their sampling distribution. (b) simulate the potential values of the mediator, (c) simulate the potential outcomes given the simulated values of the mediator, (d) compute the causal mediation effects.

For identifying the causal mediation effects, we have following assumption:

assumption 1: no unmeasured confounding of the treatment–outcome relationship

assumption 2: no unmeasured confounding of the mediator–outcome relationship.

assumption 3: no unmeasured confounding of the treatment–mediator relationship.

assumption 4: there is no mediator–outcome confounder that is affected by the exposure.

For controlling for [assumption 1] confounding of the exposure (i.e., greenness)–outcome (i.e., MAFLD) relationship, we should adjust for common causes of the exposure and the outcome such as age, gender. And for satisfy the assumption 2, we should adjust for the common cause between exposure (i.e., greenness) and mediator (i.e., air pollution and physical activity) such as region, areas. And for satisfy the assumption 3, we should adjust for the common cause of mediator (i.e., air pollution and physical activity) and outcome (i.e., MAFLD) such as age, gender, household income. Actually, we need not to distinguish in this regression approach the treatment–outcome and the mediator–outcome confounding variables and include all three types of confounding variables [2]. Assumption 4 is known as a cross-world independence assumption, because it requires that conditional on the confounding variables of mediator and outcome, the mediator under the exposure level equal x\* is independent of the outcome under the exposure level equal x (x, x\* are in cross-world).

In order to interrogate the assumptions, we have calculated a sensitivity parameter indicating the minimum strength of association on difference scale that an unmeasured confounder would need to have with both the exposure and the outcome, conditional on the measured covariates, could to diminish the nature indirect effects to zero [2 3]. Results shown in Table S1. Table S1 shows that, for example, to diminish mediating effects of PM1 between NDVI500m and MAFLD, an unmeasured confounder would be need which have a risk difference of 0.125 in association with MAFLD and exposure conditional on observed confounder and mediator.

**Table S1.** The sensitivity parameter for combination of exposure and mediator

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NDVI500m | | EVI500m | |
| ACME | sensitivity parameter | ACME | sensitivity parameter |
| PM1 | -0.0155 | 0.125 | -0.024 | 0.155 |
| PM2.5 | -0.0323 | 0.180 | -0.043 | 0.206 |
| PM10 | -0.0243 | 0.156 | -0.033 | 0.182 |
| Physical activity | -0.0143 | 0.120 | -0.018 | 0.135 |

Abbreviation: ACME: average causal mediating effect in difference scale

# Table S2 Distributions for NDVI, EVI with different buffers and exposure window.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min. | P25 | Median | Mean | P75 | Max. | IQR |
| **One-year average** |  |  |  |  |  |  |  |
| NDVI250m | -0.09 | 0.3 | 0.41 | 0.399469 | 0.5 | 0.75 | 0.2 |
| NDVI500m | -0.09 | 0.3 | 0.42 | 0.403642 | 0.5 | 0.73 | 0.2 |
| NDVI1000m | -0.02 | 0.3 | 0.44 | 0.416112 | 0.51 | 0.71 | 0.21 |
| EVI250m | -0.04 | 0.17 | 0.24 | 0.236963 | 0.3 | 0.45 | 0.13 |
| EVI500m | -0.02 | 0.17 | 0.25 | 0.239708 | 0.3 | 0.45 | 0.13 |
| EVI1000m | -0.02 | 0.18 | 0.27 | 0.247274 | 0.31 | 0.43 | 0.13 |
| **Two-year average** |  |  |  |  |  |  |  |
| NDVI250m | -0.08 | 0.29 | 0.42 | 0.397718 | 0.5 | 0.73 | 0.21 |
| NDVI500m | -0.09 | 0.29 | 0.42 | 0.401516 | 0.5 | 0.71 | 0.21 |
| NDVI1000m | -0.01 | 0.3 | 0.44 | 0.414397 | 0.51 | 0.71 | 0.21 |
| EVI250m | -0.03 | 0.17 | 0.25 | 0.238218 | 0.3 | 0.49 | 0.13 |
| EVI500m | -0.03 | 0.17 | 0.25 | 0.240673 | 0.3 | 0.45 | 0.13 |
| EVI1000m | -0.02 | 0.18 | 0.26 | 0.248662 | 0.31 | 0.42 | 0.13 |
| **Three-year average** |  |  |  |  |  |  |  |
| NDVI250m | -0.08 | 0.28 | 0.42 | 0.399111 | 0.5 | 0.74 | 0.22 |
| NDVI500m | -0.09 | 0.29 | 0.42 | 0.403503 | 0.5 | 0.71 | 0.21 |
| NDVI1000m | -0.01 | 0.3 | 0.44 | 0.416092 | 0.52 | 0.72 | 0.22 |
| EVI250m | -0.04 | 0.17 | 0.25 | 0.241064 | 0.31 | 0.49 | 0.14 |
| EVI500m | -0.03 | 0.17 | 0.26 | 0.244154 | 0.31 | 0.46 | 0.14 |
| EVI1000m | -0.02 | 0.18 | 0.27 | 0.251988 | 0.32 | 0.42 | 0.14 |

Abbreviation: NDVI250m, NDVI500m, NDVI1000m, Normalized Difference Vegetation Index within 250m, 500m, 1000m buffers, respectively; EVI250m, EVI500m, EVI1000m, Enhanced Vegetation Index within 250m, 500m, 1000m; OR, odds ratio. P25: 25th percentile. P75: 75th percentile

# Table S3 Associations between per IQR increase in one-year and two -year average value of greenness metrics within different buffers and MAFLD prevalence.

|  |  |  |
| --- | --- | --- |
| Greenness metrics | Crude ORa（95% CI） | Adjusted ORb（95% CI） |
| **One-year average** |  |  |
| NDVI250m | 0.70 (0.68, 0.73) | 0.75 (0.72, 0.78) |
| NDVI500m | 0.71 (0.69, 0.73) | 0.76 (0.74, 0.79) |
| NDVI1000m | 0.68 (0.66, 0.71) | 0.73 (0.70, 0.76) |
| EVI250m | 0.72 (0.70, 0.74) | 0.79 (0.76, 0.81) |
| EVI500m | 0.72 (0.70, 0.74) | 0.79 (0.76, 0.82) |
| EVI1000m | 0.70 (0.68, 0.72) | 0.76 (0.73, 0.79) |
| **Two-year average** |  |  |
| NDVI250m | 0.69 (0.67, 0.71) | 0.74 (0.71, 0.76) |
| NDVI500m | 0.69 (0.67, 0.72) | 0.75 (0.72, 0.77) |
| NDVI1000m | 0.68 (0.66, 0.71) | 0.73 (0.70, 0.75) |
| EVI250m | 0.72 (0.70, 0.74) | 0.78 (0.75, 0.81) |
| EVI500m | 0.71 (0.69, 0.74) | 0.78 (0.75, 0.81) |
| EVI1000m | 0.70 (0.68, 0.72) | 0.76 (0.73, 0.79) |

a No adjustment.

b Adjusted for age, male, ethnicity, region, areas, annually household income, education, smoke status, drinking status, high fat diet.

Abbreviation: NDVI250m, NDVI500m, NDVI1000m, Normalized Difference Vegetation Index within 250m, 500m, 1000m buffers, respectively; EVI250m, EVI500m, EVI1000m, Enhanced Vegetation Index within 250m, 500m, 1000m; OR, odds ratio.

# Table S4 Association between greenness metrics (per IQR increase) within different buffers and MAFLD after additional adjusting history of diseases.

|  |  |
| --- | --- |
| Greenness metrics | Adjusted OR |
| Main analyses |  |
| NDVI250m | 0.76 (0.73, 0.79) |
| NDVI500m | 0.78 (0.75, 0.81) |
| NDVI1000m | 0.76 (0.73, 0.79) |
| EVI250m | 0.80 (0.77, 0.83) |
| EVI500m | 0.81 (0.78, 0.84) |
| EVI1000m | 0.79 (0.76, 0.83) |
| Additional adjusting history of diseases |  |
| NDVI250m | 0.77 (0.74, 0.8) |
| NDVI500m | 0.79 (0.76, 0.82) |
| NDVI1000m | 0.76 (0.73, 0.79) |
| EVI250m | 0.81 (0.78, 0.84) |
| EVI500m | 0.81 (0.78, 0.85) |
| EVI1000m | 0.80 (0.77, 0.83) |

Abbreviation: NDVI250m, NDVI500m, NDVI1000m, Normalized Difference Vegetation Index within 250m, 500m, 1000m buffers, respectively; EVI250m, EVI500m, EVI1000m, Enhanced Vegetation Index within 250m, 500m, 1000m; OR, odds ratio.

# Table S5 Association between quartile greenness and MAFLD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Greenness metrics | | Crude ORa（95% CI） | P value | Adjusted ORb（95% CI） | P value |
| NDVI500m | |  |  |  |  |
| Q1[-0.09,0.29] | | Reference | - | Reference | - |
| Q2(0.29,0.42] | | 0.80 (0.76, 0.84) | <0.001 | 0.86 (0.81, 0.90) | <0.001 |
| Q3(0.42,0.5] | | 0.70 (0.67, 0.74) | <0.001 | 0.81 (0.76, 0.86) | <0.001 |
| Q4(0.5,0.71] | | 0.58 (0.55, 0.61) | <0.001 | 0.69 (0.65, 0.73) | <0.001 |
| EVI500m |  |  |  |  |
| Q1[-0.03,0.17] | Reference | - | Reference | - |
| Q2(0.17,0.26] | 0.81 (0.78, 0.85) | <0.001 | 0.90 (0.85, 0.95) | <0.001 |
| Q3(0.26,0.31] | 0.69 (0.65, 0.72) | <0.001 | 0.82 (0.77, 0.87) | <0.001 |
| Q4(0.31,0.46] | 0.60 (0.56, 0.63) | <0.001 | 0.74 (0.69, 0.79) | <0.001 |

a No adjustment.

b Adjusted for age, male, ethnicity, region, areas, annually household income, education, smoke status, drinking status, high fat diet.

Abbreviation: NDVI250m, NDVI500m, NDVI1000m, Normalized Difference Vegetation Index within 250m, 500m, 1000m buffers, respectively; EVI250m, EVI500m, EVI1000m, Enhanced Vegetation Index within 250m, 500m, 1000m; OR, odds ratio. Q1: the first quartile. Q2: the second quartile. Q3: the third quartile. Q4: the fourth quartile.

1. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods* 2010;15(4):309-34. doi: 10.1037/a0020761 [published Online First: 2010/10/20]

2. VanderWeele T. Explanation in causal inference: methods for mediation and interaction: Oxford University Press, 2015.

3. Mathur MB, Ding P, Riddell CA, et al. Web Site and R Package for Computing E-values. *Epidemiology* 2018;29(5):e45-e47. doi: 10.1097/ede.0000000000000864 [published Online First: 2018/06/19]