# Reproducibility of paper:

**Lin, L., Yi, X., Liu, H. et al. The airway microbiome mediates the interaction between environmental exposure and respiratory health in humans. Nat Med 29, 1750–1759 (2023).**

Javier Mancilla Galindo

2024-04-30

## Table of contents

**Packages used**

```r
if (!require("pacman", quietly = TRUE)) {
  install.packages("pacman")
}

pacman::p_load(
  tidyverse, # Used for basic data handling and visualization.
  dagitty, #Used in conjunction with https://www.dagitty.net/ to create
  #directed acyclic graph to inform statistical modelling.
  lavaan, #Used to create correlation matrix to assess conditional independencies.
  CBPS, #Used to calculate non-parametric propensity scores for IPW.
  WeightIt, #Used to calculate inverse probability weights.
  boot # Calculate bootstrap confidence intervals.
)
```

**Session and package dependencies**

```
R version 4.3.3 (2024-02-29 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 22631)

Matrix products: default


locale:
[1] LC_COLLATE=Spanish_Mexico.utf8  LC_CTYPE=Spanish_Mexico.utf8
[3] LC_MONETARY=Spanish_Mexico.utf8 LC_NUMERIC=C
[5] LC_TIME=Spanish_Mexico.utf8

time zone: Europe/Berlin
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] boot_1.3-30        WeightIt_1.0.0      CBPS_0.23
 [4] glmnet_4.1-8       Matrix_1.6-5        numDeriv_2016.8-1.1
 [7] nnet_7.3-19        MatchIt_4.5.5       MASS_7.3-60.0.1
[10] lavaan_0.6-17      dagitty_0.3-4       lubridate_1.9.3
```

```
[13] forcats_1.0.0      stringr_1.5.1      dplyr_1.1.4
[16] purrr_1.0.2        readr_2.1.5        tidyr_1.3.1
[19] tibble_3.2.1       ggplot2_3.5.0      tidyverse_2.0.0
[22] pacman_0.5.1
```

The summary of criticisms to the paper and rationale for the following analyses are summarized in the powerpoint presentation contained in the *mediation* directory.

## DAG

The following directed acyclic graph (DAG) was reconstructed based on assumptions available in the paper for the construction of regression models in mediation analyses, by using the DAGitty website. THe DAG is saved and sourced from the accompanying script **DAG.r**
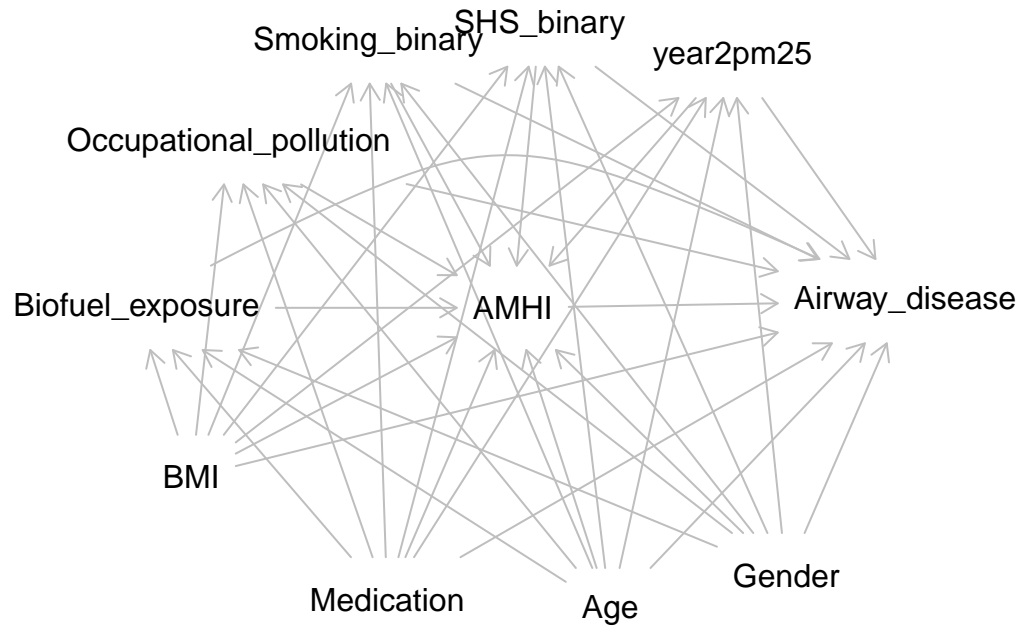


## Testing of conditional independencies in DAG:

This procedure was performed as suggested in this article.

Implied conditional independencies:

```
Age _||_ BMI
Age _||_ Gndr
Age _||_ Mdct
BMI _||_ Gndr
BMI _||_ Mdct
Bfl_ _||_ Occ_ | Age, BMI, Gndr, Mdct
Bfl_ _||_ SHS_ | Age, BMI, Gndr, Mdct
```

```
Bfl_ _||_ Smk_ | Age, BMI, Gndr, Mdct
Bfl_ _||_ y225 | Age, BMI, Gndr, Mdct
Gndr _||_ Mdct
Occ_ _||_ SHS_ | Age, BMI, Gndr, Mdct
Occ_ _||_ Smk_ | Age, BMI, Gndr, Mdct
Occ_ _||_ y225 | Age, BMI, Gndr, Mdct
SHS_ _||_ Smk_ | Age, BMI, Gndr, Mdct
SHS_ _||_ y225 | Age, BMI, Gndr, Mdct
Smk_ _||_ y225 | Age, BMI, Gndr, Mdct


                         Bfl_xp Occpt_ Smkng_ SHS_bn yr2p25    Age    BMI Medctn
Biofuel_exposure          1.000
Occupational_pollution    0.304  1.000
Smoking_binary            0.171  0.342  1.000
SHS_binary                0.154  0.296  0.052  1.000
year2pm25                -0.196 -0.122 -0.267 -0.040  1.000
Age                       0.159  0.112  0.141 -0.032 -0.005  1.000
BMI                      -0.220 -0.035 -0.125 -0.125  0.026 -0.136  1.000
Medication                0.372  0.458  0.231  0.245  0.002  0.006 -0.252  1.000
Gender                    0.026  0.328  0.939 -0.148 -0.214  0.108 -0.104  0.183
Airway_disease            0.216  0.104  0.003  0.012  0.287  0.059 -0.155  0.519
AMHI                     -0.110 -0.134 -0.025 -0.041 -0.125 -0.089 -0.078 -0.392
                         Gender Arwy_d   AMHI
Biofuel_exposure
Occupational_pollution
Smoking_binary
SHS_binary
year2pm25
Age
BMI
Medication
Gender                    1.000
Airway_disease            0.048  1.000
AMHI                      0.007 -0.548  1.000


                                   estimate      p.value        2.5%
Age _||_ BMI                     -0.136414439 9.604998e-02 -0.29047053
Age _||_ Gndr                     0.107697031 1.899115e-01 -0.05348771
Age _||_ Mdct                     0.005961319 9.423807e-01 -0.15445367
BMI _||_ Gndr                    -0.104254367 2.045743e-01 -0.26025039
BMI _||_ Mdct                    -0.252231108 1.774505e-03 -0.39715968
Bfl_ _||_ Occ_ | Age, BMI, Gndr, Mdct  0.188498630 2.252495e-02  0.02687314
```

```
Bfl_ _||_ SHS_ | Age, BMI, Gndr, Mdct  0.055390463 5.072971e-01 -0.10803097
Bfl_ _||_ Smk_ | Age, BMI, Gndr, Mdct  0.382088405 1.485078e-06  0.23422305
Bfl_ _||_ y225 | Age, BMI, Gndr, Mdct -0.234927411 4.199311e-03 -0.38334693
Gndr _||_ Mdct                          0.183276828 2.460759e-02  0.02371177
Occ_ _||_ SHS_ | Age, BMI, Gndr, Mdct  0.304120705 1.729680e-04  0.14904018
Occ_ _||_ Smk_ | Age, BMI, Gndr, Mdct  0.025186032 7.632278e-01 -0.13782960
Occ_ _||_ y225 | Age, BMI, Gndr, Mdct -0.087881979 2.920506e-01 -0.24686925
SHS_ _||_ Smk_ | Age, BMI, Gndr, Mdct  0.547029400 2.073739e-13  0.42304675
SHS_ _||_ y225 | Age, BMI, Gndr, Mdct -0.088189061 2.903589e-01 -0.24716023
Smk_ _||_ y225 | Age, BMI, Gndr, Mdct -0.209778365 1.088447e-02 -0.36038277
                                            97.5%
Age _||_ BMI                             0.02438008
Age _||_ Gndr                            0.26349879
Age _||_ Mdct                            0.16607269
BMI _||_ Gndr                            0.05695887
BMI _||_ Mdct                           -0.09584441
Bfl_ _||_ Occ_ | Age, BMI, Gndr, Mdct   0.34083839
Bfl_ _||_ SHS_ | Age, BMI, Gndr, Mdct   0.21592797
Bfl_ _||_ Smk_ | Age, BMI, Gndr, Mdct   0.51534956
Bfl_ _||_ y225 | Age, BMI, Gndr, Mdct  -0.07535500
Gndr _||_ Mdct                           0.33402330
Occ_ _||_ SHS_ | Age, BMI, Gndr, Mdct   0.44585615
Occ_ _||_ Smk_ | Age, BMI, Gndr, Mdct   0.18688523
Occ_ _||_ y225 | Age, BMI, Gndr, Mdct   0.07564652
SHS_ _||_ Smk_ | Age, BMI, Gndr, Mdct   0.66120201
SHS_ _||_ y225 | Age, BMI, Gndr, Mdct   0.07533880
Smk_ _||_ y225 | Age, BMI, Gndr, Mdct  -0.04899982
```

Local tests results plot:

x-axis label: test statistic (95% CI)

y-axis labels (top to bottom):
- _ _||_ Smk_ | Age, BMI, Gndr, Mdct
- _ _||_ y225 | Age, BMI, Gndr, Mdct
- BMI _||_ Gndr
- Age _||_ BMI
- _ _||_ Occ_ | Age, BMI, Gndr, Mdct
- _ _||_ y225 | Age, BMI, Gndr, Mdct
- _ _||_ SHS_ | Age, BMI, Gndr, Mdct
- _ _||_ Smk_ | Age, BMI, Gndr, Mdct
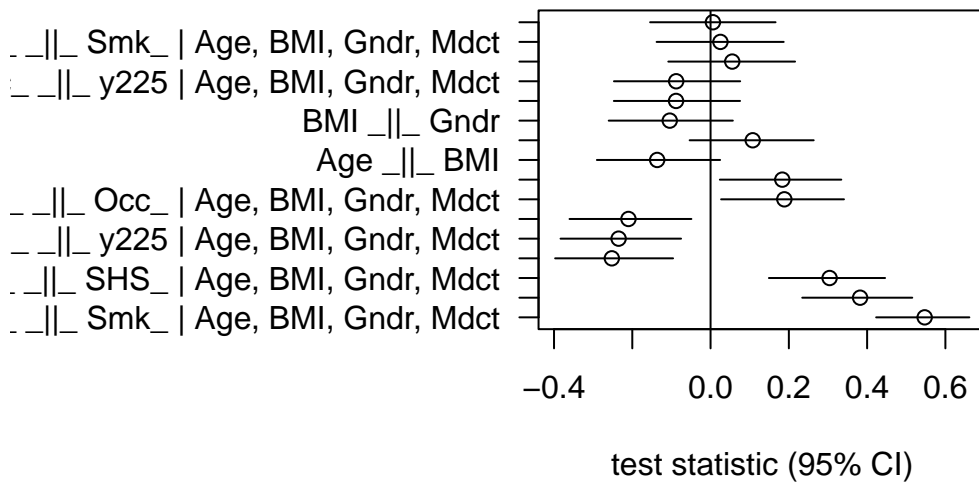
I updated the DAG according to the procedure described by Ankan, 2021.
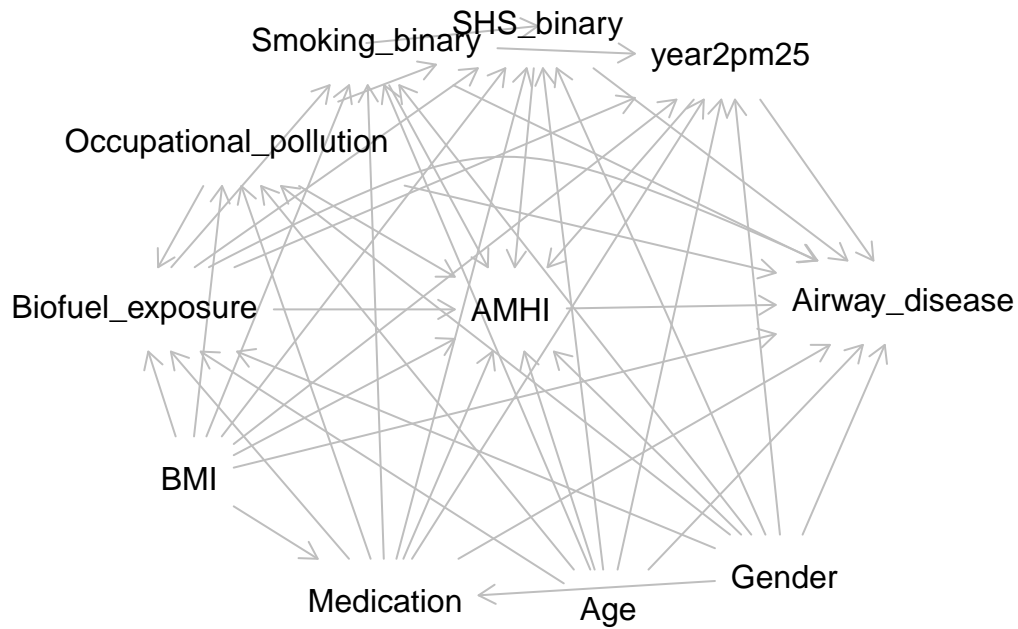


```
                                 estimate      p.value
Age _||_ BMI                   -0.136414439  0.09604998
```

```
Age _||_ Gndr                                            0.107697031 0.18991153
Age _||_ Mdct                                            0.005961319 0.94238067
BMI _||_ Gndr                                           -0.104254367 0.20457426
Occ_ _||_ Smk_ | Age, BMI, Bfl_, Gndr, Mdct            -0.051607761 0.53820853
Occ_ _||_ y225 | Age, BMI, Bfl_, Gndr, Mdct            -0.045672563 0.58600780
SHS_ _||_ y225 | Age, BMI, Bfl_, Gndr, Mdct, Smk_      -0.001613923 0.98471005
                                                             2.5%        97.5%
Age _||_ BMI                                            -0.29047053 0.02438008
Age _||_ Gndr                                          -0.05348771 0.26349879
Age _||_ Mdct                                          -0.15445367 0.16607269
BMI _||_ Gndr                                          -0.26025039 0.05695887
Occ_ _||_ Smk_ | Age, BMI, Bfl_, Gndr, Mdct           -0.21285615 0.11234779
Occ_ _||_ y225 | Age, BMI, Bfl_, Gndr, Mdct           -0.20716543 0.11821834
SHS_ _||_ y225 | Age, BMI, Bfl_, Gndr, Mdct, Smk_     -0.16515452 0.16201228
```

Local tests results plot:
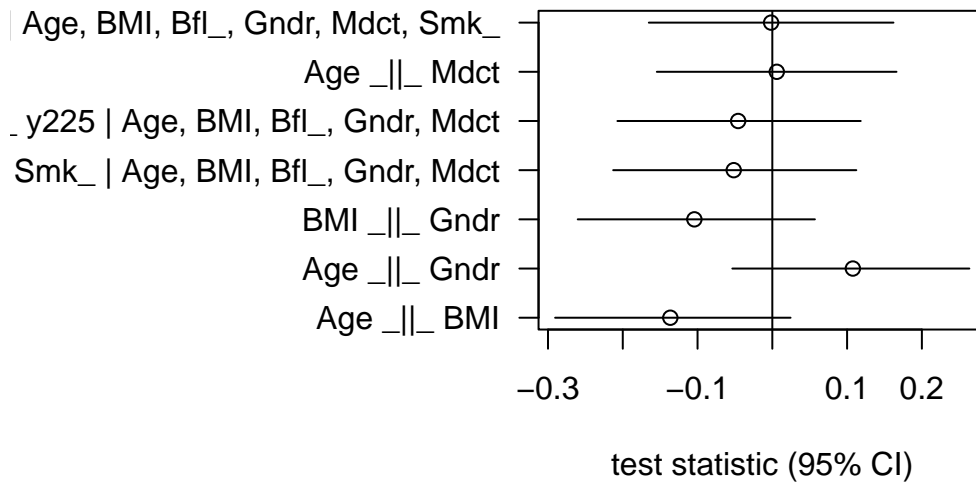
## Biofuel

### Inverse probability weights

Weights for exposure:

Weights for mediator:

Overall weight:

Confidence interval

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_results, type = "all")

Intervals :
Level      Normal              Basic
95%   (-28.0075,  29.2961 )   ( -4.0017,   4.8129 )

Level      Percentile          BCa
95%   (-4.1996,  4.6150 )   (-1.5276, 13.7922 )
Calculations and Intervals on Original Scale
```

## Occupational pollution

### Inverse probability weights

Weights for exposure:

Weights for mediator:

Overall weight:

Confidence interval

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_results, type = "all")
```

```
Intervals :
Level      Normal              Basic
95%   (-962.870,  957.094 )   ( -22.483,    8.908 )


Level      Percentile          BCa
95%   (  -16.420,    14.971 )   (-47254.558,    -1.360 )
Calculations and Intervals on Original Scale
Warning : BCa Intervals used Extreme Quantiles
Some BCa intervals may be unstable
```

## PM 2.5

**Inverse probability weights**

Weights for exposure:

Weights for mediator:

Overall weight:

Confidence interval

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_results, type = "all")

Intervals :
Level      Normal              Basic
95%   (-48.9454,  48.3801 )   ( -0.8053,    0.6633 )


Level      Percentile          BCa
95%   (-0.7096,  0.7590 )   (-0.6328,  0.9072 )
Calculations and Intervals on Original Scale
```

# Smoking

### Inverse probability weights

Weights for exposure:

Weights for mediator:

Overall weight:

Confidence interval

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_results, type = "all")

Intervals :
Level      Normal              Basic
95%   (-52.7994,  53.0275 )   ( -5.9490,   6.3779 )

Level      Percentile          BCa
95%   (-5.7300,  6.5969 )   (-2.1082, 16.5905 )
Calculations and Intervals on Original Scale
```

# Second hand smoking

### Inverse probability weights

Weights for exposure:

Weights for mediator:

Overall weight:

Confidence interval

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_results, type = "all")
```

```
Intervals :
Level      Normal                Basic
95%   (-61.014,  72.596 )   ( -2.389,  13.715 )

Level     Percentile            BCa
95%   (  -7.930,    8.173 )   (   1.141, 1095.965 )
Calculations and Intervals on Original Scale
Warning : BCa Intervals used Extreme Quantiles
Some BCa intervals may be unstable
```

## Table results

```
              Exposure ci_lower   ci_upper confidence_interval
1       Biofuel_exposure        0  4.6149729            0, 4.61
2 Occupational_pollution        0 14.9713783           0, 14.97
3                   PM25        0  0.7589639            0, 0.76
4                Smoking        0  6.5969129             0, 6.6
5    Secondhand_smoking        0  8.1734864            0, 8.17
```