Uncovering Gut Microbiota's Causal Role in Major Depressive Disorder

Adèle H. Ribeiro[1†], Leon Fehse[1†], Nils R. Winter[2†], Marius Welzel[1],
Tilo Kircher[3], Sharmili E. Thanarajah[4], Udo Dannlowski[2], Dominik Heider[1*], Tim Hahn[2*]

1 - Institute of Computer Science, Heinrich-Heine University of Düsseldorf; 2 - Institute of Translational Psychiatry, Wilhelms-University of Münster;
3 - Department of Psychiatry and Psychotherapy, Philipp-University of Marburg; 4 - Department of Psychiatry, Goethe-University of Frankfurt

INTRODUCTION

Major Depressive Disorder (MDD) is a multifaceted mental health condition. Despite numerous studies highlighting a significant association between MDD and the gut microbiome, it remains unclear whether these associations play a causal role in MDD development.

In this study, we conducted a differential abundance analysis (DAA) followed by a causal analysis of the DFG FOR2107 dataset (https://for2107.de/), which includes 1,269 patients. Our goal is to employ tools under Judea Pearl's causality framework [4] to obtain insights into whether taxa identified as significantly associated with MDD could potentially be causes of MDD.

Causal discovery algorithms capable of handling hidden confounders [8, 1], paired with tools for identifying causal effects from their outputs [3, 2], have enabled a fully data-driven approach to causal inference. This approach offers a clear advantage over methods such as Mendelian Randomization, as all model hypotheses are inherently testable from observational data.

We highlight two important contributions:

- Through a meticulous application of the FCI algorithm, we identified that *Eggerthella* and *Hungatella* causally contribute to MDD, while *Coprobacilius* indirectly causes MDD via *Eggerthella*.
- Obesity not only affects MDD but also confounds the relation between taxa variables and MDD. Using effect identification tools [3, 2], we show the interventional probability of MDD increases with the abundance of *Eggerthella* and *Hungatella*.

MICROBIOME DIFFERENTIAL ABUNDANCE ANALYSIS (DAA)

We performed a DAA to identify gut microbiome taxa with significantly different abundance levels between healthy and individuals with MDD. We employed two different techniques, ZicoSeq (Yang and Chen, 2022) and LinDA (Zhou et al., 2022), while adjusting for the potential confounding of age, sex, BMI, site, and library size.

Table shows the genera identified by both techniques at a significance level of 5% after FDR correction.

| Genus	FDR-corr. p-values	
	LinDA	ZicoSeq
Hungatella	0.0002	0.0071
Eggerthella	0.0063	0.0071
Coprobacillus	0.0070	0.0071
Lachnospiraceae FCS020 group	0.0063	0.011

CAUSAL DISCOVERY VIA FAST CAUSAL INFERENCE (FCI) ALGORITHM

We used both the complete and conservative FCI algorithms to reveal causal relations [8, 1], available at the `pcalg` R package. Conditional independencies between mixed (i.e., discrete and continuous) variables were tested using the symmetric approach in [6].

To ensure robustness against violations of the *faithfulness* assumption, we assessed marginal causal consistency across Partial Ancestral Graphs (PAGs) obtained over all variable subsets, following [5]. Additionally, we rigorously tested and validated inferred PAGs for consistency with observed data on conditional independencies.

In a PAG, arrowheads and tails represent, respectively, non-ancestral (non-causal) and ancestral (causal) relationships. Bidirected edges represent spurious associations, while circles represent undetermined relations.

Figure shows a valid PAG obtained using both FCI and conservative FCI, based on conditional independence tests developed using LinDA. The PAG shows all robustly identified causal relationships between taxa variables and MDD. Lachnospiraceae FCS020 group was consistently linked to MDD by an edge ∘→, indicating it could either be a causal factor or merely spuriously associated with MDD.
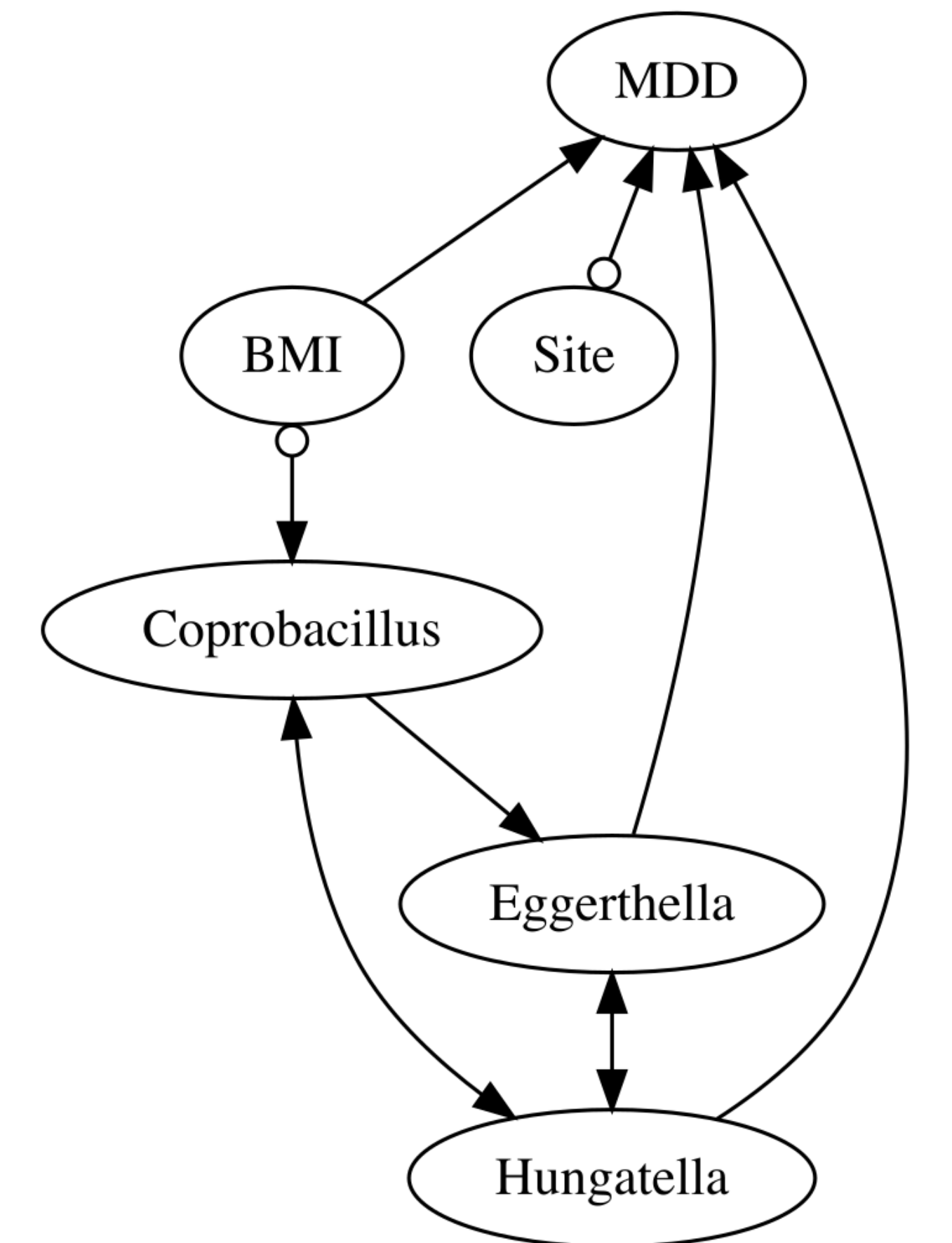
OBESITY-SPECIFIC CAUSAL EFFECTS OF *EGGERTHELLA* AND *HUNGATELLA* ON MDD

For easier interpretation, we categorized BMI values into three obesity categories.
Among the individuals, 746 were classified as Non-Obese, 334 as Obesity I, and 189 as Obesity II/III.

Obesity — Non-Obese BMI (14,25] — Obesity I BMI (25,30] — Obesity II/III BMI (30,60]
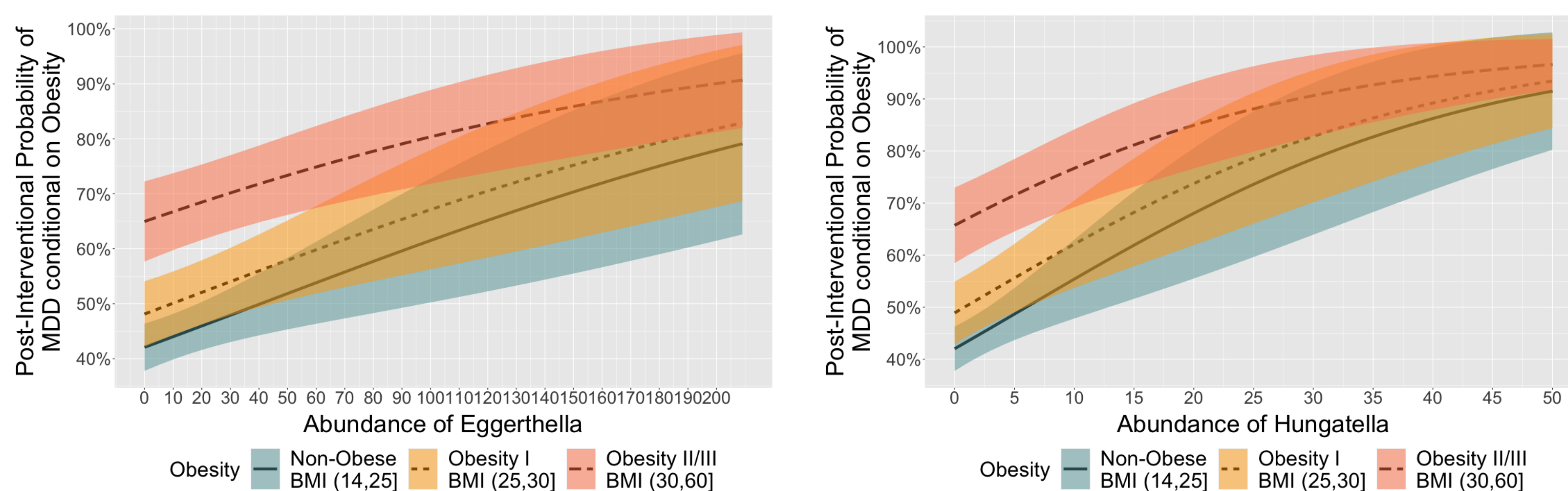
Examining the plots of post-interventional probabilities of MDD alongside the 95% confidence regions, a significant increase is evident with the abundance of both *Eggerthella* and *Hungatella*, regardless of the obesity group.

By the generalized backdoor criterion [3], the probability of developing MDD ($Y = 1$) after an intervention that sets the abundance levels of *Eggerthella* / *Hungatella* ($do(X = x)$), at a specified obesity condition ($C = c$), is:

$$P(Y=1|do(x),c) = \int_{\mathbf{z}} P(Y=1|x,c,\mathbf{z})P(\mathbf{z}|c)dbfz,$$

where $\mathbf{Z} = \{Hungatella\}$ when X is *Eggerthella* and $\mathbf{Z} = \{Eggerthella\}$ when X is *Hungatella*. Its estimation is carried out using a Monte Carlo method, with $P(Y = 1|x,c,\mathbf{z})$ modeled through logistic regression:

Parameter	Estimate	Std. Error	p-value
Intercept	-0.406	0.081	5.39e-07
Obesity I	0.259	0.134	5.29e-02
Obesity II/III	0.954	0.173	3.72e-08
Eggerthella	0.008	0.002	1.64e-03
Hungatella	0.054	0.015	3.47e-04

CONSIDERING DISCRETE LEVELS OF *EGGERTHELLA* AND *HUNGATELLA*
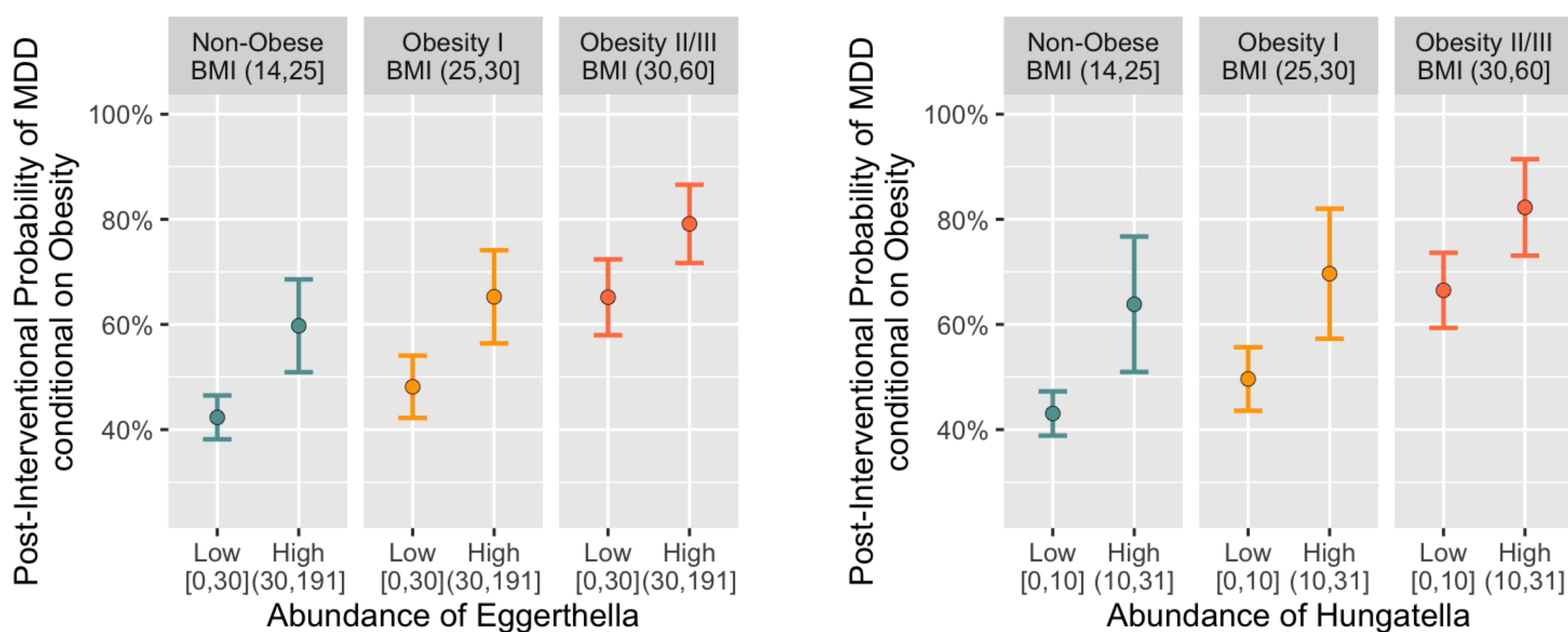
The obesity-specific effect, defined as $P(\text{MDD} = 1|do(X = \text{High}),c) - P(\text{MDD} = 1|do(X = \text{Low}),c)$ is as follows:

With X as *Eggerthella*:

Obesity	Effect	95% C.I.	
No	0.1742	0.07659	0.2718
I	0.1712	0.06460	0.2778
II/II	0.1397	0.03607	0.2433

With X as *Hungatella*:

Obesity	Effect	95% C.I.	
No	0.2078	0.07230	0.3433
I	0.2002	0.06261	0.3379
II/II	0.1577	0.04138	0.2741

The effect of *Eggerthella* is 0.1683 (CI: 0.06812, 0.2684), while that of *Hungatella* is 0.1984 (CI: 0.06607, 0.3307).

ACKNOWLEDGEMENTS

REFERENCES

[1] Colombo D, Maathuis MH. (2014). Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*

[2] Jaber A., Ribeiro A., Zhang J., and Bareinboim E. (2008). Causal identification under Markov equivalence: Calculus, algorithm, and completeness. *Adv. Neural Inf. Process. Syst., 35*

[3] Maathuis MH., Colombo D, (2015). A generalized back-door criterion. *Ann. Statist. 43 (3) 1060 - 1088.*

[4] Pearl, J (2000). Causality: Models, Reasoning, and Inference. *Cambridge University Press, New York, 2000. 2nd edition, 2009.*

[5] Anna Roumpelaki, Giorgos Borboudakis, Sofia Triantafillou, Ioannis Tsamardinos (2016). Marginal Causal Consistency in Constraint-based Causal Learning. *CFA@UAI 2016: 39-47*

[6] Tsagris, M., Borboudakis, G., Lagani, V. et al. (2018). Constraint based causal discovery with mixed data. *Int J Data Sci Anal 6*

[7] Yang L, Chen J. (2022). A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. *Microbiome., 10(1):130.*

[8] Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence, 172(16-17), pp.1873-1896.*

[9] Zhou H, He K, Chen J, Zhang X. (2022). LinDA: linear models for differential abundance analysis of microbiome compositional data. *Genome Biol., 23(1):95.*