IMPERIAL

Unpacking the Exposome's Influence on Chronic Kidney Disease: The Mediating Role of Biomarkers

Computational Epidemiology

Group 3

Dan Rose, Wanqing Fang, Berin Pinar, Yilin Chen, Titouan Guillou

Background

Chronic Kidney Disease and the Exposome

Status quo in the UK:

- 22% of adults over 35 and 50% of adults over 75 years of age have CKD (~7.2m individuals)
- A substantial proportion of CKD cases remain undiagnosed in the general population
- CKD is an established risk factor for cardiovascular disease, cognitive impairment, and stroke, which collectively cost the UK economy >£100bn p.a.

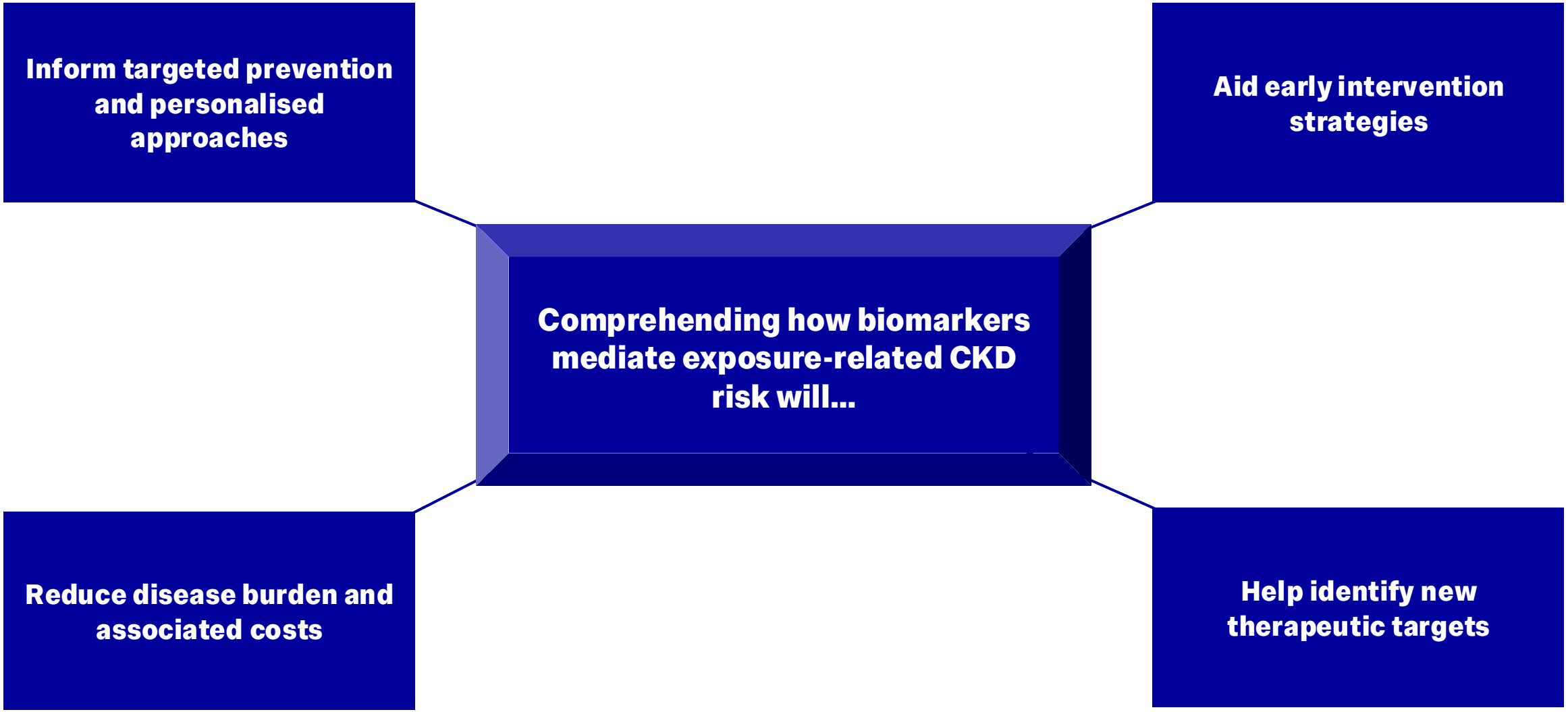
A plethora of exposures are associated with CKD risk:

- Smoking
- Alcohol Consumption
- Obesity and diet
- PM2.5
- Deprivation

Our research question:

To what extent does the exposome contribute to the development of chronic kidney disease, and how are these effects mediated via biomarkers?

Study Motivations



Data Processing

Preprocessing steps

Exclusion Criteria:

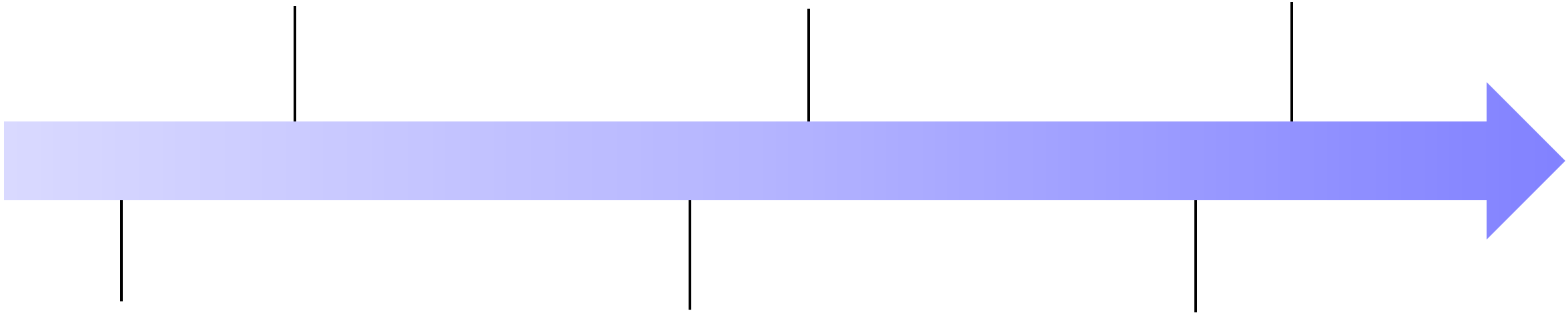
- Individuals not in UK Biobank
- Prevalent cases
- Limitations

Scaling/One-Hot Encoding:

- Reference dummy column removed to aid interpretation and reduce collinearity in analysis

Imputation (MICE):

- Test data imputed based off training data, preventing data leakage



Case Definition:

- Cases - persons diagnosed with chronic kidney disease (CKD) as reported by HES and/or registry data.

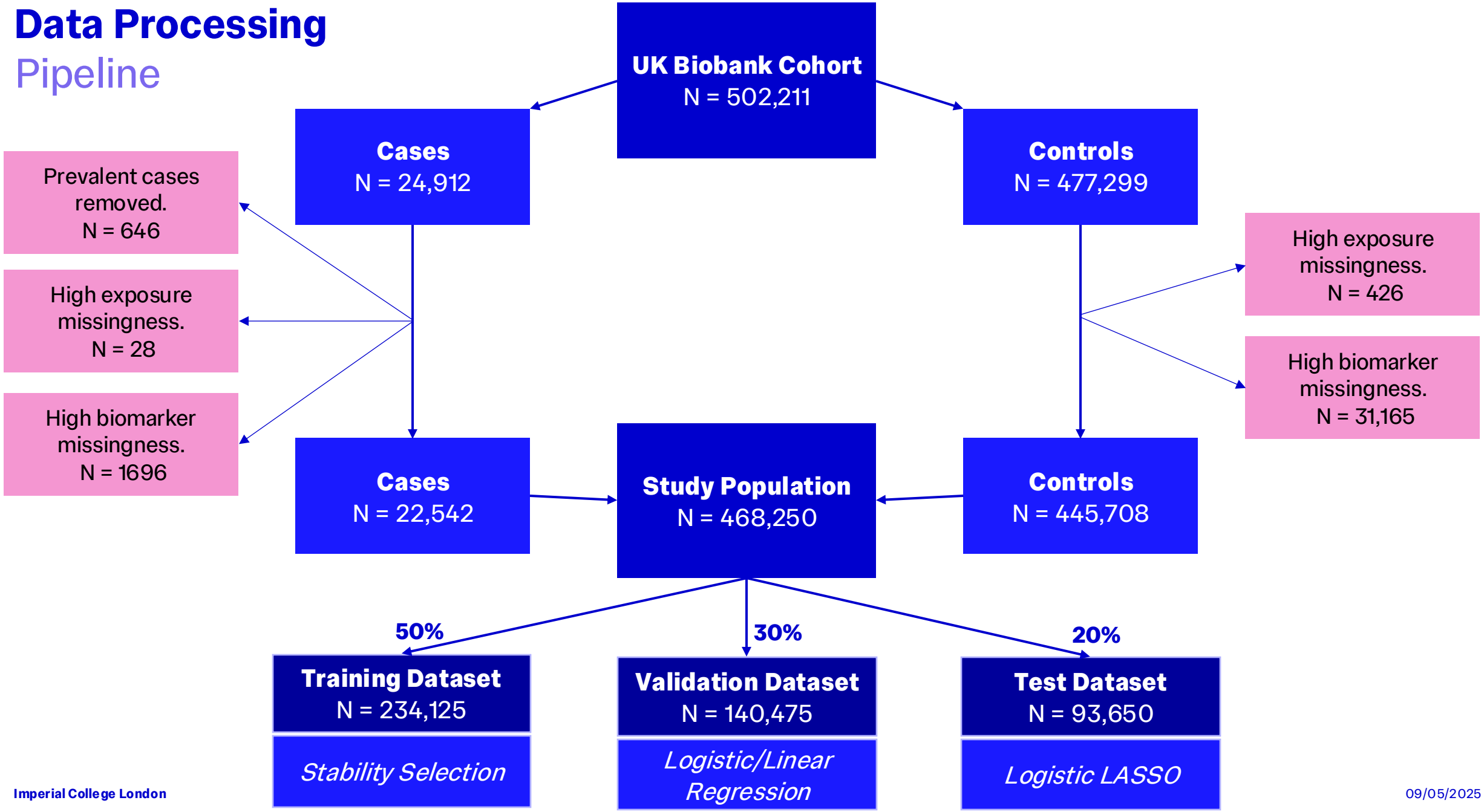
Missingness:

- Individuals with >50% missing biomarker and/or exposure data removed
- Biomarkers with >50% missingness removed (rheumatoid factor, oestradiol)

Data Splits:

- 50:30:20 train-validation-test split

Data Processing
Pipeline

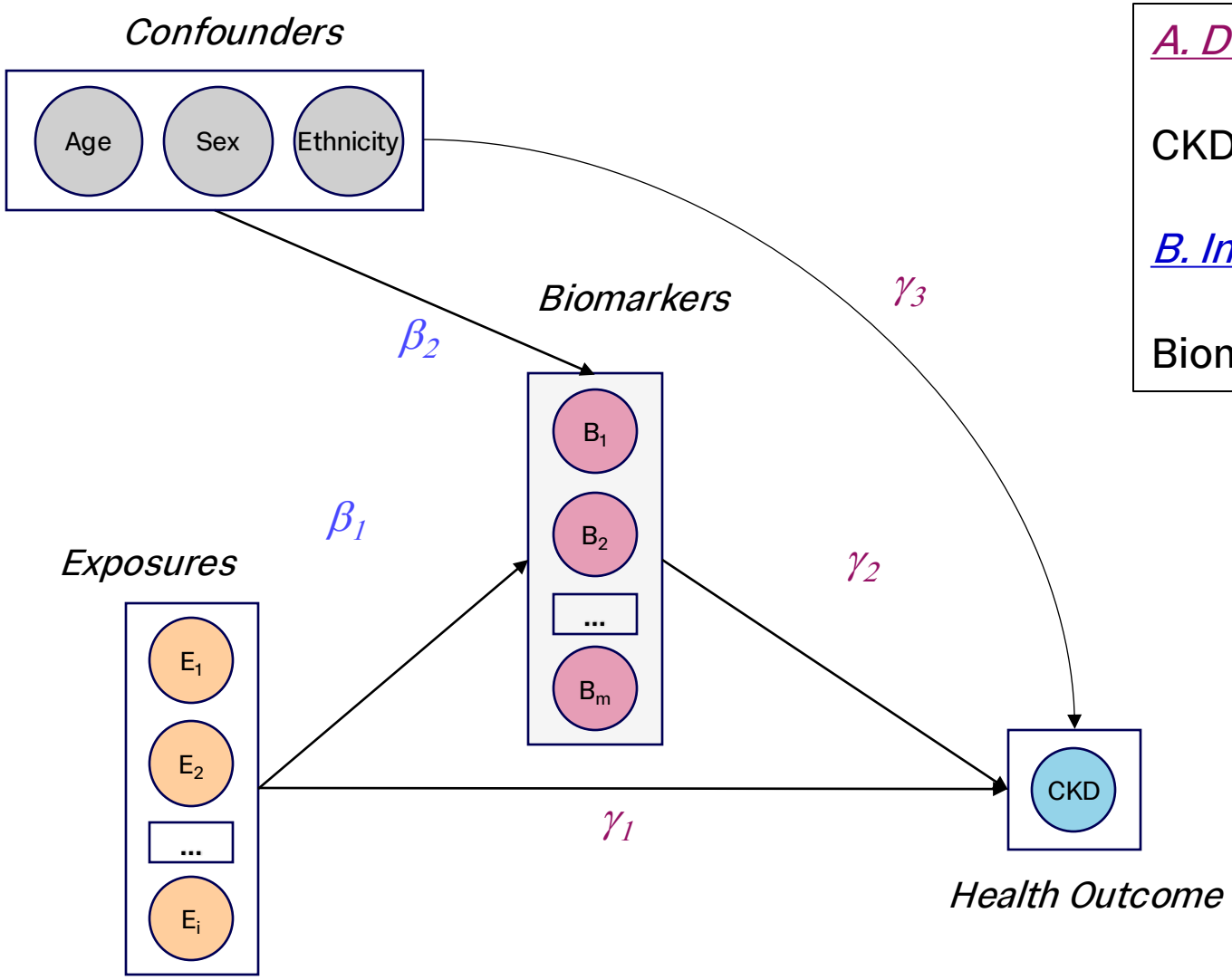


Methods

Overview

Modelling approach

Selected DAG of informative relationships



A. Direct Effect Model:

CKD = γ_1 *Exposures + γ_2 * Biomarkers + γ_3 * *Confounders*

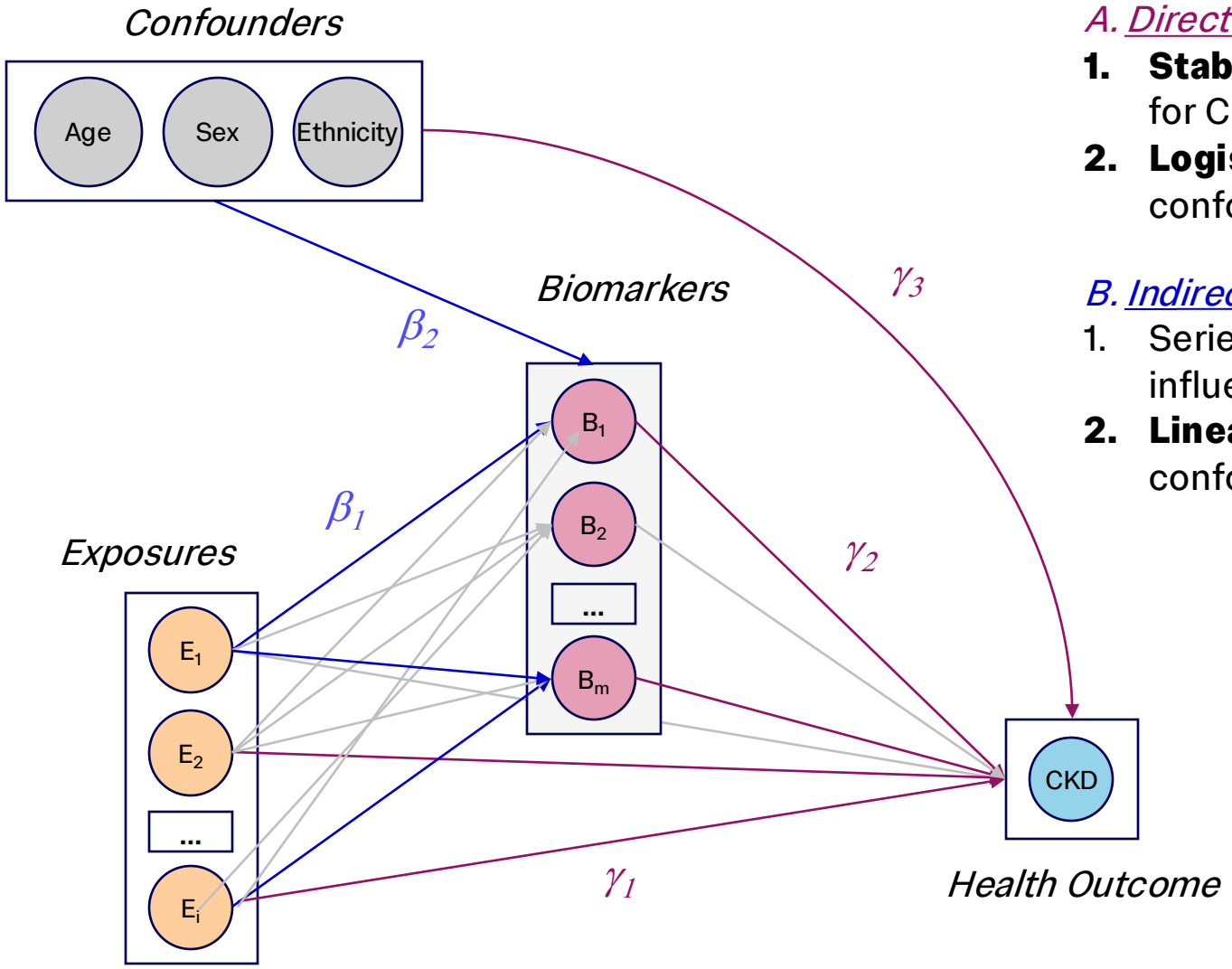
B. Indirect Effect Model:

Biomakers = β_2 * Confounders + β_1 * Exposures

Exposures: N = 34
Biomarkers: N = 41

Modelling approach

Selected DAG of informative relationships



A. Direct Effect Model:

1. **Stability LASSO** to select essential exposures and biomarkers for CKD.
2. **Logistic regression** with only the selected variables and confounders to retrieve unbiased estimates .

B. Indirect Effect Model:

1. Series of **Stability LASSOs** for the selection of exposures that influence previously selected biomarkers .
2. **Linear regression** with only the selected exposures and confounders to retrieve unbiased estimates on each biomarkers.

- **Direct Effect** of exposure i (γ_{1i})
- **Indirect Effect** of exposure i

$$\sum_{\substack{j \text{ selected} \\ biomarkers}} \gamma_{2\_j}\, \beta_{1\_ij}$$

γ_{1i}, γ_{2i} : *log-odds*
$\beta_{1_{ij}}$: *linear coefficient*

Results

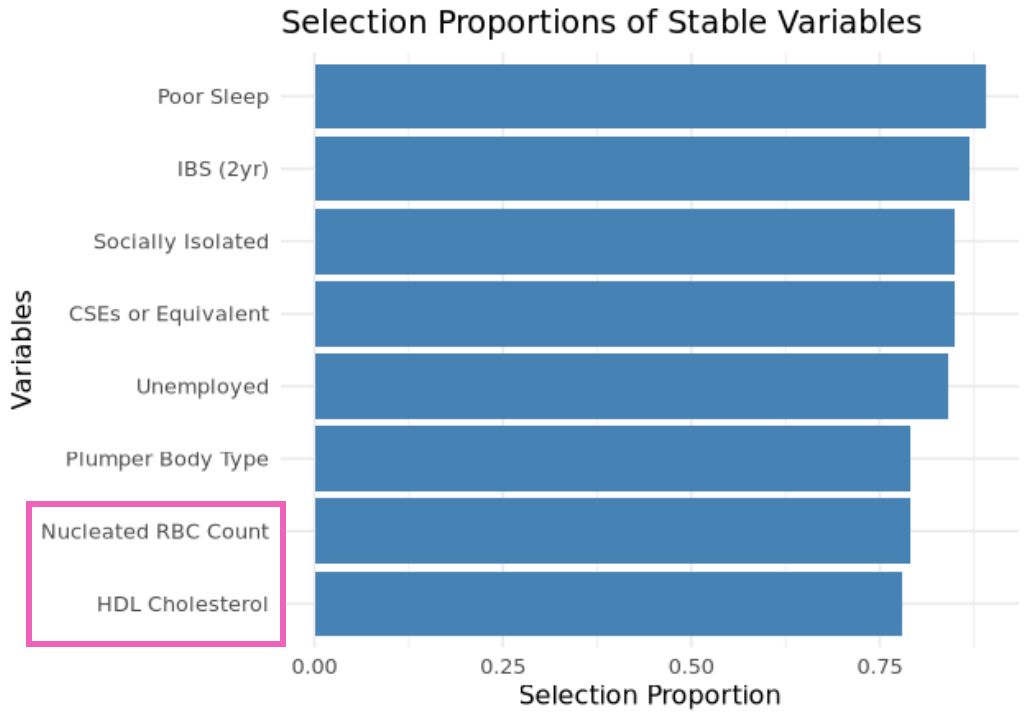
Descriptive Analysis

	Total (n = 468,250)			
	Case	Control	p-value	Missing(%)
Sample Characteristics	22,542	445,708		
Demographics				
Sex = Male (%)	11659 (51.7)	202665 (45.5)	<0.001	0
Age (median [IQR])	64.00 [60.00, 67.00]	57.00 [50.00, 63.00]	<0.001	0
Ethnicity (%)			<0.001	0.1
White (reference)	21137 (93.9)	420369 (94.4)		
Asian	561 (2.5)	9933 (2.2)		
Black	443 (2.0)	6808 (1.5)		
Other	362 (1.6)	8045 (1.8)		
Clinical Variables				
BMI (median [IQR])	28.85 [25.93, 32.56]	26.64 [24.07, 29.75]	<0.001	0.4
Systolic Blood Pressure (median [IQR])	142.00 [130.00, 155.50]	136.00 [124.50, 149.00]	<0.001	8.6
Socioeconomic Variables				
Employment Status (%)			<0.001	0.5
Employed (reference)	6399 (28.5)	252158 (56.9)		
Looking after home and/or family	309 (1.4)	12604 (2.8)		
Other	219 (1.0)	5794 (1.3)		
Retired	13808 (61.6)	151814 (34.2)		
Unable to work due to sickness	1404 (6.3)	13838 (3.1)		
Unemployed	286 (1.3)	7334 (1.7)		
Own or Rent (%)			<0.001	0
Own (reference)	18796 (83.4)	395908 (88.8)		
Rent	3091 (13.7)	39689 (8.9)		
Other	655 (2.9)	10111 (2.3)		
Lifestyle Variables				
Alcohol Intake (%)			<0.001	0.6
Never/Abstinent (Excluding previous drinkers)	1580 (7.1)	18993 (4.3)		
Previous Drinker (who currently never drink)	1407 (6.3)	15362 (3.5)		
Light (reference)	12810 (57.3)	242718 (54.8)		
Moderate	3378 (15.1)	84242 (19.0)		
Heavy*	1922 (8.6)	50260 (11.3)		
Abusive*	1258 (5.6)	31482 (7.1)		
Smoking Status (%)			<0.001	0.5
Never (reference)	10161 (45.4)	244906 (55.2)		
Previous	9651 (43.1)	152099 (34.3)		
Current	2557 (11.4)	46600 (10.5)		

Results

Stably Selected Variables – Direct Model

Selection Proportions of Stable Variables Predicting CKD (Direct Model)



Stability selection threshold (π)	0.78
Optimal lambda (λ)	0.000492

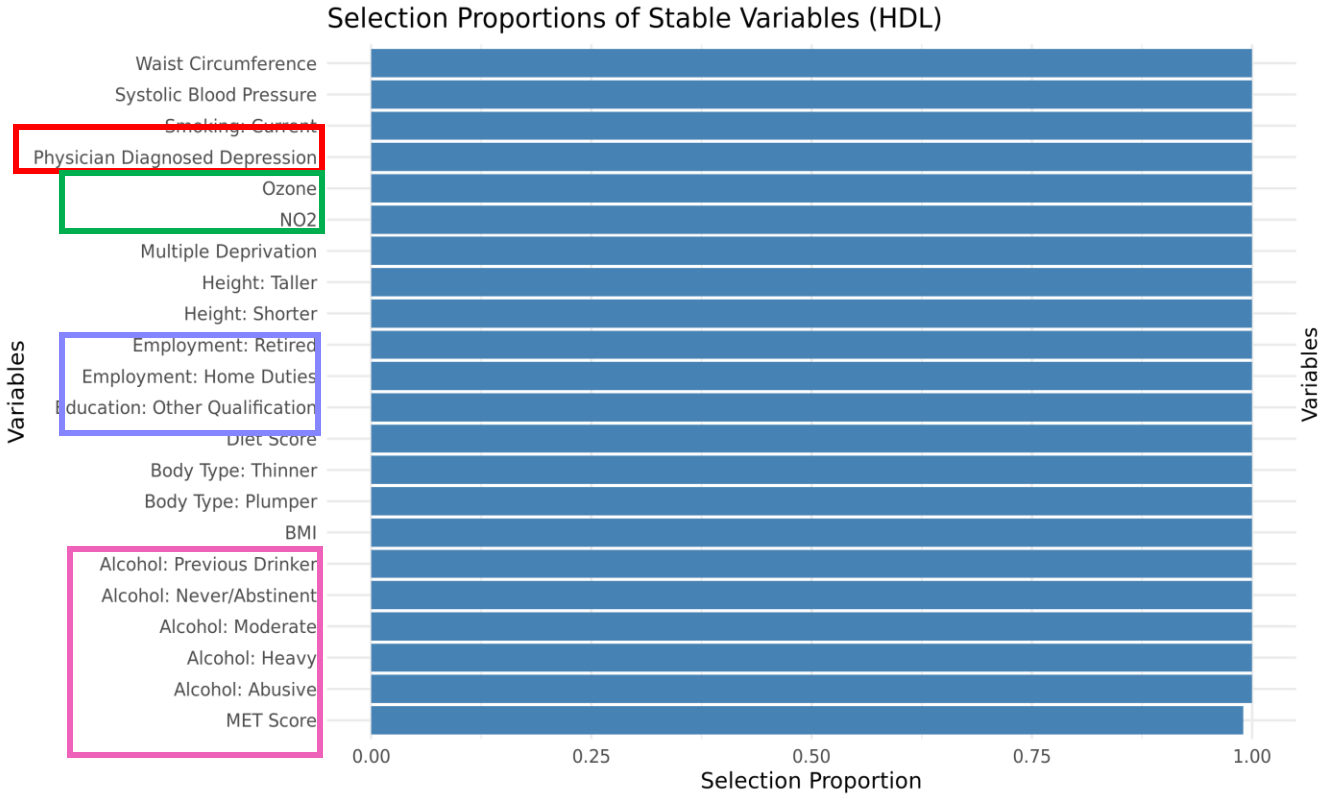
- 8 variables were selected:
  - 6 are exposures: poor sleep, having IBS in last 2 years, being socially isolated, having education up to CSEs or equivalent, being unemployed, having a plumper body composition
  - 2 are biomarkers: Nucleated RBC count, HDL cholesterol
- These variables were included in the logistic regression model to estimate direct effects on CKD risk.

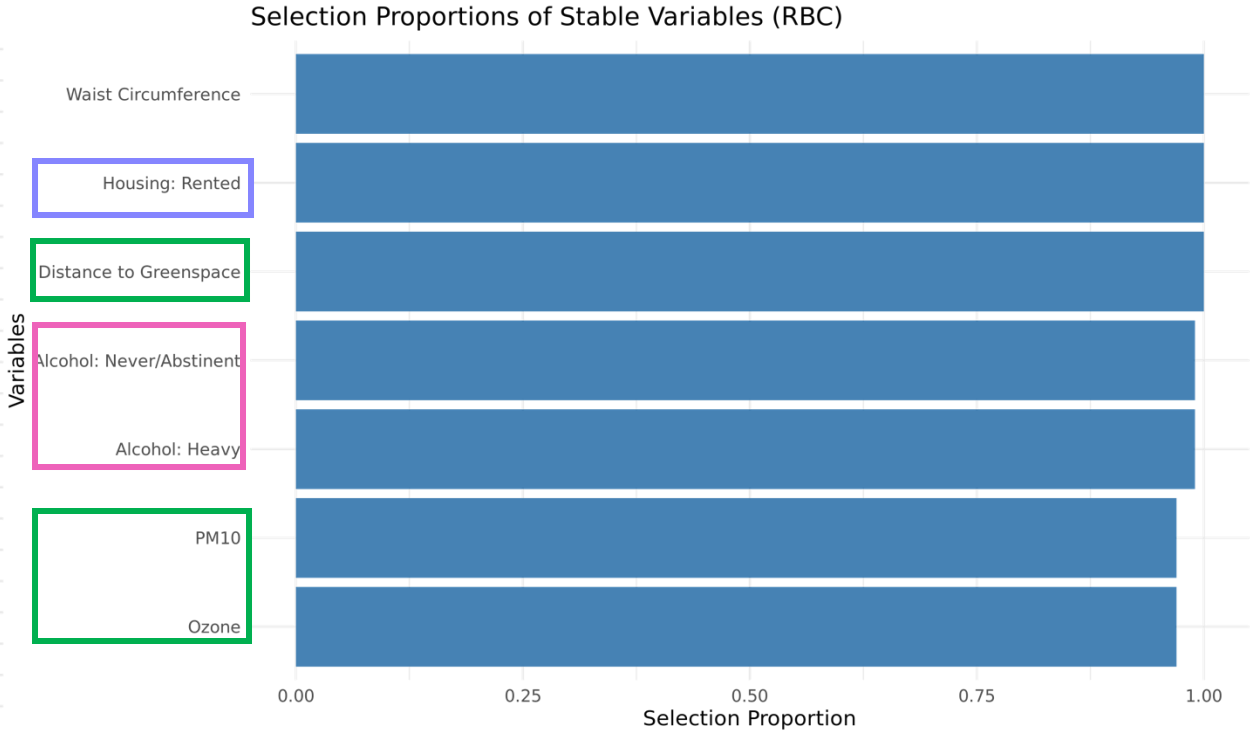
Results

Stably Selected Variables – Indirect Model

Selection Proportions of Stable Variables Predicting CKD (Indirect Model)

	HDL	RBC
Stability selection threshold (π)	0.99	0.97
Optimal lambda (λ)	0.008985489	0.006042341



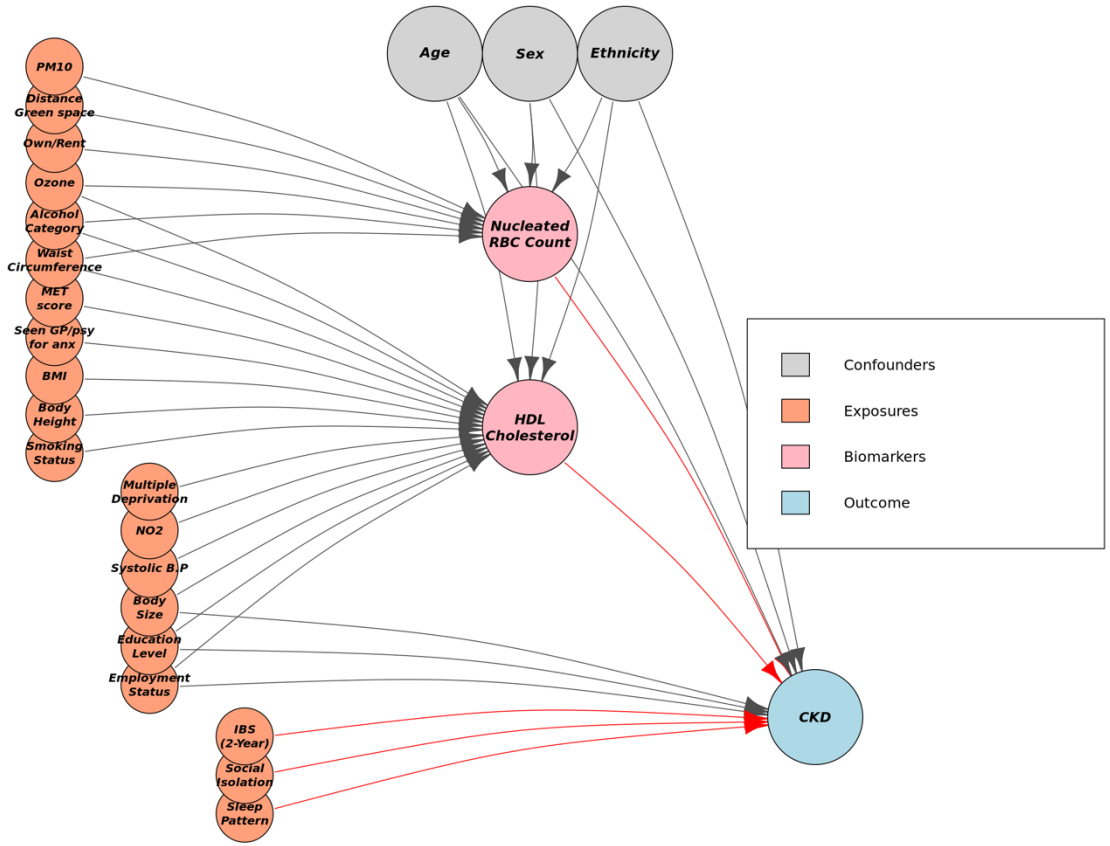


Results

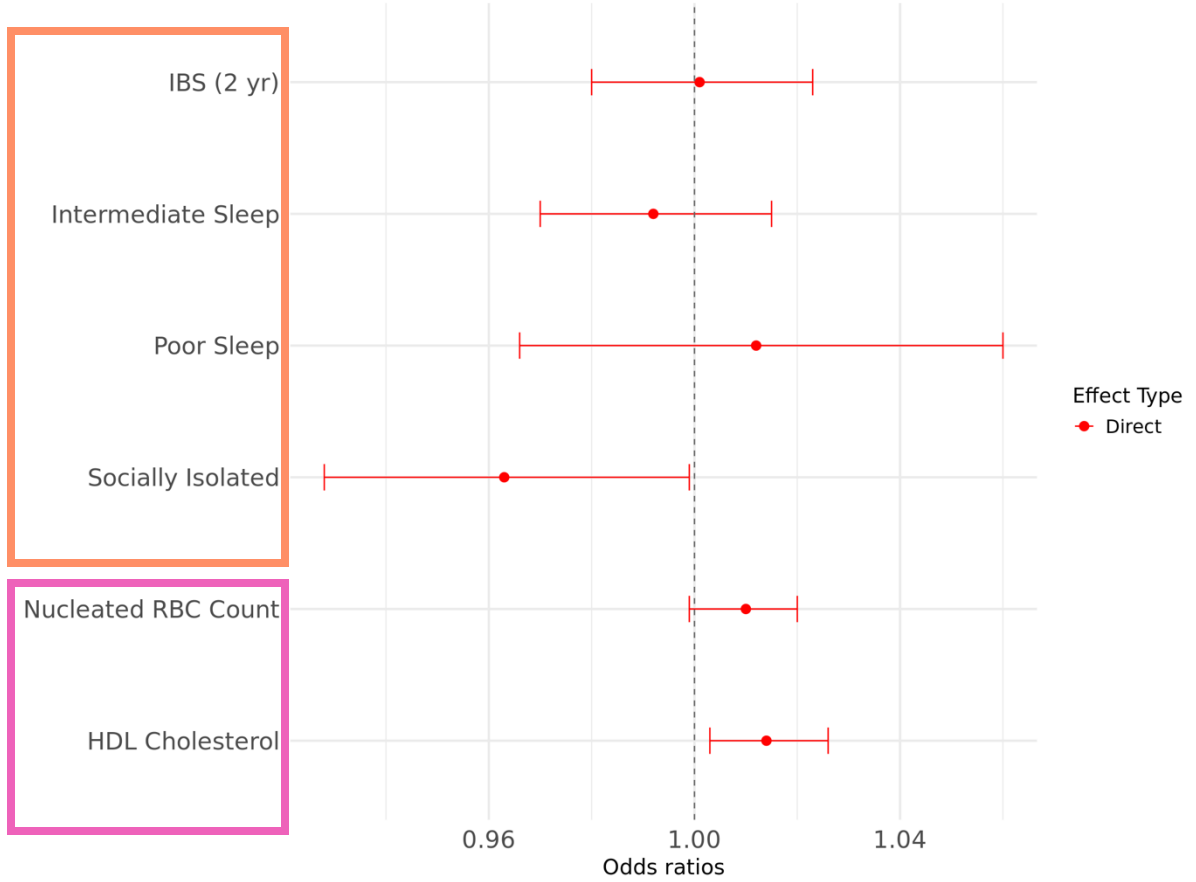
Sparse DAG

Results

Direct effect

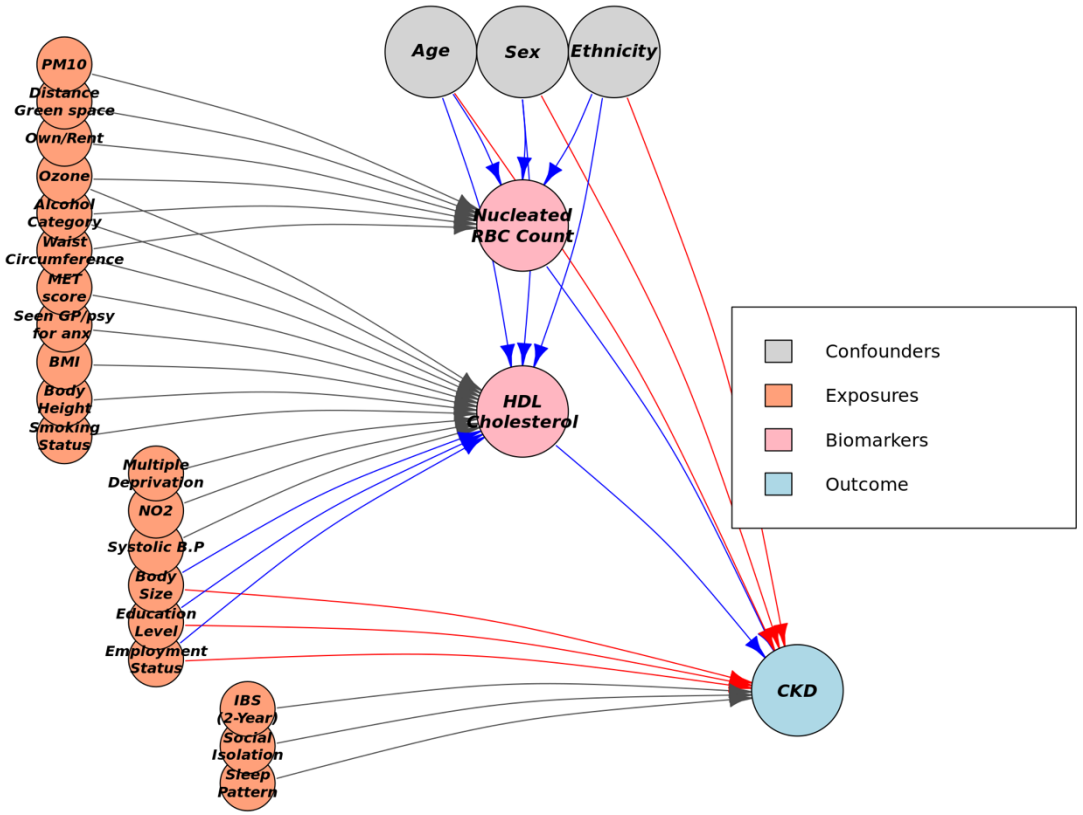


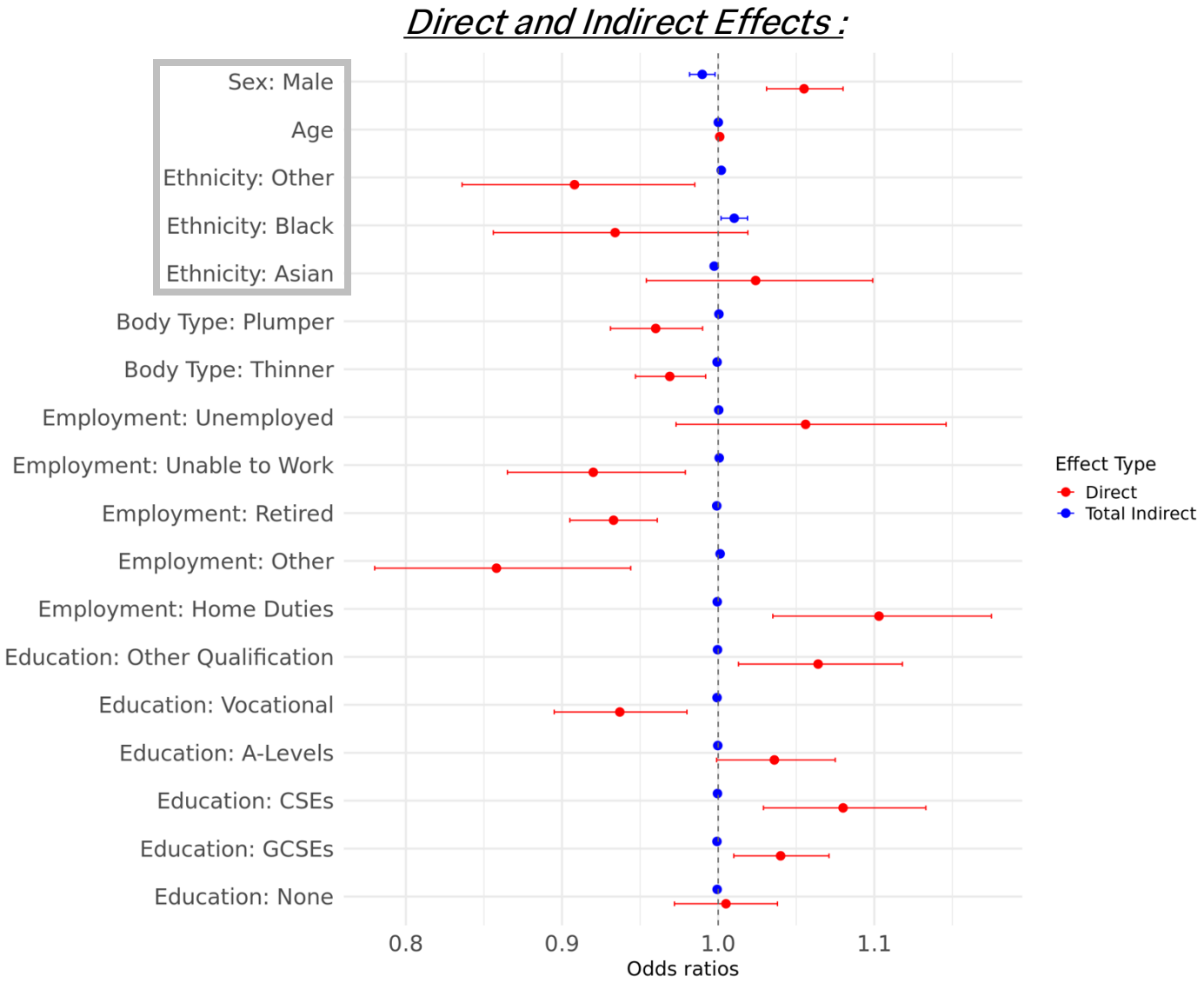
Direct Effects :



Results

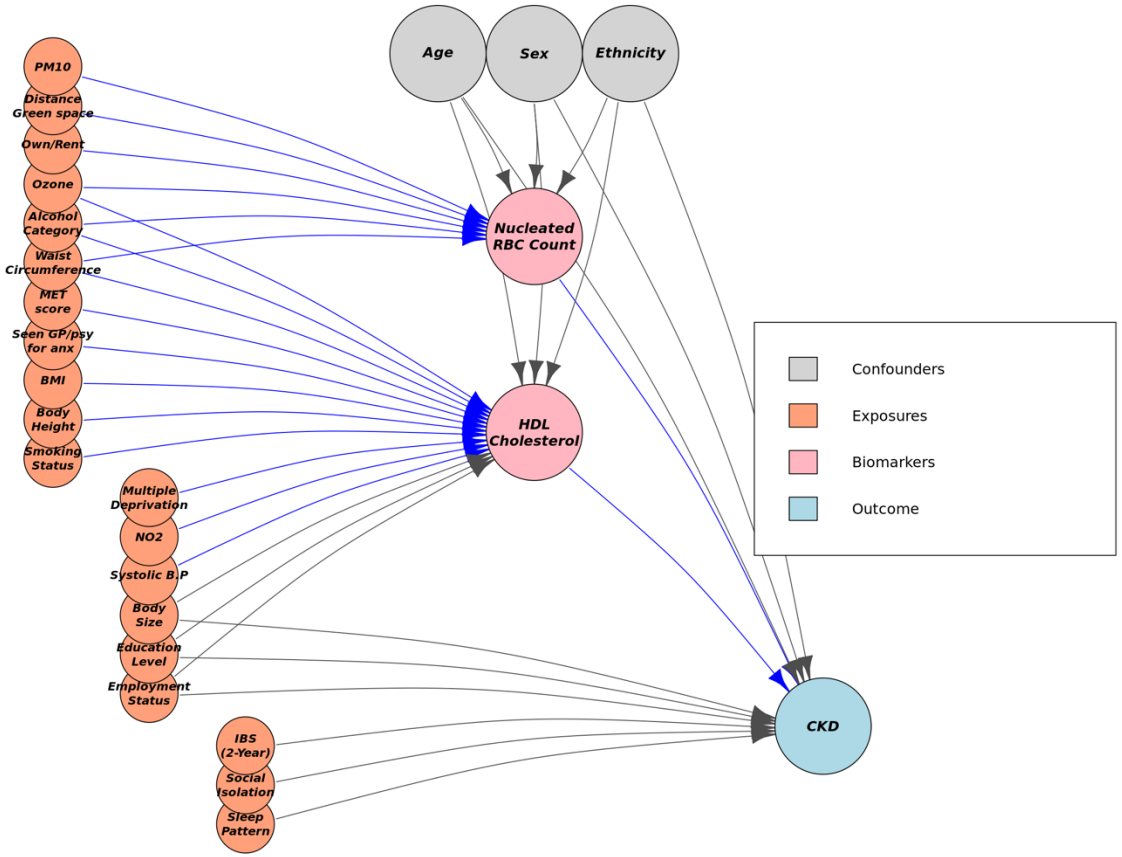
Direct and indirect effect



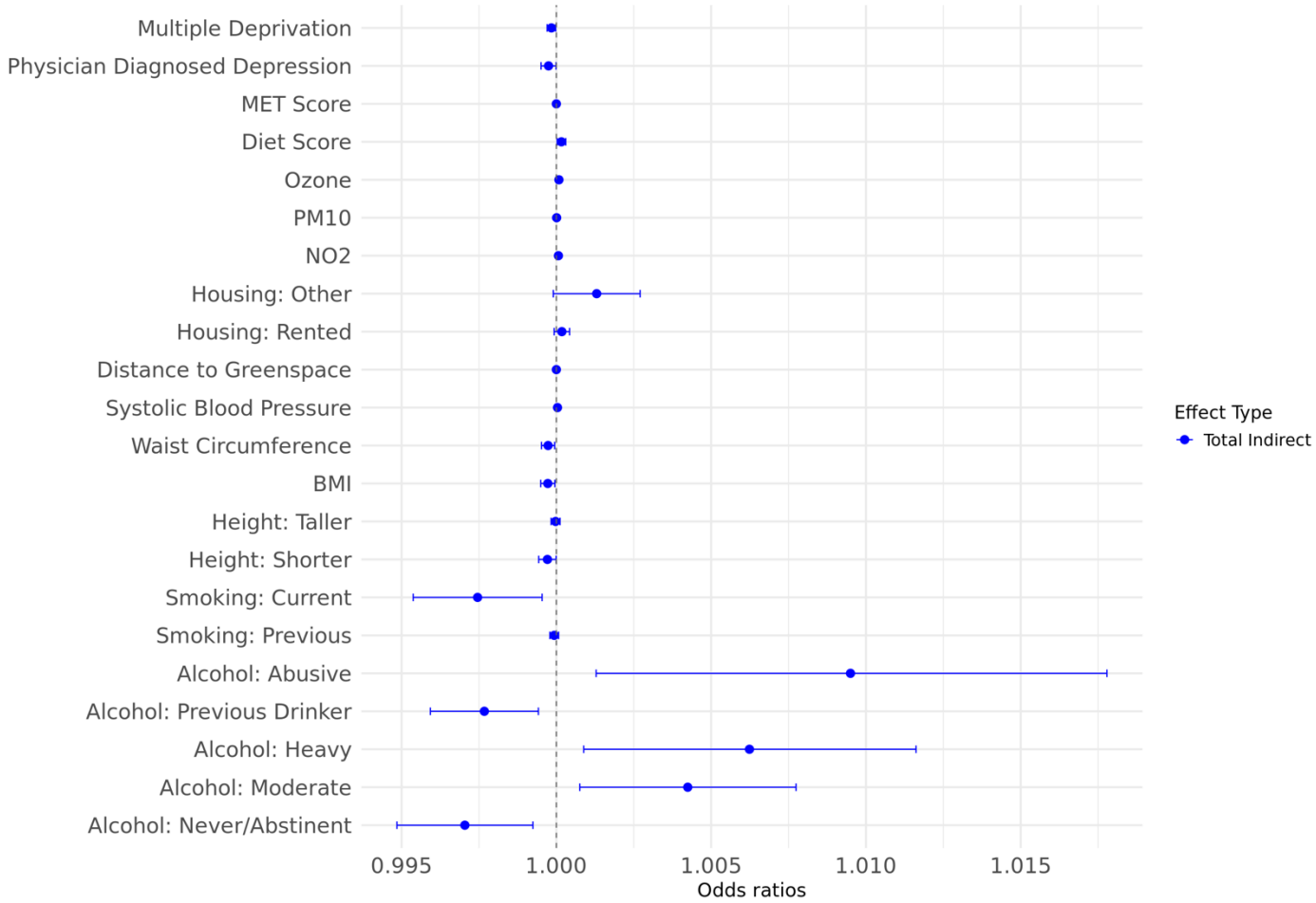


Results

Indirect/Mediated effect

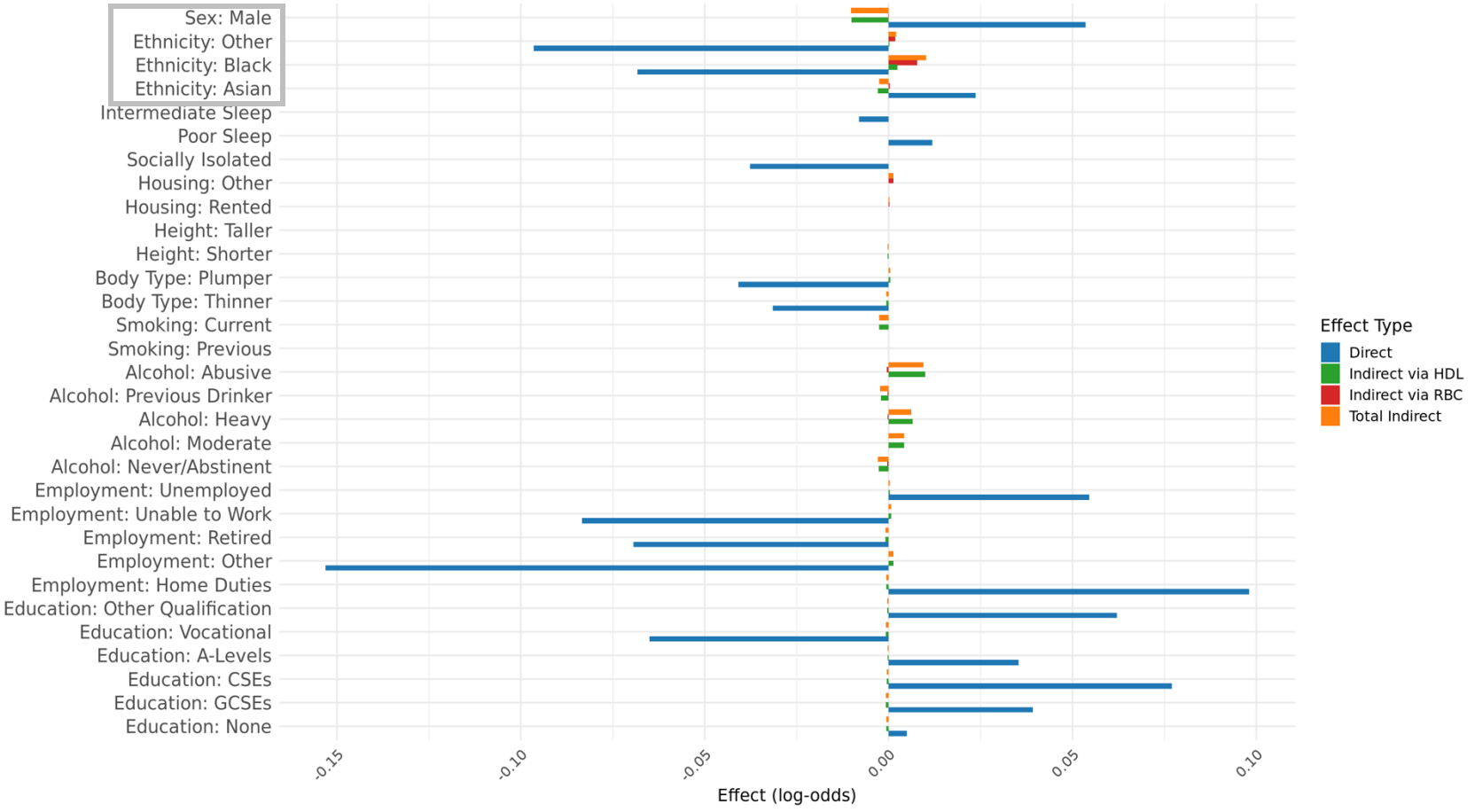


Mediated Effects :



Modelling approach

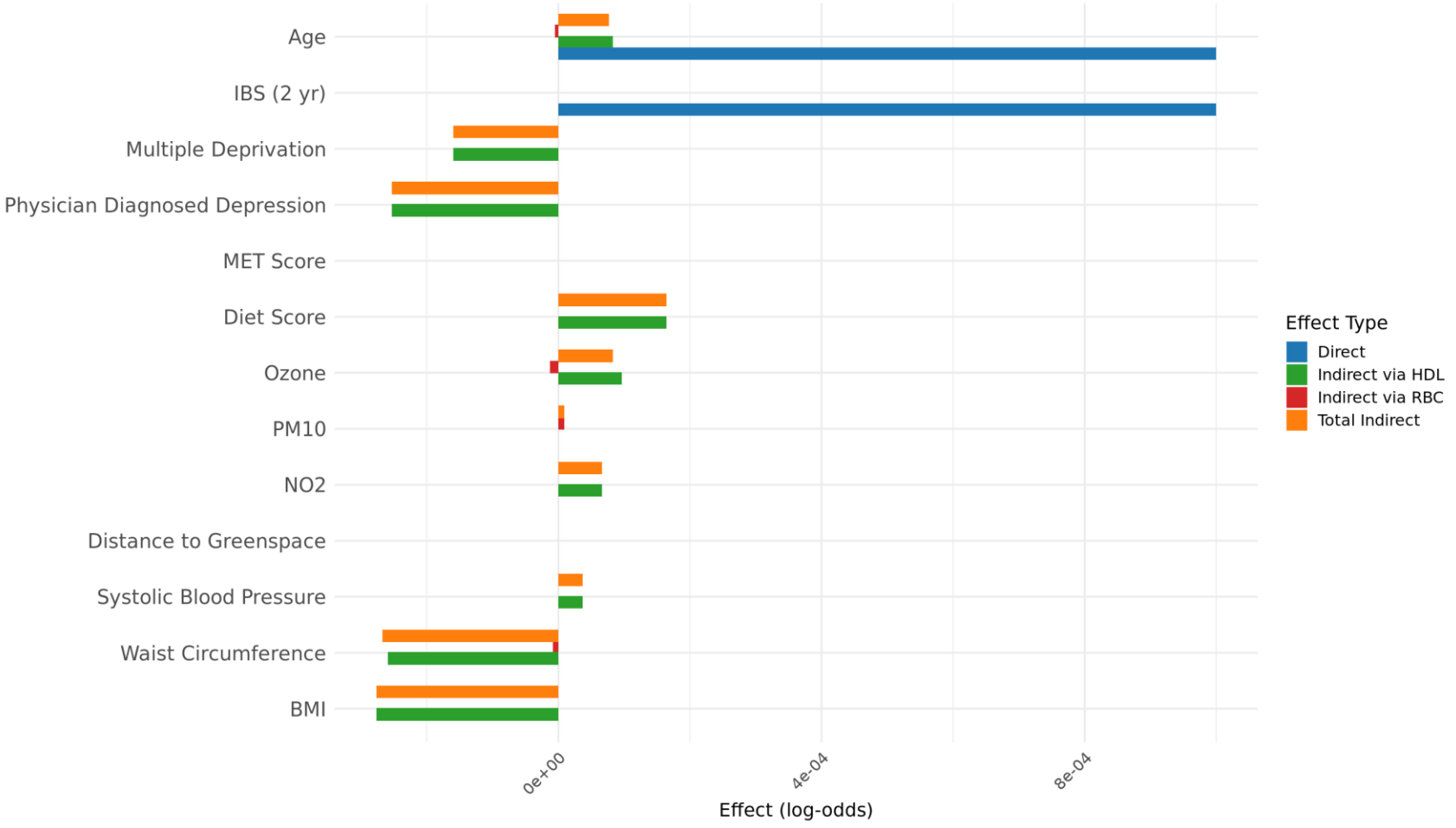
Summary of direct and indirect (mediated) effects – Categorical variables



Variable	Proportion mediated effect (%)
Sex(Male)	**16%**
Ethnicity (Other)	2%
Ethnicity (Black)	**13%**
Ethnicity (Asian)	**10%**
Body Type : Plumper	1%
Body Type : Thinner	2%
Employment (Unemployed)	1%
Employment (Unable to work)	1%
Employment (Retired)	1%
Employment (Other)	1%
Employment (Home duties)	1%
Employment (Other qualification)	**12%**
Education (Vocational)	1%
Education (A-level)	1%
Education (CSEs)	0.5%
Education (GCSEs)	2%
Education (None)	**12%**

Modelling approach

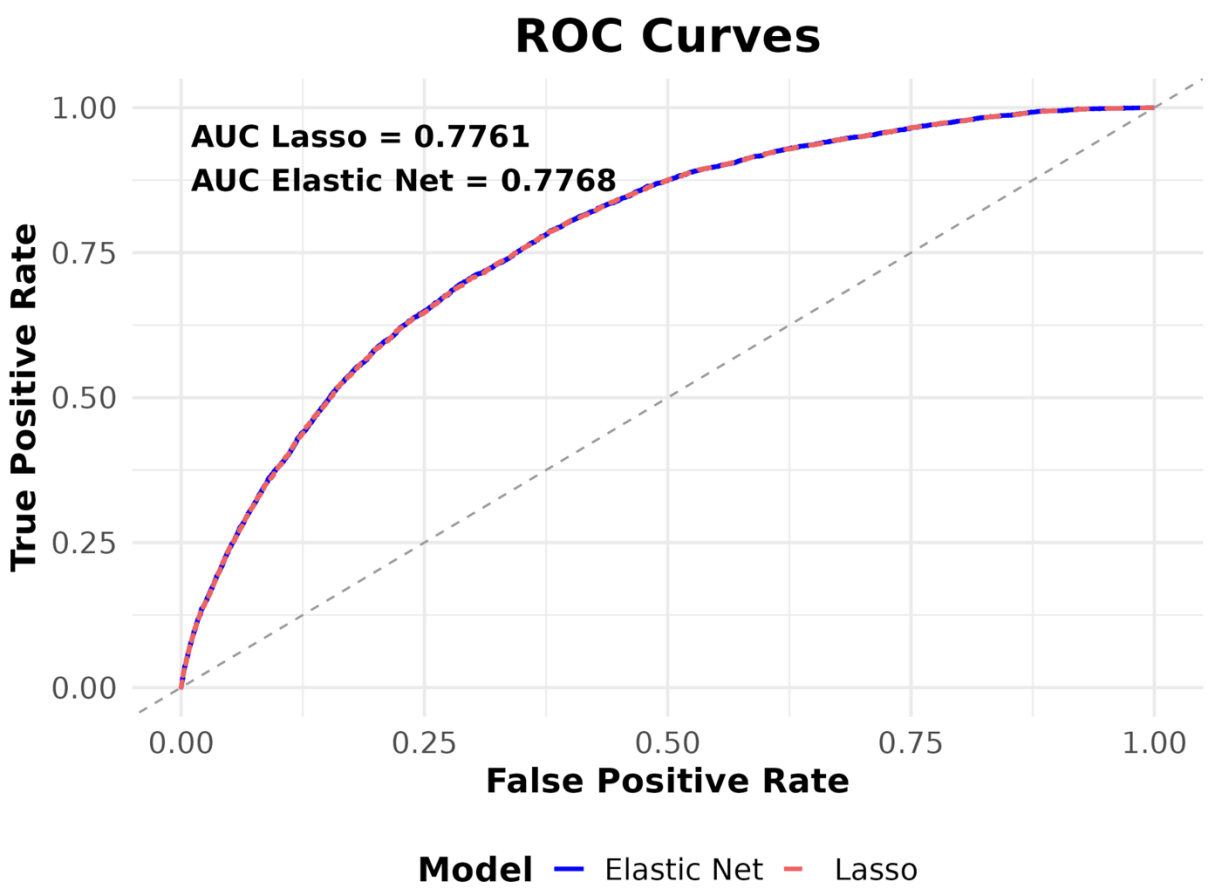
Summary of direct and indirect (mediated) effects – Continuous variables



Variable	Proportion mediated effect (%)
Age	7%

Results

Prediction Results



- Out of all 25 variables, both models selected 22
- Lasso excluded "***Distance to greenspace***" and "***Smoking status***"
- Elastic Net excluded "***Social isolation***" and "***IBS in 2 years***"
- Both models excluded "***Nucleated RBC Count***"

Model	Accuracy	Sensitivity	Specificity
Lasso	0.68	0.10	0.98
Elastic net	0.68	0.10	0.98

Discussion
Conclusions

Indirect and Direct Effect
- Direct effects of exposome on CKD were found to be stronger than effects mediated HDL and nucleated RBC count
  - Effects might be mediated through other, potentially unmeasured, biological pathways.

HDL Cholesterol
- HDL cholesterol was found to be positively associated with CKD:
  - May reflect underlying metabolic dysfunction or HDL particle impairment[6]
  - Its effect on CKD could be mediated/confounded unmeasured factors affecting HDL
- BMI, waist circumference, and alcohol consumption show significant effect via HDL
  - Affect lipid metabolism → influence kidney health

nRBC and Prediction
- nRBC emerged to be a weak mediator → also weak predictor (not selected by LASSO & ElNet)
- Predictive models are not good at identifying true cases (low sensitivity) → limits clinical use

Study Limitation

Exposure and biomarkers measurement

- Behavioural factors were measured by self-reported questionnaires
- Exposures and biomarkers were measured once at baseline
- Potential temporal mismatch between exposures and mediators

Outcome measurement

- HES captured hospitalised and death registery cases only
- CKD is often silent until late stage
- Risk of reverse causality

UK Biobank cohort

- Elder and pre-dominantly white population of UK Biobank limits the generalisability of our findings to diverse populations

Future Considerations

Enhance exposure/biomarkers measurement

- Collect exposures and biomarkers at multiple time point during the follow-up
- Expand biomarker panels (e.g KIM-1)

Subgroup analysis

- Subgroup analysis stratified by sex, ethnicity, etc
- Inform public health policy making

External validation

- Validate signatures in diverse populations

IMPERIAL

Thank you

Questions welcomed

30/04/2025

References

1. Kranzler HR, Rosenthal RN. Dual diagnosis: Alcoholism and co-occurring psychiatric disorders. Am J Addict. 2003;12 Suppl 1:S26–S40. doi:10.1080/10550490390226912. PMID: 16775042.

2. Ryu V, Hardin JW, Crawford CA. Long-Term Effects of Adolescent Alcohol Exposure on the Hippocampus and Associated Cognitive Functions: A Review of Rodent Studies. Alcohol Clin Exp Res. 2021;45(8):1615–1630. doi:10.1111/acer.14112.

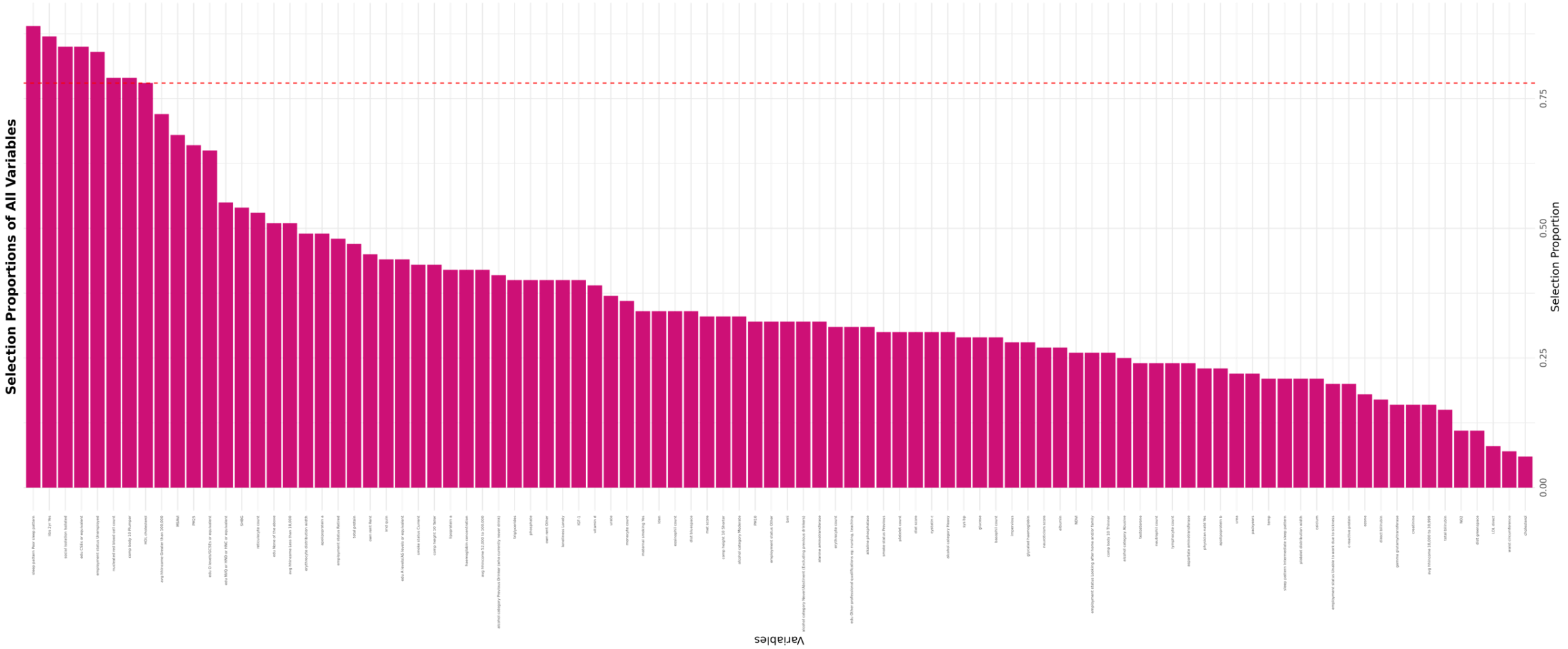
3. Bouvier d'Yvoire A, Schmidt A, Hüppi PS, Victor M, Barisnikov K. Executive Function and Psychopathology in Preterm Born Adolescents. Neonatology. 2023;112(3):121–130. doi:10.1159/000529089

4. Airagnes G, Lemogne C, Renuy A, Goldberg M, Hoertel N, Roquelaure Y, et al. Sedative–hypnotic use in the French working population: Results from the CONSTANCES cohort. PLoS One. 2021;16(1):e0245042. doi:10.1371/journal.pone.0245042. PMID: 33481947; PMCID: PMC7796538.

5. Bommersbach TJ, Rossom RC, Buchanan R, John Clark C, Patrick AR, Tollefson GD, et al. Comparative Safety of Benzodiazepines and Z-Drugs: A Population-Based Cohort Study. JAMA Netw Open. 2023;6(7):e2324876. doi:10.1001/jamanetworkopen.2023.24876. PMID: 37495291; PMCID: PMC10310512.
6. Butler MJ. Dysfunctional HDL takes its Toll on the endothelial glycocalyx. Kidney Int. 2020;97(3):450–2.

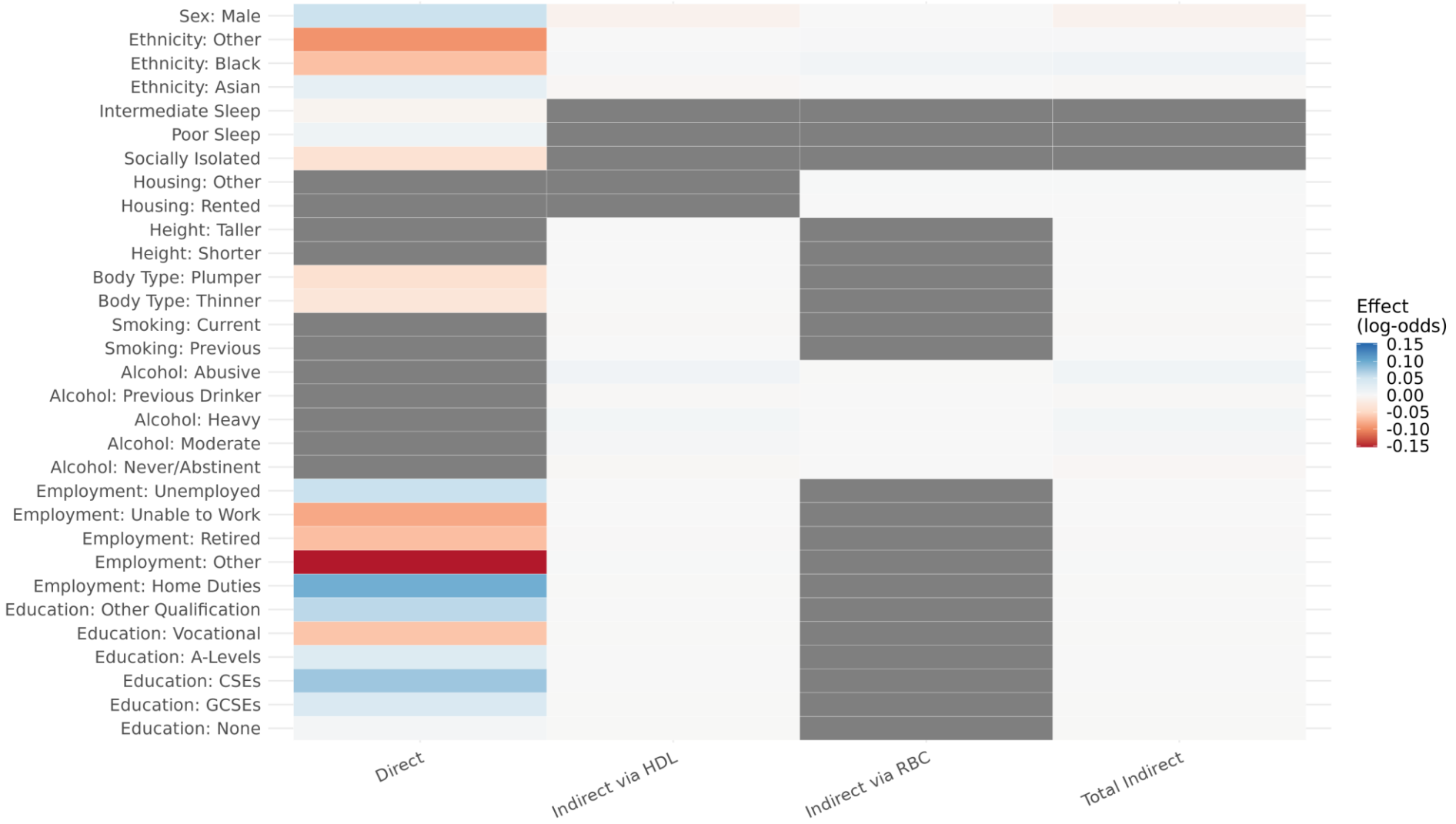
Appendix

Selection Proportion of All Variables (from Direct Model)



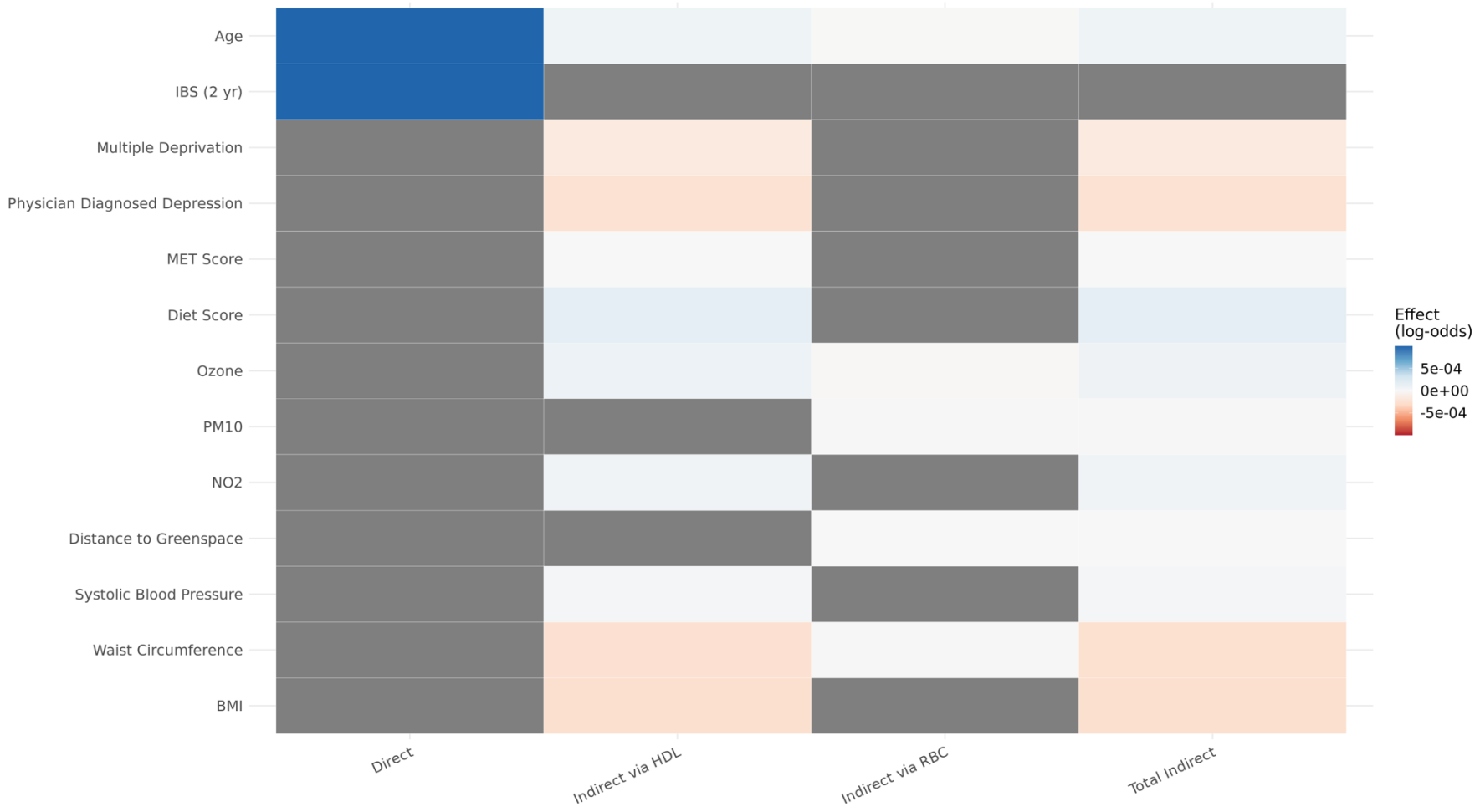
Appendix

Heatmap of the effects - categorical



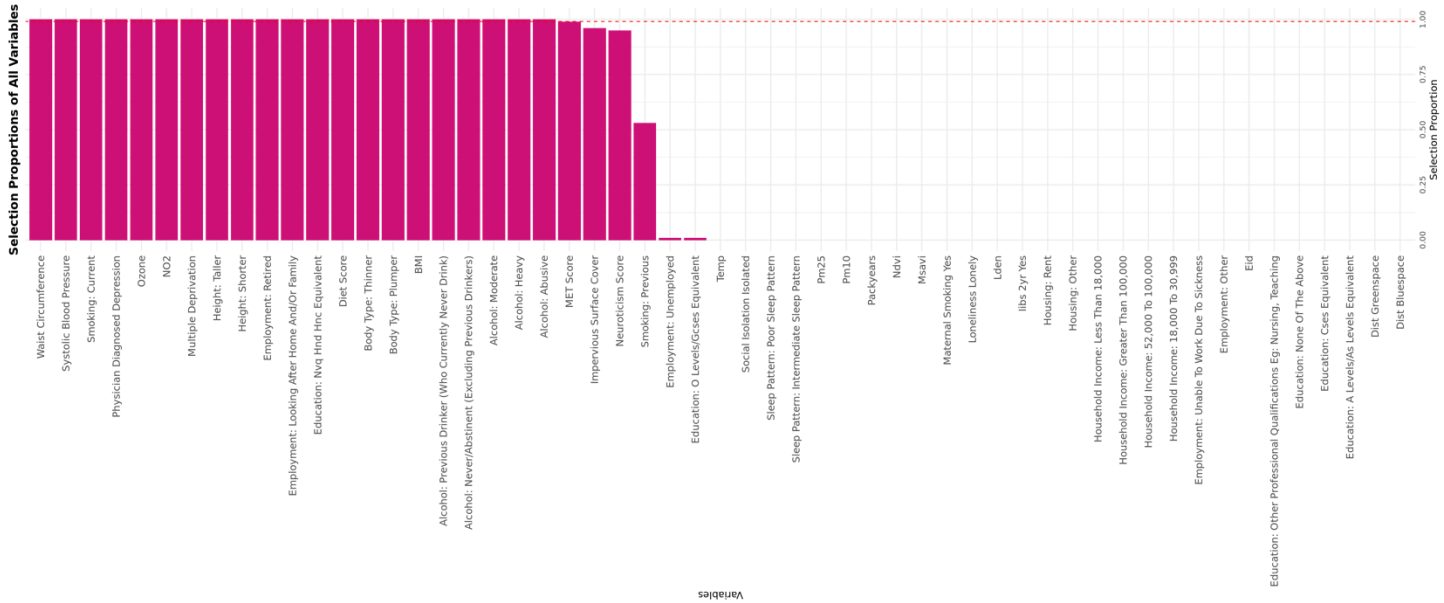
Appendix

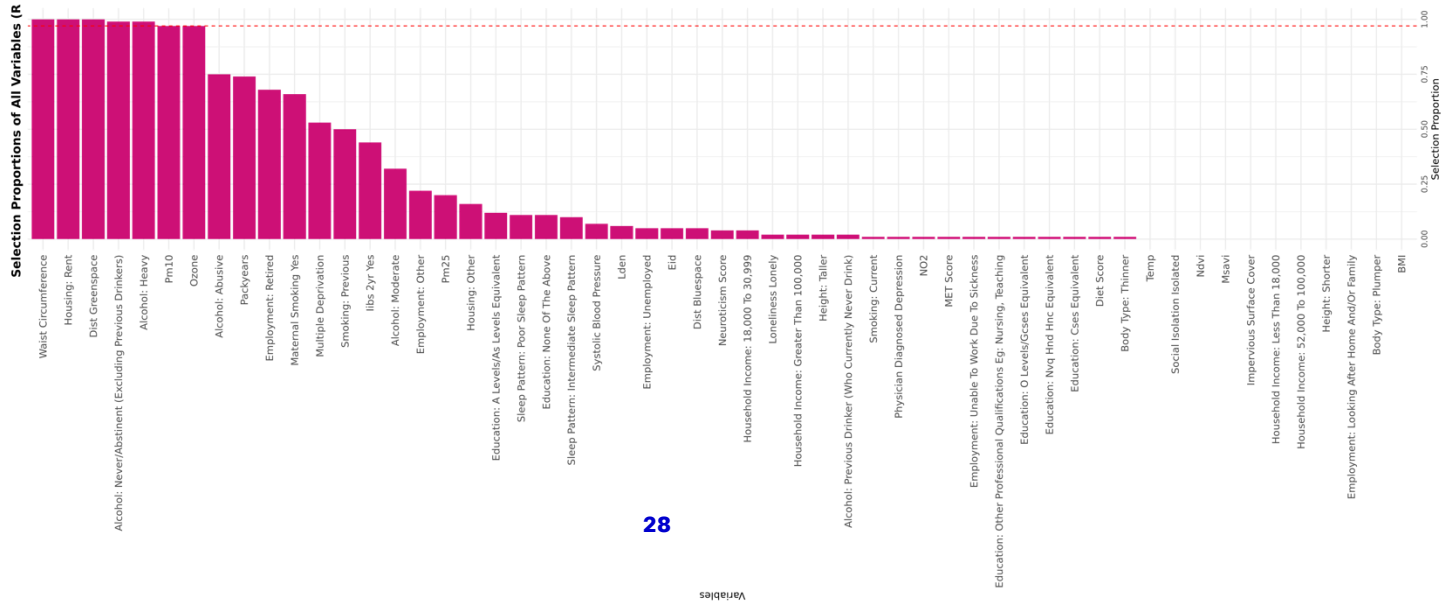
Heatmap of the effects - continous



Appendix

Selection Proportion of All Variables (from Indirect Model)





Appendix

Confusion Matrix (lasso & elastic net)

Lasso:

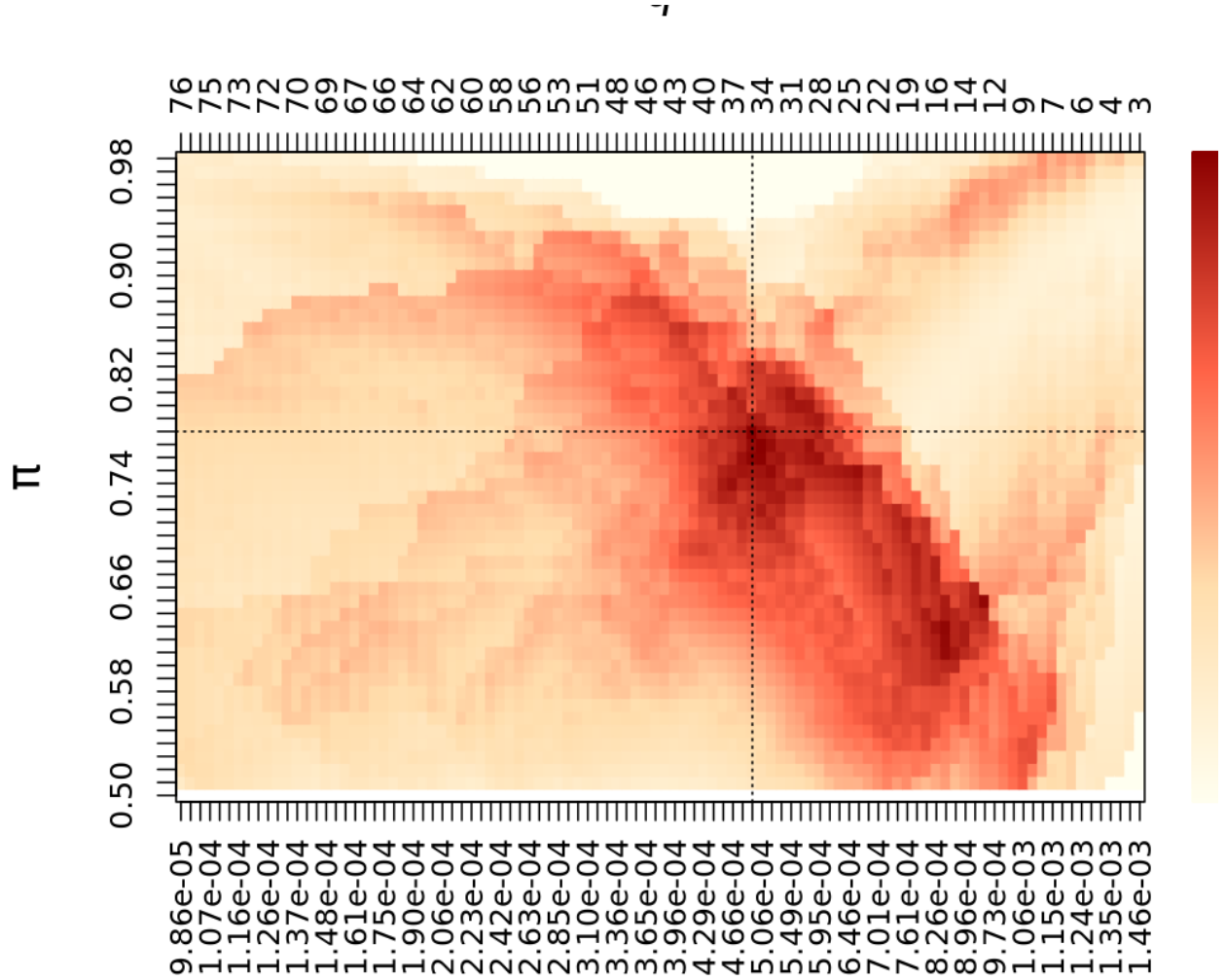
	Actual 0	Actual 1
Prediction 0	60610	1235
Prediction 1	28621	3184

Elastic net:

	Actual 0	Actual 1
Prediction 0	60453	1216
Prediction 1	28678	3303

Appendix

Stability Selection plot – direct effect



Appendix

Stability Selection plot – indirect effect