

Causal Discovery in Gut Microbes for PCOS

Mariana Paco Mendivil
mpacomendivil@ucsd.edu

Candus Shi
c6shi@ucsd.edu

Nicole Zhang
nwzhang@ucsd.edu

Biwei Huang
bih007@ucsd.edu

Jelena Bradic
jbradic@ucsd.edu

Abstract

The human gut microbiome has become a significant factor in the understanding of metabolic health...

Code:

<https://github.com/nzhang20/Causal-Discovery-for-Biomedical-Applications>

1	Introduction	2
2	Methods	2
3	Results	2
4	Discussion	2
5	Conclusion	2
	References	3
	Appendices	A1

1 Abstract

The human gut microbiome has become a significant factor in understanding metabolic health, influencing conditions such as type 2 diabetes (T2D) and polycystic ovary syndrome (PCOS). Despite its recognized impact, much of the current research on the human gut microbiome and T2D remains limited to associative studies, leaving gaps in understanding the underlying causal relationships. This study addresses these gaps by enhancing causal discovery algorithms and integrating them with predictive modeling to investigate microbial contributions to T2D and PCOS.

Using cross-sectional datasets, including the Stanford T2D Gut Microbiome dataset and a PCOS Lancet dataset, we apply ...blah blah blah...

Our methodology incorporates ... blah blah blah

Our results reveal distinct microbial patterns associated with disease states and show the importance of causal discovery for understanding microbiome-disease interactions. This work provides a foundation for further research on treatment strategies influenced by the microbiome.

2 Introduction

The human gut microbiome has gained significant attention in recent years for its important role in metabolic health. While there has been extensive research that links the microbiome to health disorders such as type 2 diabetes (T2D) and polycystic ovary syndrome (PCOS), the majority of these studies remain correlational, leaving causal relationships undiscovered. Understanding these relationships is essential to improving and personalizing medical treatments for such diseases.

This study builds on recent advancements in causal discovery algorithms to investigate how microbial taxa influence metabolic disorders. Our goal is to find patterns that conventional association-based methods might miss by focusing on datasets with extensive metadata and utilizing tools made for high-dimensional, clustered data. Our work emphasizes the importance of addressing data complexities, such as clustering from varied study designs, to ensure robust causal inference.

[briefly talk about the methods we will use]

2.1 Literature Review

Polycystic ovary syndrome (PCOS) is a complex endocrine disorder linked to metabolic diseases such as obesity and type 2 diabetes (T2D). Current diagnostic methods use hormone and metabolic biomarkers, but these techniques are insufficient to differentiate between different PCOS subtypes, such as those characterized by hyperandrogenism. Due to inconsistent study findings, regional differences, and a lack of causative evidence, the association

between PCOS and gut microbiota is still unclear. This study aims to uncover specific gut microbiota characteristics in PCOS patients.

The study combined information from 14 clinical studies that satisfied strict inclusion requirements (e.g., 16S rRNA sequencing, reproductive-aged women, raw microbiota data). Statistical methods included PERMANOVA for beta diversity, Wilcoxon tests for alpha diversity, and NetMoss for microbiota network analysis. Regional and testosterone-level differences were also assessed.

From 563 retrieved studies, 14 were included (513 PCOS patients, 435 healthy controls). The main findings are as follows: PCOS patients had slightly lower levels of Bacillota and higher levels of Actinobacteriota; PCOS patients in China had lower alpha diversity, whereas PCOS patients in Europe had higher diversity; PCOS patients with high testosterone (HT) had different microbial patterns compared to those with low testosterone (LT), including lower levels of Faecalibacterium and higher levels of Prevotella; altered gut microbiota affected bile acid metabolism and inflammation, which led to PCOS pathogenesis.

With biomarkers like Prevotella and Faecalibacterium, PCOS subtypes have distinct gut microbiota compositions that are impacted by geography and testosterone levels. These results highlight the possibility of personalized treatments based on microbiota. However, to handle population variety and improve strain-level assessments, extensive, global research is required.

2.2 Data

To answer our research question, we used data from an individual participant data (IPD) meta analysis and systematic review conducted by (Yang et al. 2024). This means the dataset is an aggregation of the 14 studies that were included in the systematic review, but at the individual level. This is different from a meta analysis which analyzes aggregated data or statistics from multiple different studies. Each row of this PCOS dataset represents one sample of gut microbe abundance measurements as well as the sample's study's region, the sample's classification as a PCOS patient or a healthy control (HC), and if they were a PCOS patient, whether they had low (LT) or high (HT) testosterone levels. This granularity gives us more data and statistical power behind our results rather than using just one PCOS study. Although IPD meta analyses have these advantages as well as advantages over regular aggregated meta analyses, there are some glaring issues that must be addressed before we work with this dataset as is. In particular, the dataset is clustered because it consists of participants from different studies where each study had their own unique recruitment and sampling methods. Thus, we must account for clustering while running all analyses on this large aggregated dataset (Riley, Lambert and Abo-Zaid 2010). Two ways to do this in a typical meta analysis is via a fixed-effect model or a random-effects model (?). The choice of one over the other depends on the situation at hand, and since our PCOS dataset is comprised of 11 studies from China, 1 study from Poland, 1 study from Austria, and 1 study from Russia, the appropriate model appears to be a random effects model as there are many other areas of the world that we are interested in. (Not sure if any of this makes sense in the realm of causal discovery, but will leave here for now). (Alternatively) How-

ever, if we take a look at the scatter plots between pairs of microbes, there are no clear clusters representing different study regions or the individual studies. (Perhaps we can't see clusters because it's in such high dimension. We must run some clustering method to see if there is any clustering. Are the PCOS patients from different studies more alike or are all patients from one study whether PCOS or HC more alike than a different study?)

3 Methods

4 Results

5 Discussion

6 Conclusion

References

- Riley, Richard D, Paul C Lambert, and Ghada Abo-Zaid. 2010. "Meta-analysis of individual participant data: rationale, conduct, and reporting." *BMJ* 340. [\[Link\]](#)
- Yang, Yanan, Jiale Cheng, Chongyuan Liu, Xiaopo Zhang, Ning Ma, Zhi Zhou, Weiyang Lu, and Chongming Wu. 2024. "Gut microbiota in women with polycystic ovary syndrome: an individual based analysis of publicly available data." *eClinicalMedicine* 77. [\[Link\]](#)

Appendices