

國立陽明交通大學

統計學研究所

碩士論文

Institute of Statistics

National Yang Ming Chiao Tung University

Master Thesis

腸道微生物資料中介效應分析—

以帕金森氏症與便秘之因果關係探討為例

Exploring the Causal Effect of Constipation on Parkinson's
Disease Through Mediation Analysis of Microbial Data

研究生：吳佩樺（WU, PEI-HUA）

指導教授：林聖軒（LIN, SHENG-HSUAN）

中華民國一一一年七月

July 2022

腸道微生物資料中介效應分析—

以帕金森氏症與便秘之因果關係探討為例

Exploring the Causal Effect of Constipation on Parkinson's
Disease Through Mediation Analysis of Microbial Data

研究生：吳佩樺

Student：Pei-Hua Wu

指導教授：林聖軒 博士

Advisor：Dr. Sheng-Hsuan Lin

國立陽明交通大學

統計學研究所

碩士論文

陽明交大
NYCU

A Thesis

Submitted to Institute of Statistics

College of Science

National Yang Ming Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

July 2022

Taiwan, Republic of China

中華民國 一一一年七月

誌 謝

時光飛逝，轉眼間，於研究所的學習生涯即將進入尾聲。在交大統計所的兩年間滿是收穫，與同儕和師長們的回憶也是細數不盡，對我來說是一趟珍貴且難忘的旅程。

本篇論文能夠順利完成，首先要感謝的是林聖軒教授與傅詩宸博士的指導，在研究過程中悉心指導並給予寶貴的建議，也不斷地給予我肯定與包容，讓我在學術知識以及個人發展都獲得許多充實與成長。再者，也感謝百忙之中撥冗參與的口試委員們—李建樂教授以及王秀瑛教授，承蒙口試委員們對本篇論文的詳細審閱，提供諸多寶貴意見，使本篇論文更加完善嚴謹，在此致上最誠摯的謝意。

碩士班修業期間，感謝在林聖軒教授研究室中一起同甘共苦的夥伴們，尤其是 Microbiome 組的宜真與 BBS 組的林全和泓儒，在研究上互相砥礪及給予協助，私下也能一起玩樂、聊天交流，為我的研生活增添了許多歡樂的回憶。

此外，也要感謝我的家人們，不斷地給予我鼓勵和支持，讓我擁有持續努力的動力，也更堅定地朝向自己的目標前進。

這兩年的碩士生活雖然受疫情影響而有許多變數，過程中仍受到許多人的支持與關心，也對我的人生影響深遠。最後，謹以此向所有鼓勵我的人致上最深的謝意，並將這份成果與喜悅呈現給你們。碩士畢業也並不代表結束，而是另一個嶄新的開始，未來的我會更加努力，繼續學習與成長。

吳佩樺 謹誌於

國立陽明交通大學 統計學研究所

中華民國一一一年七月

本篇論文之部分內容已發表於 *Frontiers in Cellular and Infection Microbiology*，其連結如下：

https://www.frontiersin.org/articles/10.3389/fcimb.2022.871710/full?utm_source=dlvr.it&utm_medium=twitter

歡迎點擊下方連結或掃描下方 QRCode 觀賞本篇論文之說明影片：

https://www.youtube.com/watch?v=1H06f_flkTs



陽明交大
NYCU

中 文 摘 要

帕金森氏症 (Parkinson's disease, PD) 是一種全球性的慢性神經退化疾病，有鑑於世界人口快速老化，此疾病對全球所帶來的負擔也日益增加。帕金森氏症有許多危險因子，其中便秘症狀即為重要的危險因子之一。在 Braak 提出的腸腦軸假說中，胃腸道被認為是帕金森氏症的起源，並且有越來越多證據顯示腸道微生物變化在帕金森氏症的致病機制中具有一定的作用。在本研究中，我們旨在透過中介效應分析探討腸道微生物改變在便秘症狀引發帕金森氏症的機制中所扮演的角色，檢定及量化腸道微生物變化能夠解釋便秘症狀與帕金森氏症之間的因果效應之程度。在腸道微生物之相對豐度差異分析結果中，顯示有便秘症狀及無便秘症狀兩組受試者中，有 18 個菌屬、7 個菌種及 11 條功能路徑存在顯著差異。而在這些顯著的結果中，有 9 個菌屬、4 個菌種及 1 條功能路徑於便秘症狀引發帕金森氏症中具顯著中介效應。綜觀上述結果，我們也發現所有具有顯著中介效應之腸道微生物皆屬於與帕金森氏症有密切相關的短鏈脂肪酸 (short-chain fatty acid, SCFA) 生成菌。中介效應之結果亦顯示在便秘症狀引發帕金森氏症的效應中，被腸道微生物變化所中介的比例高達 76.56 %。我們的研究結果也闡明腸道微生態失衡在便秘症狀誘發帕金森氏症的致病機制中扮演極為重要的角色，其中主要都是透過減少短鏈脂肪酸生成菌。此結論更表明未來可以期待使用短鏈脂肪酸生成菌相關之益生菌預防或治療便秘症狀引發之帕金森氏症。然而，本篇研究存在三項研究限制。首先，原始資料數據集中並沒有提供其他可能必須校正之潛在干擾因子；再者，便秘症狀與帕金森氏症可能會相互影響，但本篇研究是基於便秘症狀為帕金森氏症之病因的假設所進行；最後，我們需要更進一步的研究以解釋便秘症狀引發帕金森氏症中剩餘的 23.44% 之效應。

關鍵字：帕金森氏症、便秘、微生物組、腸道微生物變化、中介效應分析

英 文 摘 要

Parkinson's disease (PD) is a worldwide neurodegenerative disease with an increasing global burden, while constipation is an important risk factor for PD. The gastrointestinal tract has been proposed as the origin of PD in Braak's gut-brain axis hypothesis, and there is increasing evidence indicating that intestinal microbial alteration has a role in the pathogenesis of PD. In this study, we aim to investigate the role of intestinal microbial alteration in the mechanism of constipation-related PD by mediation analysis, which can test and quantify the extent by which the intestinal microbial alteration explains the causal effect of constipation on PD incidence. Results of the microbial differences analysis showed that 18 bacterial genera, 7 species, and 11 functional pathways are significantly different between groups of constipated and non-constipated subjects. And among these bacteria and functional pathways, 9 genera, 4 species, and 1 pathway had a significant mediation effect between constipation and PD. In the above results, all of bacteria that had a significant mediation effect were short-chain fatty acid (SCFA)-producing bacteria that were substantially related to PD. Results from the mediation analysis showed that up to 76.56% of the effect of constipation on PD was mediated through intestinal microbial alteration. Our findings support that gut dysbiosis plays a critical role in the pathogenesis of constipation-related PD, mostly through the decreasing of SCFA producing bacteria, indicating that probiotics with SCFA-producing bacteria may be promising in the prevention and treatment of constipation-related PD. However, there are three limitations in this study. First, there are several potential confounders that should be adjusted were not provided in the original dataset. Second, our study was conducted based on the assumption of constipation being the etiology of PD; however, constipation and PD may mutually affect each other. Last, further studies are necessary to explain the remaining 23.44% effect leading to PD by constipation.

Keyword: Parkinson's disease, constipation, microbiome, intestinal microbial changes, mediation analysis

目 錄

中文摘要	i
英文摘要	ii
目錄	iii
圖目錄	v
表目錄	vi
第一章、前言	1
1.1 研究背景	1
1.1.1 帕金森氏症介紹與常見危險因子	1
1.1.2 便秘症狀與帕金森氏症之關係	1
1.1.3 腸道微生物之簡介	2
1.1.4 帕金森氏症與腸道微生物之關係	2
1.1.5 現有研究之缺漏	3
1.2 研究目的	3
1.3 論文架構	4
第二章、材料與方法	5
2.1 資料取得及道德規範聲明	5
2.2 資料採集方式及篩選	5
2.3 腸道微生物資料處理	6
2.3.1 16S rRNA 資料處理	6
2.3.2 功能路徑預測	7
2.4 統計分析方法	8
2.4.1 人口統計特徵	8
2.4.2 Alpha 多樣性及 Beta 多樣性	8
2.4.3 相對豐度差異分析	11
2.4.4 功能路徑豐度差異分析	12
2.5 中介效應分析	12

2.5.1	資料前處理	12
2.5.2	中介效應分析之簡介	13
2.5.3	單一中介因子之模型建立及效應估計	14
2.5.4	多重中介因子之模型建立及效應估計	17
第三章、	結果	20
3.1	人口統計特徵分析結果	20
3.2	Alpha 多樣性及 Beta 多樣性分析結果	24
3.3	相對豐度差異分析結果	26
3.4	功能路徑豐度差異分析結果	28
3.5	中介效應分析結果	29
第四章、	討論	32
4.1	相關文獻回顧	32
4.2	本篇研究與過往研究之呼應	32
4.3	腸道微生物菌屬層級之結果探討	33
4.4	功能路徑之結果探討	34
4.5	研究限制	34
4.6	未來展望	35
第五章、	結論	36
	參考文獻	37
	附錄（一）：資料集	45
	附錄（二）：程式碼—Alpha 多樣性	46
	附錄（三）：程式碼—Beta 多樣性	48
	附錄（四）：程式碼—火山圖	51
	附錄（五）：程式碼—中介效應分析	55

圖 目 錄

圖（一）：腸道微生物資料處理之流程圖	6
圖（二）： SS_T 之示意圖	10
圖（三）： SS_W 定義之示意圖	11
圖（四）：單一中介因子之因果關係圖	15
圖（五）：多重中介因子之因果關係圖	17
圖（六）：Alpha 多樣性各項指標之結果比較箱型圖	24
圖（七）：主座標分析散佈圖－基於非加權 UniFrac 距離計算	25
圖（八）：主座標分析散佈圖－基於加權 UniFrac 距離計算	25
圖（九）：主座標分析散佈圖－基於加權坎貝拉距離計算	26
圖（十）：有無便秘症狀兩組受試者之相對豐度差異（菌屬層級）	27
圖（十一）：有無便秘症狀兩組受試者之相對豐度差異（菌種層級）	28
圖（十二）：有無便秘症狀兩組受試者之微生物組的功能差異	29

陽明交大
NYCU

表 目 錄

表（一）：Lasso 與 Ridge 回歸模型之優缺點比較	18
表（二）：比較有無便秘症狀兩組受試者之人口統計特徵	20
表（三）：比較有無罹患帕金森氏症兩組受試者之人口統計特徵	22
表（四）：中介效應分析結果（腸道微生物之測量單位：菌屬層級）	30
表（五）：中介效應分析結果（腸道微生物之測量單位：菌種層級）	31
表（六）：中介效應分析（腸道微生物之測量單位：功能路徑）	31

陽明交大
NYCU

第一章、前言

1.1 研究背景

1.1.1 帕金森氏症介紹與常見危險因子

帕金森氏症（Parkinson's disease, PD）是一種影響中樞神經系統的慢性神經退化疾病，主要影響運動神經系統。其表現症狀可分為運動症狀（motor symptoms）及非運動症狀（non-motor symptoms），舉例來說，運動症狀有顫抖、肢體僵硬、動作遲緩、姿態不穩...等；非運動症狀則有便秘、快速眼動睡眠障礙、憂鬱...等(Rai, Chaturvedi et al. 2020, Rai and Singh 2020, Rai, Singh et al. 2021)。醫學上常會依照 Hoehn-Yahr 分級表（Hoehn and Yahr scale）(Hoehn and Yahr 1998)作為帕金森氏症之病情嚴重程度的評估，此分級方式主要分為五期，由較輕微的 1 期（單邊／側身體受影響，但沒有影響平衡）至較為嚴重的 5 期（在沒有他人幫助的情況下，只能臥床或坐輪椅）。帕金森氏症目前仍無法完全治癒，但是可利用左旋多巴（Levodopa）及諾康停（Entacapone）等藥物控制病情，有效減緩運動症狀發生之風險。相關研究顯示，帕金森氏症是目前全球第二常見的神經退化疾病，全球的盛行率約為 0.3%，在台灣則為 0.15%(黃靖鈞, 黃義騰 et al. 2018)，且盛行率也會隨著年齡增長而提升，有研究顯示 60 歲以上的人口每增加 1 歲，其盛行率便會提高 1%(De Lau and Breteler 2006)。除了年齡以外，也有其他研究指出性別、基因遺傳、飲食習慣...等也是帕金森氏症的危險因子，意即年長男性及有帕金森氏症之家族病史的人有較高機率會罹患帕金森氏症(Stirpe, Hoffman et al. 2016)。此外，環境風險因素也是值得注意的危險因子，目前相關實驗提出接觸過多殺蟲劑、除草劑及重金屬的化學物質也會增加罹患帕金森氏症的風險(De Lau and Breteler 2006)。抽菸習慣更是目前於最多研究被探討且少數獲得一致結果之危險因子，其研究結果皆指出有抽菸習慣者罹患帕金森氏症的風險會降低(De Lau and Breteler 2006)。

1.1.2 便秘症狀與帕金森氏症之關係

除了上述提及的幾點，便秘症狀也是帕金森氏症的危險因子之一，便秘症狀之判定並非單一的定義，大多是由多個症狀來定義是否有便秘症狀，如：糞便過硬、排便不順及排便次數過少（通常指一周少於三次）等(Lembo and Camilleri 2003)。過去就有研究證明帕金森氏症患者在診斷出典型的運動症狀之前，就會出現便秘症狀，甚至有案例在出現典型運動症狀的前 20 年就被觀察有便秘症狀(Savica, Carlin et al. 2009)。有一項根據臺灣

全民健康保險研究資料庫（Taiwan National Health Insurance Research Database）所進行的回溯式世代研究也指出在 551,324 個正常樣本中，隨著便秘症狀的嚴重程度的不同（輕度、中度至重度），罹患帕金森氏症的調整風險比依序為 3.28（95%信賴區間為 2.14 至 5.03）、3.83（95%信賴區間為 2.51 至 5.84）及 4.22（95%信賴區間為 2.95 至 6.05）(Lin, Lin et al. 2014)。另外也有一項統合分析（meta-analysis）發現在其收集的 741,593 個樣本中，有便秘症狀與沒有便秘症狀的受試者們，罹患帕金森氏症的總體勝算比約為 2.27（95%信賴區間為 2.09 至 2.46）；此外，若便秘症狀發生於罹患帕金森氏症前至少十年，其總體勝算比則為 2.13（95%信賴區間為 1.78 至 2.56）(Adams-Carr, Bestwick et al. 2016)。

1.1.3 腸道微生物之簡介

人體是由複雜的微生物基因體所構成，近年來有許多研究顯示微生物基因體對健康的影響相當明顯且重要。而人體胃腸道的腸道微生物群更是含有比人體細胞多 10 倍的微生物細胞，為一個複雜的生態群落，主要分為厚壁菌門（Firmicutes）、擬桿菌門（Bacteroidetes）、放線菌門（Actinobacteria）、梭桿菌門（Fusobacteria）、變形菌門（Proteobacteria）、疣微菌門（Verrucomicrobia）及藍細菌門（Cyanobacteria）七類，其中擬桿菌門及厚壁菌門之比例就高達 90% 以上(Huang, Liao et al. 2021)。腸道微生物對於人類生物學的各個方面深具影響，包括營養物質、維生素、藥物和有毒化合物的吸收和代謝、免疫系統的發展及預防病原體之入侵等，其主要透過細胞代謝活動及宿主交互作用（host interactions）影響正常生理及感病性（disease susceptibilities）(Lozupone, Stombaugh et al. 2012)，因此腸道微生物群組成之改變與許多疾病息息相關，如：發炎、代謝相關疾病、神經相關疾病及腫瘤(Arnold, Roach et al. 2016)。

1.1.4 帕金森氏症與腸道微生物之關係

此外，人體的腸道與大腦之間存在著緊密的關係，過去有學者提出「腸腦軸線」（Gut-brain axis）的理論(Braak, Del Tredici et al. 2003, Braak, Rüb et al. 2003)，腸腦軸線是一道連結中樞神經系統(Central Nervous System, CNS)及腸神經系統(Enteric Nervous System, ENS)之間的雙向橋梁，主要負責監測與整合腸道功能，並連結大腦的情緒與認知中樞及腸道外圍的功能與機制（如：腸道通透性、免疫活化...等）。還有相關實驗證據證明

腸道微生物對腸腦軸有重要的影響，不僅能直接與腸神經系統相互作用，也能透過神經內分泌及代謝路徑與中樞神經系統相互作用(Carabotti, Scirocco et al. 2015)。一般而言，腸道微生物群處於穩定平衡的狀態，但若人體服用抗生素、飲食習慣改變，甚至是罹患疾病時，腸道微生物群將會失去穩定，也就是出現腸道微生態失衡（dysbiosis of intestinal microbiota）之現象(Lozupone, Stombaugh et al. 2012)。

1.1.5 現有研究之缺漏

如上所述，目前有眾多證據顯示便秘症狀與帕金森氏症之間有強烈的關聯，但便秘症狀引發帕金森氏症的詳細機制仍尚未釐清。而其中一個致病的機制可能是透過腸道微生物的改變。目前至少有 17 項研究顯示帕金森氏症患者有腸道微生態失衡之現象(Keshavarzian, Engen et al. 2020)，其中有一項針對 36 位帕金森氏症患者進行為期兩年的追蹤性研究，其研究結果發現雙歧桿菌屬（*Bifidobacterium*）、脆弱類桿菌（*Bacteroides fragilis*）及柔嫩梭菌屬（*Clostridium leptum*）與帕金森氏症之嚴重程度有一定的關聯性(Minato, Maeda et al. 2017)。在近期發表的評論性文獻中，有提出數項腸道微生態失衡引發帕金森氏症的機制，例如：腸道微生物的改變會造成腸道障壁及血腦障壁的通透性增加、氧化壓力的增加及多巴胺生成變化等(Huang, Liao et al. 2021)。有趣的是，具有慢性便秘症狀的患者也會發生腸道微生態失衡的現象(Ohkusa, Koido et al. 2019, Zhang, Wang et al. 2021)，且這種便秘相關的腸道微生態失衡可透過比沙可啶（Bisacodyl）有效治療，比沙可啶為一種瀉劑，服用後可清空患者的腸道並改善便秘症狀(Khalif, Quigley et al. 2005)。另外也有一項老鼠試驗顯示便秘會誘發腸道微生態失衡，甚至加劇多發性硬化症(Lin, Liu et al. 2021)（Experimental autoimmune encephalomyelitis, EAE）。雖然有許多相關的證據顯示便秘症狀、腸道微生物變化及帕金森氏症之間是有關連的，但卻沒有一項研究去進行完整的分析以釐清腸道微生物變化在便秘症狀與帕金森氏症之間的機制所扮演的角色。

1.2 研究目的

為了研究便秘症狀、腸道微生物變化及帕金森氏症之間的關聯，本篇研究中，我們預計採用神經遺傳學研究協會（NeuroGenetics Research Consortium, NGRC）所蒐集之資料，根據過去的文獻，在此我們假設便秘症狀會透過誘發腸道微生物變化而導致帕金森氏症

的發病，進而建立中介效應模型以檢定及量化腸道微生物變化能夠解釋便秘症狀與帕金森氏症之間的因果效應之程度，透過探討該機制中重要的中介因子，找出具有顯著中介效應之腸道微生物，期許未來能作為帕金森氏症患者在服用抗帕金森氏症藥物的同時，用以輔助之益生菌治療藥物，對帕金森氏症之臨床治療有所貢獻。

1.3 論文架構

本篇論文於接下來之章節中介紹腸道微生物資料分析及其結果探討。首先，第二章將會依序介紹資料數據與其採集方法、腸道微生物資料處理方式及其常見統計分析方法，包括資料之人口統計特徵分析、Alpha 及 Beta 多樣性分析、相對豐度差異分析、功能路徑豐度差異分析，以及針對便秘症狀、腸道微生物變化及帕金森氏症所進行的中介效應分析。再者，第三章將敘述及解釋統計分析方法及中介效應分析之結果，並於第四章進行中介效應分析結果之討論以及說明未來可能進行的研究方向。最後，第五章則會針對本篇論文做出結論，並說明此研究對帕金森氏症之預防與治療所帶來之貢獻。

陽明交大
NYCU

第二章、材料與方法

2.1 資料取得及道德規範聲明

本篇研究所分析的序列資料皆可由歐洲核苷酸檔案庫（European Nucleotide Archive，ENA）取得，專案代碼為 ERP016332，且所有資料皆為開放取用及去識別化，因此無任何倫理審查之需求。

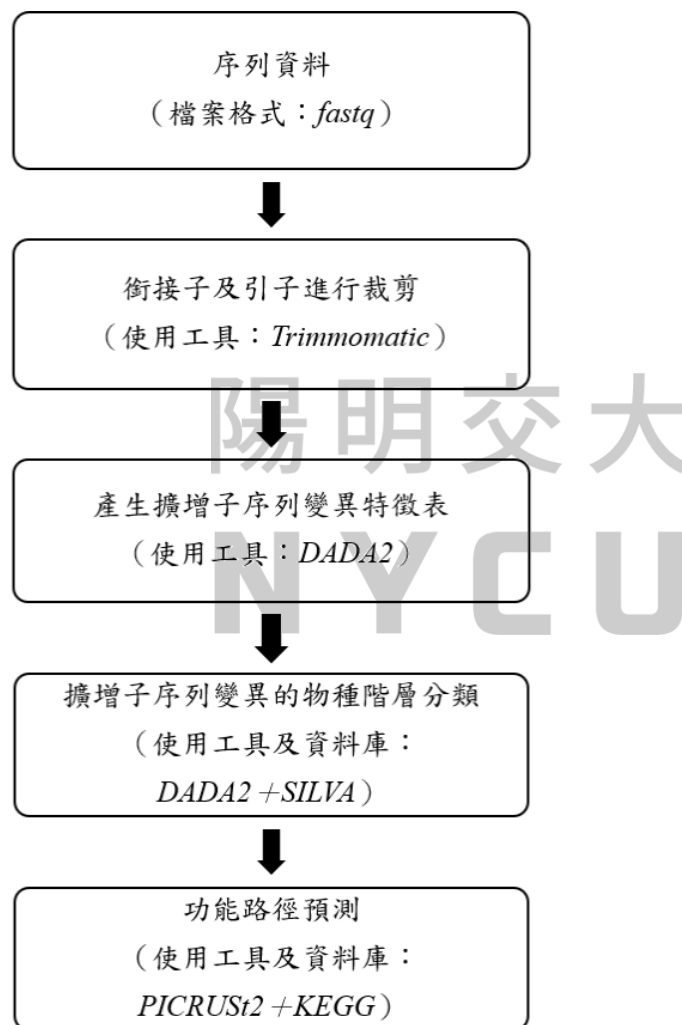
2.2 資料採集方式及篩選

本篇研究採用的資料來自 Hill-Burns 等學者在 2017 年所發表的論文(Hill-Burns, Debelius et al. 2017)，其中的 330 位受試者（185 位男性，145 位女性，平均年齡 69.2 歲）分別來自華盛頓西雅圖、喬治亞亞特蘭大及紐約阿伯尼，且皆為 2014 年 3 月至 2015 年 1 月期間神經遺傳學研究協會之成員。有關神經遺傳學研究學會資料集之相關方法學、臨床及基因相關特徵皆在 Hamza 等學者的論文(Hamza, Zabetian et al. 2010)中詳細敘述。至於資料採集方式，首先是由受試者於家中使用可拋棄式無菌拭子採集其糞便樣本，並立即將其以環境溫度保存下透過美國郵政署運送至實驗室(Hill-Burns, Debelius et al. 2017)。接著則在相同的實驗室進行 DNA 萃取，並將測序基因切成小片段後接上銜接子(adapter)及 16S 專屬引子(primer)，再進行聚合酶連鎖反應(Polymerase chain reaction, PCR)(Mullis, Faloona et al. 1986)，此為一種能在短時間內擴增 DNA 片段之複製技術，主要針對 16S rRNA 的高度變異區進行複製，以確保擴增後的序列(擴增子)，其長度至少能判定細菌種類至「屬」的等級，接著就使用定序機器(Illumina MiSeq)進行定序，採取效率較高及成本較低的次世代定序(Next Generation Sequencing, NGS)技術。

在該資料中的 330 位受試者當中，有 199 位(133 位男性，66 位女性，平均年齡 68.4 歲)是由行動障礙治療專家依據英國帕金森氏病學會腦庫建立的臨床診斷標準(UK Brain Bank criteria)之修訂版診斷為帕金森氏症的患者，至於剩下的 131 位控制組(52 位男性，79 位女性，平均年齡 70.4 歲)則是自我匯報無罹患任何的神經退化疾病。且受試者之間並無親屬血緣關係。資料中的便秘症狀紀錄則由受試者填寫問卷調查取得，由於其中 6 位受試者並未提供便秘狀態的任何資訊，因此本篇研究將此 6 位受試者排除，以剩下的 324 位受試者進行最後的資料分析。

2.3 腸道微生物資料處理

樣本之腸道微生物序列資料龐大且複雜，因此在進行後續分析前，首要步驟必須處理所採集之糞便樣本的 16S rRNA 資料。本章節將會分為兩部分介紹腸道微生物資料的處理方法：一為如何處理 16S rRNA 序列資料，包括定序方法及其菌種鑑定步驟和工具之介紹；二為透過標記基因序列預測人體代謝功能。其流程可參考圖（一），詳細介紹則於下面兩小節進行詳述。



圖（一）：腸道微生物資料處理之流程圖

2.3.1 16S rRNA 資料處理

微生物研究上的定序方法主要可分為兩大類：全基因組霰彈槍定序法（shotgun metagenomic sequencing）及目標擴增子定序法（targeted amplicon sequencing）。本篇研究中採用的方法為後者，其主要是針對 16S rRNA（細菌中是普遍存在的基因片段）中高度

變異區 (Highly Variable Region) 的序列進行定序，執行上相較方便、容易且迅速，但因為只拿特定片段的序列分析，物種分類的程度會受到限制。

序列檔案常見的格式是 fastq，為文字文件，其中包含序列 ID、鹼基序列及品質資訊，首先我們透過 Trimmomatic v0.39(Bolger, Lohse et al. 2014)將序列資料不需用到的銜接子及引子進行裁剪，接著則使用 DADA2 1.16 版本(Callahan, McMurdie et al. 2016)產生擴增子序列變異 (Amplicon sequence variants, ASVs) 之特徵表格。過去在標記基因序列的分析通常會從建構操作分類單位 (Operational taxonomic units, OTU) 叢集分群開始，雖然該作法能降低物種分類錯誤的情形，但同時也會造成一個操作分類單位無法完全代表單個物種，因此 Rosen 等人在提出 DADA 該演算法時(Rosen, Callahan et al. 2012)，選擇捨棄 OTU，改而採用解析度較高的擴增子序列變異，該演算法主要是藉由降噪 (denoising)，也就是更正擴增時定序錯誤的序列及移除嵌合體，進而建立擴增子序列變異，不同於操作分類單位，單一物種有可能會對應到一至多個擴增子序列變異。

採用 DADA2 進行 16S rRNA 資料處理的步驟如下：首先是透過繪圖觀察序列資料，並利用其中的 filterAndTrim 函數進行過濾及修剪，接著藉由 removeBimeraDenovo 函數將嵌合體移除，進而得到最終產物——擴增子序列變異特徵表。擴增子序列變異特徵表通常會透過 phyloseq 儲存在於 otu_table 中，此表格中的值是該擴增子序列變異在該樣本的絕對豐度 (absolute abundance)，也就是讀數 (reads) 數量。此外，我們也可以進一步透過 DADA2 及 SILVA 進行擴增子序列變異的物種階層分類，得到每個擴增子序列變異隸屬的界、門、綱、目、科、屬、種，而上述提及的 SILVA 為目前常見的微生物資料庫，是一個提供三大界 (細菌、古細菌、真核) 微生物 rRNA 基因序列的綜合資料庫，能處理的序列量相當大，且資料庫更新的頻率快，此篇研究使用的版本為 SILVA 132。

2.3.2 功能路徑預測

為了理解腸道微生物群是如何影響著人體內的代謝功能，我們利用擴增子序列變異特徵表，預測功能路徑在樣本體內之豐度，其後再參照京都基因與基因組百科全書 (Kyoto Encyclopedia of Genes and Genomes, KEGG) (Kanehisa and Goto 2000) 將功能路徑的 KO 編號轉換成常見的功能路徑名稱。而此部份我們主要是利用 PICRUSt2 (Phylogenetic

Investigation of Communities by Reconstruction of Unobserved States) 2.4.1 版本(Douglas, Maffei et al. 2020) 進行操作，PICRUSt 是一個基於標記基因序列預測功能豐度的軟體，於 2013 年發表並於 2020 年發表第二版 PICRUSt2，二版之資料庫相較一版擴增了十倍以上，使用上也不再限於 OTU，而是可以使用 ASV 的序列資料。

2.4 統計分析方法

經過微生物資料處理後，我們將由不同角度進行分析，分別為剖析有無便秘症狀兩組受試者之間的差異，及本篇研究最重要的中介效應分析。前者又可分為兩方面進行，一為針對此兩組受試者之人口統計特徵；二為針對此兩組受試者之腸道微生物資料，本章節將進行 Alpha 多樣性及 Beta 多樣性、菌種的相對豐度，及功能路徑的豐度之比較分析。

2.4.1 人口統計特徵

我們分別以有無便秘症狀及是否罹患帕金森氏症將受試者分為兩組，並各自比較兩組在人口統計特徵上是否有所差異，本篇研究中欲比較的數項特徵可分為四類，一為基本資訊，包含：年齡、性別、種族、身高、體重、身體質量指數 (Body Mass Index, BMI) 及地理區域 (經度、緯度及居住地)；二為生活習慣，包含：酒精攝取量及抽菸習慣；三為飲食習慣，包含：咖啡攝取量、是否有每日攝取蔬果、是否有每日攝取穀物、是否有每日攝取肉類、是否有每日攝取堅果類及是否有每日攝取優格；四為疾病狀況，包含是否罹患除了帕金森氏症外的其他神經性疾病。其中，針對連續型變數採用無母數方法中的 Wilcoxon 等級和檢定 (Wilcoxon Rank Sum Test)；類別型變數則採用卡方檢定 (Chi-square test)，並將結果以表格呈現。

2.4.2 Alpha 多樣性及 Beta 多樣性

為了解釋便秘症狀與腸道微生物群之間的關聯性，利用不同角度去看便秘症狀到腸道微生物群這條路徑是否存在。第一個需進行分析的即為 Alpha 多樣性，Alpha 多樣性用於反映單一環境內微生物豐富度 (Richness) 及均勻度 (Evenness) (Whittaker 1972)，其中豐富度指的是一個特定環境中有多少物種，而均勻度則是一個特定環境中物種間比例的關係 (分布均勻性)。本篇研究中利用擴增子序列變異估計四種常見的 Alpha 多樣性指標，分別為 Observed、Chao1 (Chao 1984)、Shannon (Magurran 1988) 及 Simpson (Rosenzweig 1995)，前二者僅用到豐富度的概念，值越大代表 Alpha 多樣性越

大，Observed 更是計算上最簡單之指標，也就是計算樣本中擴增子序列變異的種類數量，Chao1 則又再多考慮了稀有種之權重；後兩者則同時考慮了豐富度及均勻度的概念，其值的範圍皆為 0 至 1 之間，值越大亦代表 Alpha 多樣性越大。以上指標皆使用 R 套件 phyloseq 1.32.0(McMurdie and Holmes 2013)的 `estimate_richness` 函數計算得出，接著則採用 ANOVA 檢定查看有無便秘症狀的兩群人的 Alpha 多樣性是否有差異，再採用箱型圖呈現其結果，以清楚表達兩組受試者於各指標之間的分佈差異，該圖則是使用 R 套件 phyloseq 1.32.0 的 `plot_richness` 函數計算得出，其程式碼請參考附錄(二)。

第二個需進行分析的為 Beta 多樣性，Beta 多樣性則為衡量「環境」與「環境」間物種組成之差異的指標，在此可以將其解釋作比較兩群樣本間腸道微生物組成之相似程度。在 Beta 多樣性之計算上，通常會利用根據物種組成的差異計算單一樣本間的距離矩陣接著利用無母數方法中的 PERMANOVA (Permutational multivariate analysis of variance) 檢定 (又可稱作 ADONIS 檢定)，檢定群之間的距離是否相近，亦即檢定兩群的相似程度。首先，本篇研究所採用的距離計算方法可分為兩類：一類為基於系統發育樹 (Phylogenetic tree) 下，也就是考慮不同擴增子序列變異間的親緣關係下計算的距離矩陣，常見的 UniFrac 距離 (Unique Fraction distance) 便是此類，而其又可再依「是否採計群體規模」細分為加權 UniFrac 距離 (Weighted UniFrac distance) 及非加權 UniFrac 距離 (Unweighted UniFrac distance) (Lozupone, Lladser et al. 2011)，以上距離矩陣可使用 R 套件 phyloseq 1.32.0(McMurdie and Holmes 2013)的 `UniFrac` 函數計算得出；另一類之距離計算方法則是未基於系統發育樹，亦即將各個擴增子序列變異間視作無親緣關係下計算的距離矩陣，坎貝拉距離 (Canberra distance) (Lance and Williams 1967)便是此類，該距離矩陣則可使用 R 套件 vegan 2.5.7 的 `vegdist` 函數計算得出。在計算完距離後，我們則利用上述所提及的 PERMANOVA 檢定查看有無便秘症狀的兩群人在腸道微生物組成上是否有所差異，PERMANOVA 檢定是一種針對距離矩陣的多變量變異數分析 (MANOVA)，其統計檢定量為 F 值，該值之計算如下：

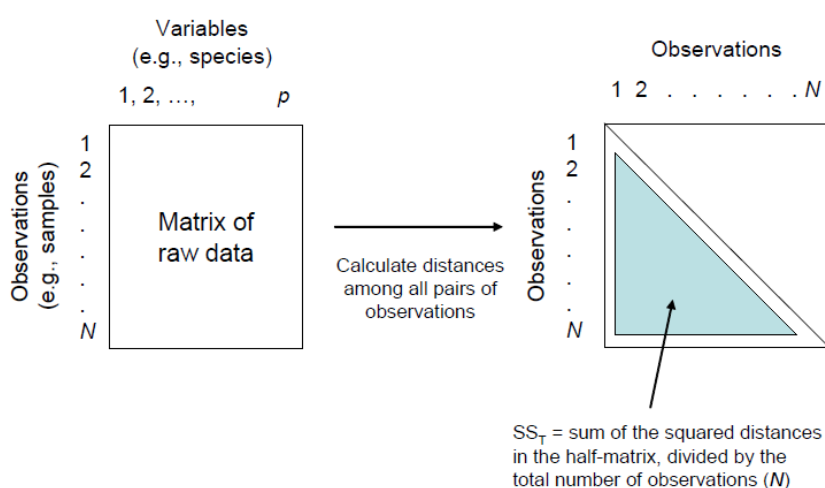
$$F = \frac{SS_A/(a - 1)}{SS_W/(N - a)}$$

其中， SS_A 為組間的距離平方和，由 $SS_T - SS_W$ 計算而得出， SS_T 為下三角矩陣 (或上三角

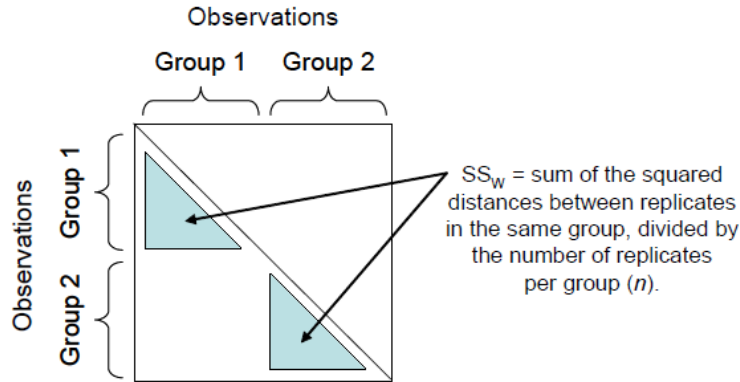
矩陣)的距離平方和除以觀察值總數(N)； SS_W 為兩觀測值在同組別的距離平方和(組內的距離平方和)除以每組的置換數(n)，兩者示意圖如圖(二)及圖(三)所示(Anderson 2005)。接著再透過對樣本進行隨機置換(permutation)，計算各組合之 F 值，取得檢定統計量之分布，進而求得 p -value：

$$p - value = \frac{(No. of F^{\pi} \geq F^0) + 1}{(Total no. of F^{\pi}) + 1}$$

其中 F^{π} 為每次隨機置換後所求得的 F 檢定統計量，而 F^0 則為依照實際資料(原始分組方式下)計算之檢定統計量。此檢定方法與傳統的多變量變異數分析相比，有兩項優勢，分別為不需要常態分佈之假設以及可使用由任何距離度量計算而出的距離矩陣進行分析，更適合運用在生態學資料。而這部份我們則是使用R套件vegan 2.5.7中的adonis函數計算得出。同時，在Beta多樣性之呈現上，我們經常使用一種微生物生態學中特定類型的降維方法，稱為主座標分析(Principle Co-ordinate Analysis, PCoA)，該分析是透過一系列的特徵值和特徵向量排序從多維資料中找出最主要的元素和結構，基於不同的距離量測值進行並選取排名前二高的維度繪製散佈圖以利觀察(Palmer 2004)。Beta多樣性之程式碼請參考附錄(三)。



圖(二)： SS_T 定義之示意圖(Anderson 2005)



圖（三）： SS_W 定義之示意圖(Anderson 2005)

2.4.3 相對豐度差異分析

接著，我們將針對有無便秘症狀的兩群人進行腸道微生物的豐度差異分析（Differential abundance analysis），豐度差異分析是微生物組數據統計分析的核心，進行該分析的目的在于找出有無便秘症狀的兩群人中豐度有明顯差異的腸道微生物，也可以藉此確定便秘症狀會造成腸道微生物變化，但在進行分析前，由於樣本本身的差異及測序過程涉及的複雜化學性質都有可能影響樣本的總豐度，因此我們會先針對擴增子序列變異的絕對豐度表格（Absolute abundance table）進行正規化（Normalization）以移除上述的差異。常見的正規化方法有兩種，一為稀疏化（rarefaction），也就是使每個樣本序列抽取相同數量的序列讀數，讓每個樣本維持於同樣的資訊水平，但此做法可能會使某些資訊被遺漏，也有文獻提出該方法並不理想，因為其統計檢定力可能會依據數據的刪除數量而降低(Weiss, Xu et al. 2017)；因此本篇研究就採用另外一種正規化方法—總和比例化（Total-sum scaling）(Lin and Peddada 2020)，就如我們於章節 2.3 所提到的，我們利用 DADA2 所獲得的擴增子序列變異特徵表中的值為該擴增子序列變異在樣本的絕對豐度，而總和比例化透過計算擴增子序列變異在該樣本的絕對豐度除上該樣本的總豐度，將絕對豐度表格轉換為相對豐度表格（Relative abundance table），以移除上述提及的樣本總豐度不一致的疑慮。接著則採用無母數方法中的 Wilcoxon 等級和檢定（Wilcoxon Rank Sum Test），檢定有無便秘症狀的兩組人的腸道微生物之豐度是否存在差異。而最後我們使用火山圖（volcano plot）呈現此檢定結果，其程式碼請參考附錄（四）。火山圖以 $\log_2(\text{變化倍數})$ 為橫軸， $-\log_{10}(p\text{-value})$ 為縱軸，將兩組受試者的相對豐度差異之結果視覺化，其中的變化倍數是透過計算該擴增子序列變異於有便秘症狀組別的平均相對豐度除上該擴增子序列變異於無便秘症狀組別的平均相對豐度而得出，p-value 則為

Wilcoxon 等級和檢定之結果。值得注意的是，在計算變化倍數時，可能會發生該菌屬（或菌種）於無便秘症狀組別的平均相對豐度為 0 的狀況，進而造成變化倍數發散至無限大，因此我們選擇對所有的擴增子序列變異之相對豐度加上一個極小值（ 2×10^{-6} ，此值由所有相對豐度的最小值除上 100 計算得出）以解決上述問題。

2.4.4 功能路徑豐度差異分析

經由章節 2.3.3 之資料處理後，我們取得功能路徑相對豐度表格，接下來我們將比較有便秘症狀及無便秘症狀兩組受試者在功能路徑豐度是否有差異。這部份我們採用無母數方法中的 White's 無母數 t 檢定（White's non-parametric t-test），此為 White 等人於 2009 年為臨床宏基因組資料提出之無母數檢定方法(White, Nagarajan et al. 2009)，主要是透過置換過程移除傳統 t 檢定（t-test）所需之常態分佈假設，而這邊我們將其假設為一雙尾檢定，且重複次數設為 1000 次，並把檢定結果之 p-value 使用 Storey FDR 校正，校正後之值稱為 q-value，以 0.05（5%）作為顯著水準，若 q-value < 0.05，則表示有無便秘症狀兩組受試者在該功能路徑之相對豐度有顯著差異，反之則否。以上分析我們皆使用 STAMP（Statistical Analysis of Metagenomic Profiles）進行並繪製圖表呈現最終結果。

2.5 中介效應分析

此章節我們將進行腸道微生物資料中介效應。並會分為四個部分詳細介紹：首先是資料前處理，由於腸道微生物資料維度較高且相對複雜，因此在建立模型之前對其做特定方法之處理是必要的；再者是中介效應分析之簡介，介紹中介效應分析之背景及基本名詞解釋；接下來會單一中介因子之模型建立與效應估計，將探討各個腸道微生物分別參與的中介效應；最後則是多重中介因子之模型建立與效應估計，探討所有腸道微生物所參與之整體中介效應。

2.5.1 資料前處理

腸道微生物資料（擴增子序列變異特徵表或是功能路徑資料）本身是一種組合資料（Compositional data），恆和限制式（Constant-sum constraint, CSC）是組合資料的一個屬性，也就是所有變數在同一樣本中的總和為一單位值（1 或 100%）(Kucera and Malmgren 1998)，但也因為有上述的限制式，我們可得知腸道微生物之變數間並非獨立的，只要其中一個菌種的相對豐度提高，必定有其他菌種的相對豐度會降低，又加上腸

道微生物組合資料是由多個稀疏向量 (Sparse vector) 所組成，其中稀疏向量為其中有許多元素值為 0 的向量，因此變數分布之偏度為正 (亦即右偏)，這並不符合後續建立模型的基本假設(Aitchison 1982)，因此我們需要對其做對數轉換以解決上述問題(Lin and Peddada 2020)，但是就如前面提到的，變數向量中有許多值為 0，所以再進行對數轉換前，我們一樣需對所有值加上一個極小且大於 0 的常數值，稱為偽計數 (Pseudo-count) (Xia, Chen et al. 2013, Mandal, Van Treuren et al. 2015)，而在本篇研究中，我們就以所有相對豐度的最小值再除以 100 所得到的值作為偽計數，即 2×10^{-6} ，針對擴增子序列變異加上該偽計數後進行對數轉換，上述之轉換公式舉例如下：

$$M = \log(\text{擴增子序列變異之相對豐度} + 2 \times 10^{-6})$$

轉換後再合併至菌屬及菌種的層級，功能路徑的處理方式也與上述做法一致。此外，為了確保我們留下的腸道微生物是出現在大部份樣本中的，我們參照 Hill-Burns 該篇論文中的做法(Hill-Burns, Debelius et al. 2017)，僅保留至少出現在 10% 樣本的擴增子序列變異及功能路徑，則個數分別剩下 283 個及 102 個，其中我們再針對擴增子序列變異合併至菌屬和菌種的層級，個數則會剩下 107 個及 79 個。

2.5.2 中介效應分析之簡介

如在前言中所提及的，假設便秘症狀會造成誘發腸道微生物變化，進而導致帕金森氏症的發病，而我們將會進行中介效應分析以量化腸道微生物變化對便秘症狀與帕金森氏症之間的因果效應之解釋程度。中介效應分析為統計分析中的新興領域，其主要目的是探討因果效應透過哪些機制所造成，以及每種機制所解釋的比例為何。傳統的差異法與結構方程式局限於線性模型；中介效應分析突破此一限制，得以廣泛應用於公共衛生以及生物醫學的研究領域，像是應用在解釋疾病發生的致病機制等，舉例來說，我們都知道抽菸可能會導致罹患肺癌進而造成死亡，但是除了肺癌之外，抽菸也有可能是透過其他機制造成死亡，像是口腔癌等，這時我們就可以透過中介效應分析探討肺癌是否為一個重要的中介因子。中介效應分析最早由 Baron 與 Kenny 在 1986 年提出，而傳統的中介效應分析，主要著重在單一中介因子，檢定暴露因子 (Exposure) 與結果變量 (Outcome) 的因果關係中，該中介因子所參與的效應大小 (Baron and Kenny 1986, MacKinnon 2012)。隨著資料蒐集技術的成熟，可以獲得更多的變量資料，Boca 等人就在 2014 年提出多重

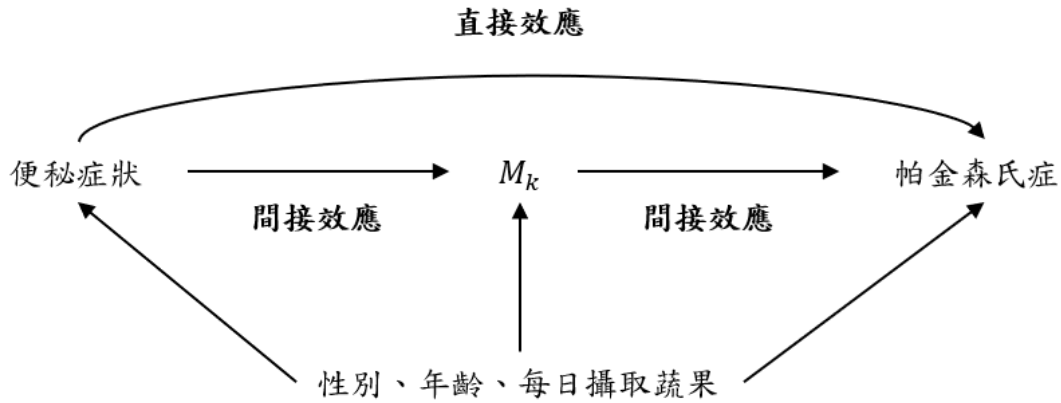
中介效應分析方法(Boca, Sinha et al. 2014)，而在多重中介因子下，又可依照中介因子之間關係的獨立性與否分成「平行中介因子(Parallel mediator)」及「有序中介因子(Ordering mediator)」，前者為相互平行的多重中介因子，後者則為橫向多重且隨時間變化的中介因子。在本篇研究中，便秘症狀及帕金森氏症分別為暴露因子及結果變量，而根據中介之定義，中介因子成立之條件需滿足以下兩項因果關係：中介因子為暴露因子之結果以及結果變量為中介因子之結果，因此我們於簡介中引述多篇文獻證明便秘症狀會造成腸道微生物變化以及腸道微生物之變化會引發帕金森氏症兩項因果關係，所以腸道微生物變化便為此篇研究中欲探討的中介因子。

我們將暴露因子對結果變量造成的影響稱為總效應(Total effect)，並透過中介效應分析將其分解為兩部分，分別是暴露因子透過中介因子對結果變量造成的效應，稱為間接效應(Indirect effect)，又可稱之為中介效應(Mediation effect)；及暴露因子未透過中介因子對結果變量所造成的效應，稱為直接效應(Direct effect)，又可稱之為替代效應(Alternative effect)。於本篇研究中，總效應即為便秘症狀對帕金森氏症的因果效應；間接效應(中介效應)為便秘症狀透過腸道微生物變化引發帕金森氏症之效應；直接效應則為便秘症狀不透過腸道微生物變化引發帕金森氏症之效應(可能是透過其他機制造成之效應)。並利用間接效應(中介效應)與總效應的比值，也就是中介比例(Proportion mediated, PM)，衡量中介因子所解釋的比例。如果比例越接近 1，代表該中介因子所代表的路徑可以完全解釋總效應所有的機制；如果比例接近 0，則代表該中介因子所代表的路徑並不是總效應的機制。此外，在估計直接效應與間接效應(中介效應)時，需建立在幾項假設上，2009 年及 2010 年就有學者提出中介效應分析所需要的辨識假設，其中包括暴露因子、中介因子及結果變量之間所有的干擾因子(Confounder)都要被測量且校正(VanderWeele and Vansteelandt 2009, Imai, Keele et al. 2010)，根據之前研究發現，性別、年齡及是否有每日攝取蔬果這三個變數都會影響便秘症狀、帕金森氏症及腸道微生物(De Lau and Breteler 2006, Yatsunenko, Rey et al. 2012, Xu and Knight 2015, Hill-Burns, Debelius et al. 2017, Picillo, Palladino et al. 2021)，因此我們將其設為中介效應分析中需進行校正的干擾因子。

2.5.3 單一中介因子之模型建立及效應估計

首先，我們將探討單個腸道微生物之變化於便秘症狀引發帕金森氏症的機制中所參與

的中介效應，其各自之因果關係圖如圖（四）所示：



圖（四）：單一中介因子之因果關係圖：假設共有 K 個腸道微生物，其中的 M_k 代表第 k 個腸道微生物之變化

為了估計其中介效應，我們首先需建立以下三個模型：

$$\Pr(Y = 1|A, C) = \alpha_0 + \alpha_A A + \alpha_C C \quad (\text{模型一})$$

$$\Pr(Y = 1|A, C, M) = \omega_0 + \omega_A A + \omega_M M_k + \omega_C C, k = 1, \dots, K \quad (\text{模型二})$$

$$E[M_k | A, C] = \beta_0 + \beta_A A + \beta_C C, k = 1, \dots, K \quad (\text{模型三})$$

其中 Y 為帕金森氏症， A 為便秘症狀， $(M_1, M_2, \dots, M_K)$ 分別代表 K 個腸道微生物（又分為三個測量單位：菌屬（genus）層級、菌種（species）層級及功能路徑）， C 為干擾因子向量（包含性別、年齡及是否有每日攝取蔬果三個變數）。

於模型一中，我們建立一個以帕金森氏症為應變數（Dependent Variable）而便秘症狀與干擾因子向量為自變數（Independent Variable）的羅吉斯回歸模型（Logistic regression model）。而模型二一樣是一個以帕金森氏症為應變數的羅吉斯回歸模型，但自變數除了便秘症狀與干擾因子向量以外，另外加上了各個腸道微生物，由於有 K 個腸道微生物，因此我們需建立 K 個模型二，且其係數（ ω_0 、 ω_A 、 ω_M ）皆不同。同樣地，

我們也需要建立 K 個模型三，以各個腸道微生物為應變數而便秘症狀與干擾因子為自變數之線性回歸模型（Linear regression model），且其係數（ β_0 、 β_A 、 β_C ）一樣皆不同。

模型解釋上，模型一中便秘症狀的估計係數（ $\widehat{\alpha}_A$ ）為便秘症狀對帕金森氏症的效應；每個模型二中便秘症狀的估計係數（ $\widehat{\omega}_A$ ）為直接效應（不透過各腸道微生物變化的中介下便秘症狀對帕金森氏症的效應）；每個模型三中便秘症狀的係數（ $\widehat{\beta}_A$ ）則為便秘症狀對各個腸道微生物的效應。此篇研究中，單一中介因子之效應（便秘症狀透過各個腸道微生物變化進而影響帕金森氏症的效應）是透過差異法（Difference method）測量，差異法為中介效應分析中常見的係數估計方法之一，且常用於流行病學領域上。而差異法的應用如下：當我們建立兩個模型，分別為加入中介因子及未加入中介因子，如下所示：

$$E[Y|A = a, C = c] = \theta_0 + \theta_1 a + \theta_c c$$

$$E[Y|A = a, M = m, C = c] = \theta_0^* + \theta_1^* a + \theta_3^* m + \theta_c^* c$$

其中 A 為暴露因子；Y 為結果變量；M 為中介因子；C 為干擾因子，中介效應則為暴露因子之估計係數相減，也就是 $\widehat{\theta}_1 - \widehat{\theta}_1^*$ (VanderWeele 2015)。而在本篇研究所建立的模型，單一中介效應便是以模型一及各個模型二中便秘症狀的估計係數差（ $\widehat{\alpha}_A - \widehat{\omega}_A$ ）作為其估計值，中介比例則為 $\frac{\widehat{\alpha}_A - \widehat{\omega}_A}{\widehat{\alpha}_A}$ 。此外，該中介效應的標準誤（Standard error）則是使用 Delta 方法（Delta Method）估計，此方法主要是假設若存在一個隨機變數序列服從以下分布：

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

假設 g 為一連續函數，且於 θ 可微分（意即 $g'(\theta)$ 存在且不為 0），則：

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} N(0, \sigma^2 \cdot [g'(\theta)]^2)$$

因此，我們可得知該中介效應之標準誤為 $\sqrt{\sigma^2_{\widehat{\alpha}_A} + \sigma^2_{\widehat{\omega}_A}}$ 。

最後，單一中介因子效應之 p-value 則是使用聯合顯著檢定 (Joint significance test) (Huang 2018) 計算，意即假設沒有未觀測的干擾因子 (Unmeasured confounder)，在常態假設下計算便秘症狀至各腸道微生物之效應 ($\widehat{\beta}_A$) 及各腸道微生物至帕金森氏症之效應 ($\widehat{\omega}_M$) 的 p-value，以兩者中較大的值作為完整路徑之 p-value，該檢定方法不僅相當具有統計檢定力，在計算上也是具備有效性。以上的中介效應分析皆使用 R 軟體執行，羅吉斯回歸模型及線性回歸模型則使用其中的 stats 套件 4.1.0，其程式碼請參考附錄 (五)。

2.5.4 多重中介因子之模型建立及效應估計

接著，我們將討論所有腸道微生物變化同時於便秘症狀引發帕金森氏症的機制中所參與的中介效應，其因果關係圖如圖 (五) 所示：

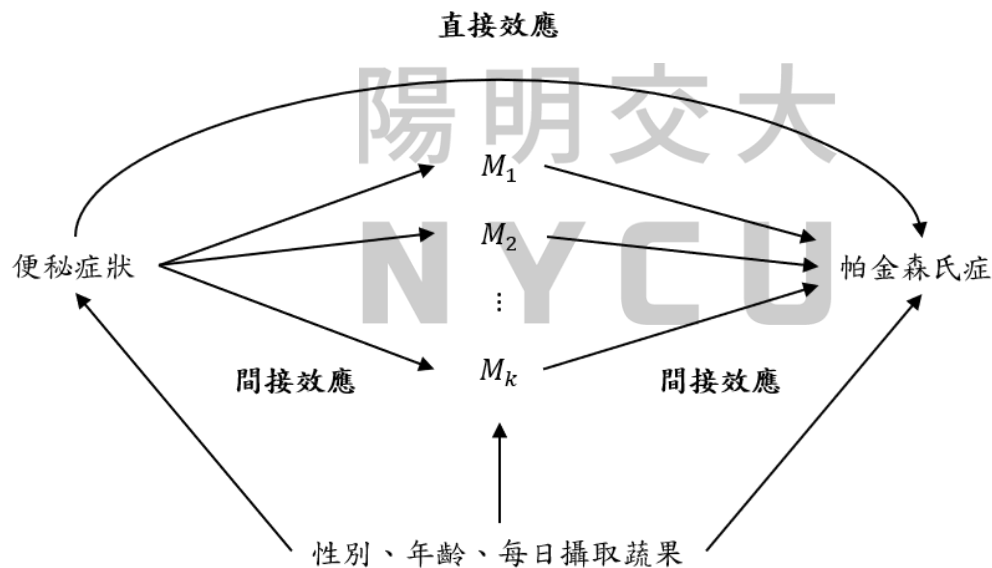


圖 (五)：多重中介因子之因果關係圖：假設共有 K 個腸道微生物，其中的 M_k 代表第 k 個腸道微生物之變化

此部分的中介效應估計，除了於單一中介效應估計中所建立的模型一，我們還需要另外建立另一個模型如下：

$$\Pr(Y = 1 \mid A, M, C) = \gamma_0 + \gamma_A A + \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 M_3 + \cdots + \gamma_K M_K + \gamma_C C \quad (\text{模型四})$$

其中 Y 為帕金森氏症， A 為便秘症狀， $(M_1, M_2, \dots, M_K)$ 代表 K 個腸道微生物（又可分為三個測量單位：菌屬層級、菌種層級及功能路徑）， C 為干擾因子向量（包含性別、年齡及是否有每日攝取蔬果三個變數）。

而模型四為一個以帕金森氏症為應變數的回歸模型，但自變數除了便秘症狀與干擾因子以外還加上所有的腸道微生物。值得注意的是，經由章節 2.5.1 前處理後的菌屬個數、菌種個數及功能路徑之個數分別為 107 個、79 個及 102 個，這也表示我們所設定的中介因子為高維度的變數向量，若使用簡單線性回歸模型將會產生模型過度配適的問題而造成係數估計之變異數過大，因此針對模型二我們採用正規化回歸模型（Regularized regression model），也就是針對係數估計的目標函數加上懲罰參數（Penalty parameter），透過懲罰參數的限制有效地降低係數波動與不穩定性，意即降低模型的變異程度。依懲罰參數的不同，正規化回歸模型又可再分做 Lasso 回歸模型（Lasso regression model）及 Ridge 回歸模型（Ridge regression model），此外，這兩種回歸模型對於不相干的變數之處理方式也不同，Lasso 回歸模型中會強制較無影響力的變數之估計係數為 0，以進行變數篩選；但在 Ridge 回歸中，較無影響力的變數之估計係數則是趨近於 0，也可以降低資料集中的雜訊。此兩種回歸模型之優缺點比較可參考下表（一）。

表（一）：Lasso 與 Ridge 回歸模型之優缺點比較

	Lasso 回歸模型	Ridge 回歸模型
優點	1、降低係數波動與不穩定性 2、具變數篩選之功能	1、降低係數波動與不穩定性 2、降低資料集中的雜訊
缺點	1、移除變數的同時會犧牲模型之正確性	1、不具變數篩選之功能

基於較無影響力的腸道微生物仍帶有些許資訊供後續分析的參考，不應完全忽略的考量下，我們選擇 Ridge 回歸模型(Saleh, Arashi et al. 2019)。Ridge 回歸模型中，係數估計之目標函數如下：

$$(Y - X\beta)^T(Y - X\beta) + \lambda\|\beta\|^2$$

可以透過最小化該目標函數得到估計參數：

$$\hat{\beta} = (X^T X + \lambda I_p)^{-1} X^T Y$$

此處我們透過拔靴法（Bootstrap）進行模擬以取得統計量的近似分配，而經由我們實驗發現，其分布在模擬 500 次至 1000 次時接近收斂，因此最終我們採用模擬 1000 次之結果，其中的 λ 值則是透過交叉驗證法（Cross-validation）篩選最適值，選取交叉驗證中使得均方誤差（Mean-square error, MSE）達到最小的 λ 值。值得注意的是，針對我們想討論之變數（例如：便秘症狀以及在單一中介因子分析中顯著之腸道微生物）以及干擾因子向量，可能不適用懲罰項之限制，但於本篇研究之做法僅有便秘症狀不受懲罰項之限制。以上的多重中介效應分析皆使用 R 軟體執行，其中 Ridge 回歸模型則使用套件 glmnet 4.1-2，其程式碼請參考附錄（五）。

模型解釋上，模型一中便秘症狀的估計係數（ $\hat{\alpha}_A$ ）一樣為便秘症狀對帕金森氏症的效應；模型四中便秘症狀的估計係數（ $\hat{\gamma}_A$ ）為直接效應（不透過腸道微生物變化的中介下便秘症狀對帕金森氏症的效應）。多重中介因子之效應一樣是透過差異法測量，其估計值也就是 $\hat{\alpha}_A - \hat{\gamma}_A$ ，中介比例則為 $\frac{\hat{\alpha}_A - \hat{\gamma}_A}{\hat{\alpha}_A}$ ，中介效應之標準誤亦是使用 Delta 方法估計，為

$$\sqrt{\sigma^2_{\hat{\alpha}_A} + \sigma^2_{\hat{\gamma}_A}}。$$

第三章、 結果

3.1 人口統計特徵分析結果

首先，我們比較數個有便秘症狀（N=31）及沒有便秘症狀（N=293）兩組之特徵，其結果統整於表（二）中。由該結果可看出此兩組於年齡、性別、種族、身高、體重、身體質量指數、地理區域（經度、緯度及居住地）、生活習慣（酒精攝取量及是否有抽菸習慣）、飲食習慣（咖啡攝取量、是否有每日攝取蔬果、是否有每日攝取穀物、是否有每日攝取肉類、是否有每日攝取堅果類及是否有每日攝取優格）與是否罹患除了帕金森氏症外的其他神經性疾病這些變數中無明顯差異，但有便秘症狀的組別罹患帕金森氏症的比例則是顯著高於沒有便秘症狀的組別（分別為 93.5% 及 57%， $p\text{-value} < 0.001$ ）。

再者，我們比較數個有罹患帕金森氏症（N=199）及沒有罹患帕金森氏症（N=131）兩組之特徵，其結果統整於表（三）中。由該結果可看出此兩組於年齡、性別、身高、身體質量指數、咖啡攝取量及是否有每日攝取蔬果這些變數中檢定結果之 $p\text{-value} < 0.05$ ，意即存在明顯差異。

表（二）：比較有無便秘症狀兩組受試者之人口統計特徵

	無便秘症狀 (N=293)	有便秘症狀 (N=31)	P-value
年齡			
平均數（標準差）	69.2 (8.77)	69.6 (9.86)	0.992
性別			
女性	133 (45.4%)	9 (29.0%)	0.120
男性	160 (54.6%)	22 (71.0%)	
種族			
白人	290 (99.0%)	30 (96.8%)	0.354
黑人或非裔美國人	2 (0.7%)	1 (3.2%)	
多於一個種族	1 (0.3%)	0 (0%)	
身高（英吋）			
平均數（標準差）	67.2 (3.95)	68.6 (3.95)	0.101
遺漏值	3 (1.0%)	0 (0%)	
體重（磅）			
平均數（標準差）	174 (40.10)	187 (35.70)	0.062
遺漏值	7 (2.4%)	2 (6.5%)	

身體質量指數

平均數 (標準差)	27.0 (5.52)	28.1 (4.92)	0.193
遺漏值	7 (2.4%)	2 (6.5%)	

經度

平均數 (標準差)	-20.6 (61.90)	-11.8 (62.40)	0.319
-----------	---------------	---------------	-------

緯度

平均數 (標準差)	-33.0 (81.10)	-43.1 (83.20)	0.751
-----------	---------------	---------------	-------

居住地

喬治亞亞特蘭大	42 (14.3%)	2 (6.5%)	0.459
紐約阿伯尼	119 (40.6%)	13 (41.9%)	
華盛頓西雅圖	132 (45.1%)	16 (51.6%)	

酒精攝取量

平均數 (標準差)	0.966 (0.83)	0.933 (0.79)	0.853
遺漏值	3 (1.0%)	1 (3.2%)	

是否有吸菸習慣

否	275 (93.9%)	28 (90.3%)	0.591
是	16 (5.5%)	3 (9.7%)	
遺漏值	2 (0.7%)	0 (0%)	

咖啡攝取量

平均數 (標準差)	1.22 (0.85)	1.30 (0.79)	0.670
遺漏值	4 (1.4%)	1 (3.2%)	

是否有每日攝取蔬果

否	48 (16.4%)	9 (29.0%)	0.141
是	241 (82.3%)	22 (71.0%)	
遺漏值	4 (1.4%)	0 (0%)	

是否有每日攝取穀物

否	90 (30.7%)	10 (32.3%)	0.998
是	196 (66.9%)	20 (64.5%)	
遺漏值	7 (2.4%)	1 (3.2%)	

是否有每日攝取肉類

否	117 (39.9%)	13 (41.9%)	1
是	171 (58.4%)	18 (58.1%)	
遺漏值	5 (1.7%)	0 (0%)	

是否有每日攝取堅果類

否	215 (73.4%)	25 (80.6%)	0.606
是	73 (24.9%)	6 (19.4%)	
遺漏值	5 (1.7%)	0 (0%)	

是否有每日攝取優格

否	259 (88.4%)	28 (90.3%)	1
是	24 (8.2%)	3 (9.7%)	
遺漏值	10 (3.4%)	0 (0%)	
是否罹患其他神經性疾病			
否	263 (89.8%)	28 (90.3%)	0.494
是	13 (4.4%)	0 (0%)	
遺漏值	17 (5.8%)	3 (9.7%)	
是否罹患帕金森氏症			
否	126 (43.0%)	2 (6.5%)	<0.001
是	167 (57.0%)	29 (93.5%)	

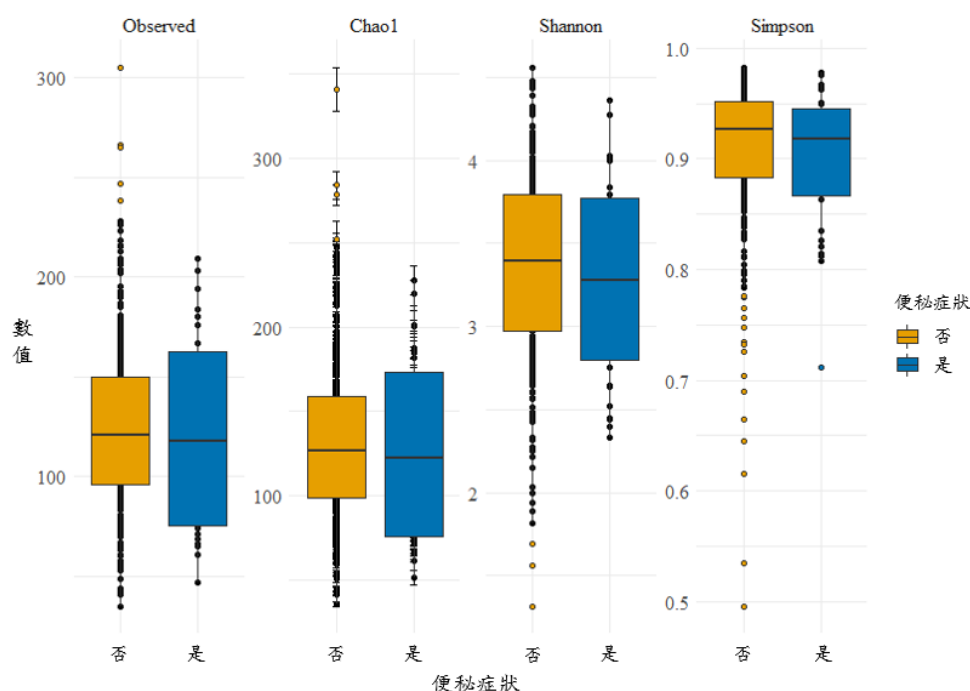
表（三）：比較有無罹患帕金森氏症兩組受試者之人口統計特徵

	未罹患帕金森氏症 (N=128)	罹患帕金森氏症 (N=196)	P-value
年齡			
平均數（標準差）	70.3 (8.52)	68.5 (9.03)	0.039
性別			
女性	77 (60.2%)	65 (33.2%)	<0.001
男性	51 (39.8%)	131 (66.8%)	
種族			
白人	128 (100%)	192 (98.0%)	0.266
黑人或非裔美國人	0 (0%)	1 (0.5%)	
多於一個種族	0 (0%)	3 (1.5%)	
身高（英吋）			
平均數（標準差）	66.4 (3.65)	68.0 (4.04)	<0.001
遺漏值	1 (0.8%)	2 (1.0%)	
體重（磅）			
平均數（標準差）	178 (40.70)	173 (39.30)	0.498
遺漏值	3 (2.3%)	6 (3.1%)	
身體質量指數			
平均數（標準差）	28.4 (5.72)	26.3 (5.14)	0.004
遺漏值	3 (2.3%)	6 (3.1%)	
經度			
平均數（標準差）	-20.8 (61.60)	-19.1 (62.30)	0.812
緯度			
平均數（標準差）	-31.7 (81.50)	-35.4 (81.20)	0.384
居住地			

喬治亞亞特蘭大	13 (10.2%)	31 (15.8%)	0.226
紐約阿伯尼	58 (45.3%)	74 (37.8%)	
華盛頓西雅圖	57 (44.5%)	91 (46.4%)	
酒精攝取量			
平均數 (標準差)	1.02 (0.77)	0.927 (0.86)	0.326
遺漏值	0 (0%)	4 (2.0%)	
是否有吸菸習慣			
否	123 (96.1%)	180 (91.8%)	0.321
是	5 (3.9%)	14 (7.1%)	
遺漏值	0 (0%)	2 (1.0%)	
咖啡攝取量			
平均數 (標準差)	1.35 (0.83)	1.14 (0.84)	0.022
遺漏值	0 (0%)	5 (2.6%)	
是否有每日攝取蔬果			
否	15 (11.7%)	42 (21.4%)	0.030
是	113 (88.3%)	150 (76.5%)	
遺漏值	0 (0%)	4 (2.0%)	
是否有每日攝取穀物			
否	41 (32.0%)	59 (30.1%)	0.877
是	85 (66.4%)	131 (66.8%)	
遺漏值	2 (1.6%)	6 (3.1%)	
是否有每日攝取肉類			
否	48 (37.5%)	82 (41.8%)	0.394
是	80 (62.5%)	109 (55.6%)	
遺漏值	0 (0%)	5 (2.6%)	
是否有每日攝取堅果類			
否	90 (70.3%)	150 (76.5%)	0.181
是	37 (28.9%)	42 (21.4%)	
遺漏值	1 (0.8%)	4 (2.0%)	
是否有每日攝取優格			
否	113 (88.3%)	174 (88.8%)	0.757
是	12 (9.4%)	15 (7.7%)	
遺漏值	3 (2.3%)	7 (3.6%)	
是否罹患其他神經性疾病			
否	120 (93.8%)	171 (87.2%)	0.245
是	8 (6.3%)	5 (2.6%)	
遺漏值	0 (0%)	20 (10.2%)	

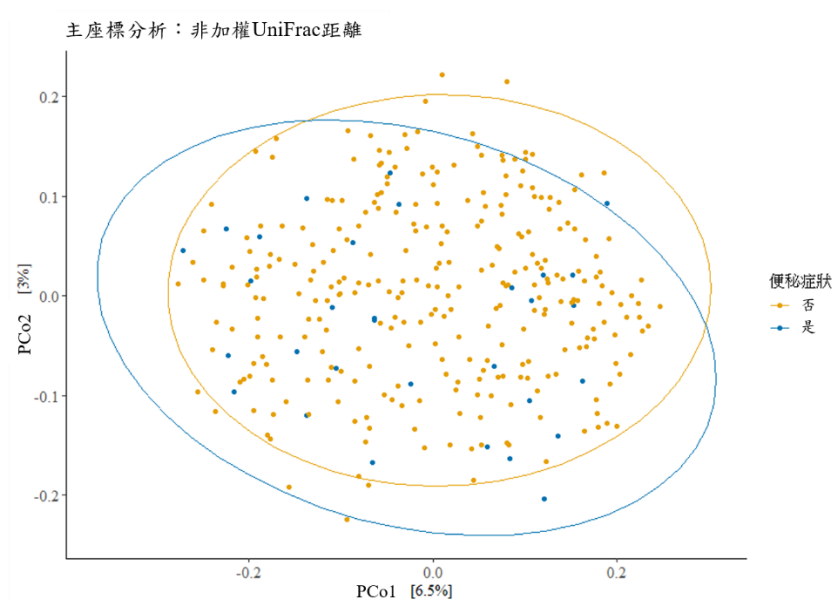
3.2 Alpha 多樣性及 Beta 多樣性分析結果

有便秘症狀及沒有便秘症狀兩組之腸道微生物群的 Alpha 多樣性及 Beta 多樣性的結果分別如下：首先是 Alpha 多樣性，如在上述方法中所提到的，此篇研究中採用四種結合物種豐富度及均勻度的 Alpha 多樣性指標（Observed、Chao1、Shannon 及 Simpson 指標），如圖（六）所示，而所有結果皆指出有無便秘症狀的兩組受試者之 Alpha 多樣性並無顯著差異，其 p-value 分別為 0.6258、0.5654、0.5335 及 0.7536，由於我們的受試者皆為人體，基本上人體內的腸道微生物之豐富度及均勻度並不會有明顯差異，因此 Alpha 多樣性無顯著差異之結論相當合理。我們接著應用基於系統發育（UniFrac）及非系統發育（坎貝拉）之距離計算方法分析有無便秘兩組樣本的微生物組成之相似程度，不同於 Alpha 多樣性之結果，此處採用的三個距離測量值（非加權、加權 UniFrac 及坎貝拉距離）皆顯示有無便秘症狀的兩組受試者之微生物組成有顯著的差異，其 p-value 分別為 0.0001、0.0003 及 0.0001。將其結果以主座標分析之散佈圖視覺化呈現，可發現：當我們基於非加權 UniFrac 距離進行分析，排名前二高的維度，也就是 PCo1 及 PCo2，分別解釋總變異的 6.5%及 3%，如圖（七）所示；當我們基於加權 UniFrac 距離進行分析，PCo1 及 PCo2 分別解釋總變異的 12.9%及 7.8%，如圖（八）所示；而當我們基於坎貝拉距離進行分析，PCo1 及 PCo2 則分別解釋總變異的 3%及 1.8%，如圖（九）所示。

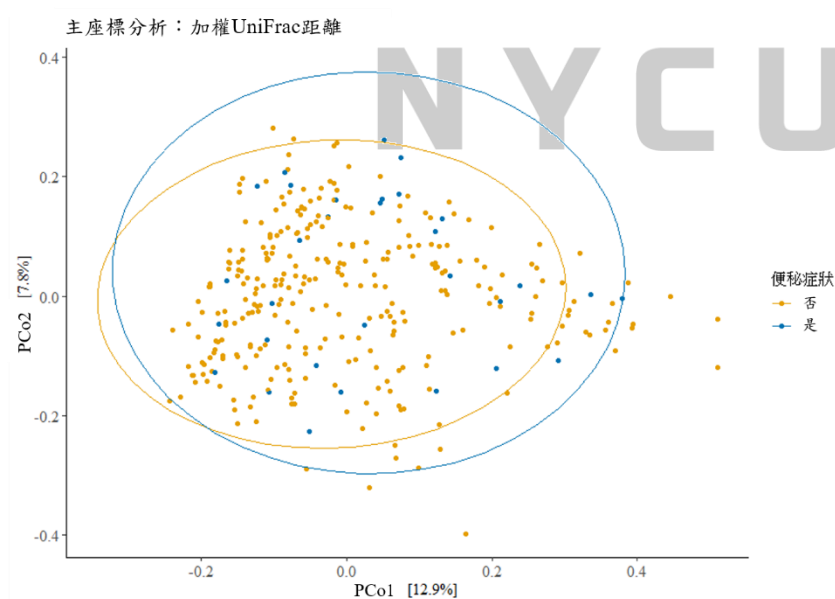


圖（六）：Alpha 多樣性各項指標之結果比較箱型圖：此箱型圖呈現 Observed、

Chao1、Shannon 及 Simpson 四項 Alpha 多樣性指標之平均數、中位數及上下四分位數。

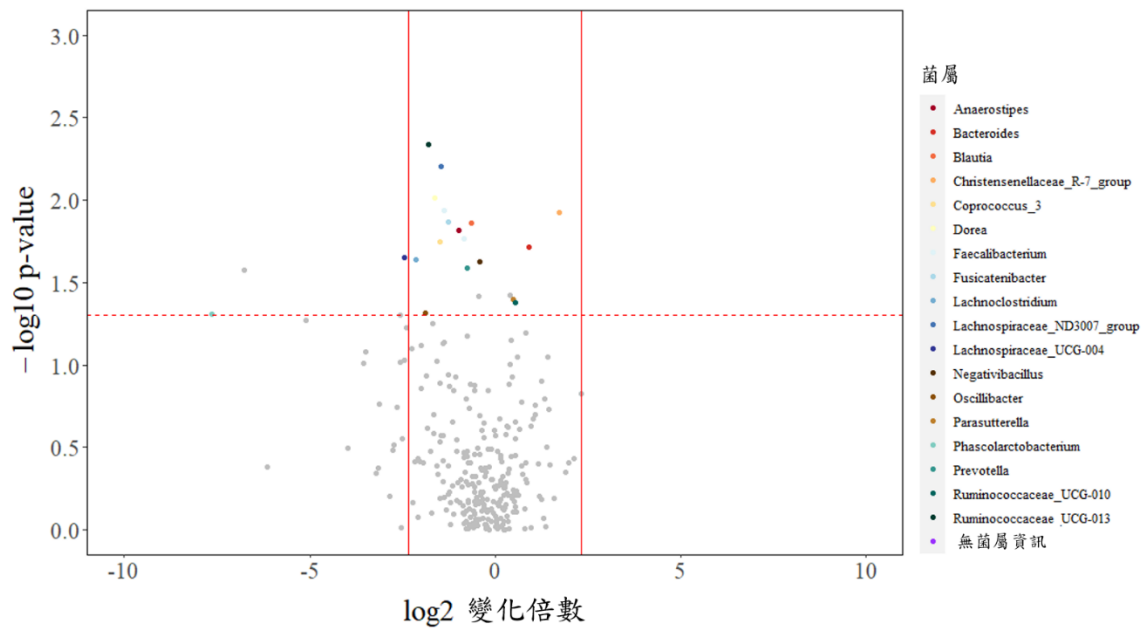


圖（七）：主座標分析散佈圖—基於非加權 UniFrac 距離計算

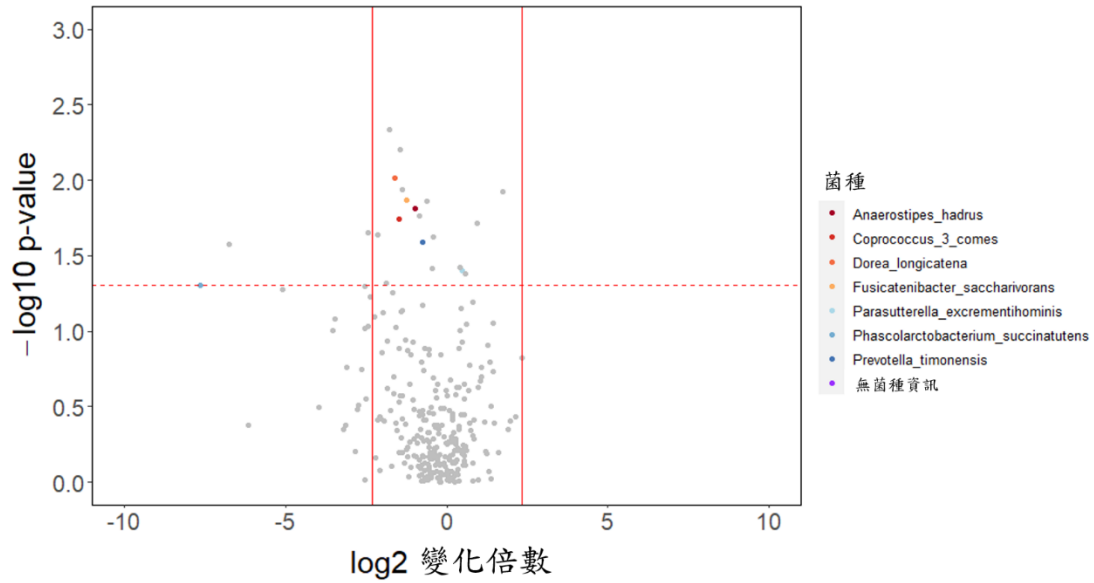


圖（八）：主座標分析散佈圖—基於加權 UniFrac 距離計算

Fusicatenibacter_saccharivorans、*Phascolarctobacterium_succinatutens* 及 *Prevotella_timo-*
nensis，如圖（十一）所示。



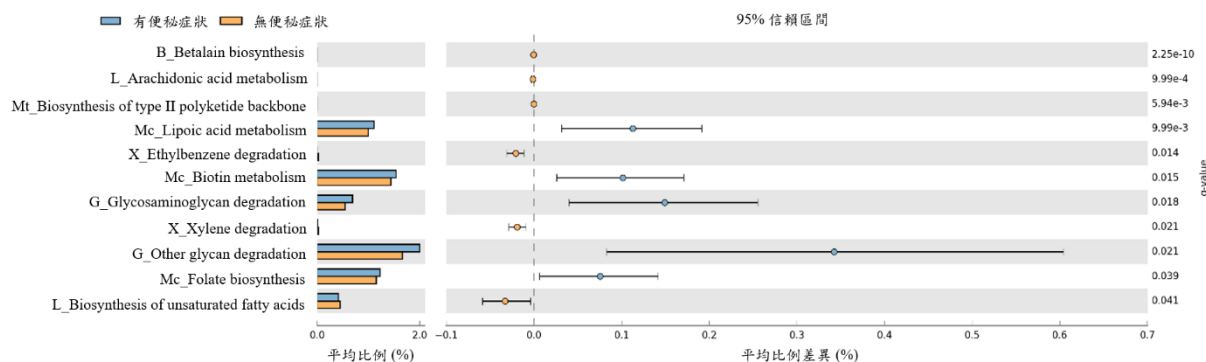
圖（十）：有無便秘症狀兩組受試者之相對豐度差異（菌屬層級）：任一點代表一個擴增子序列變異，其中 x 軸為 $\log_2$ (擴增子序列變異之變化倍數)之值；y 軸為檢定結果之 p-value。紅色虛線為 $p\text{-value}=0.05$ 之界線，若其 $p\text{-value}>0.05$ 則不上色（以灰色呈現），反之若該擴增子序列變異在有無便秘症狀兩組受試者有顯著差異則根據其菌屬上色。



圖（十一）：有無便秘症狀兩組受試者之相對豐度差異（菌種層級）：任一點代表一個擴增子序列變異，其中 x 軸為 $\log_2$ (擴增子序列變異之變化倍數) 之值；y 軸為檢定結果之 p-value。紅色虛線為 $p\text{-value}=0.05$ 之界線，若其 $p\text{-value}>0.05$ 則不上色（以灰色呈現），反之若該擴增子序列變異在有無便秘症狀兩組受試者有顯著差異則根據其菌種上色。

3.4 功能路徑豐度差異分析結果

為了評估有無便秘症狀兩組受試者之微生物組的功能差異，如方法中所提及到的，我們使用了一種計算工具，也就是 PICRUSt2，其利用 16S rRNA 擴增子數據，並基於 KEGG 數據庫下預測腸道微生物的功能路徑。同時也使用 STAMP 檢定並視覺化兩組之間功能路徑之豐度差異。在 119 條代謝途徑中，有無便秘症狀兩組受試者之間有 11 條存在顯著差異，如圖（十二）所示。根據 KEGG 分級 2 級分類，11 條途徑涉及脂肪酸代謝（Lipid metabolism）、萜類和聚酮化合物代謝（Metabolism of terpenoids and polyketides）、輔因子及維生素代謝（Metabolism of cofactors and vitamins）、外源生物降解和代謝（Xenobiotics biodegradation and metabolism）、聚醣生物合成和代謝（Glycan biosynthesis and metabolism）及其他次生代謝物的生物合成（Biosynthesis of other secondary metabolites）。



圖（十二）：有無便秘症狀兩組受試者之微生物組的功能差異：兩組受試者之間有 11 條顯著存在差異。若該路徑於有便秘症狀組之平均豐度高於無便秘症狀組，其平均比例差異則為正（95%信賴區間以藍色圓圈表示）；反之則為負（95%信賴區間以橘色圓圈表示）。（B = 其他次生代謝物的生物合成、G = 聚醣生物合成和代謝、L = 脂肪酸代謝、Mc = 輔因子及維生素代謝、Mt = 萜類和聚酮化合物代謝、X = 外源生物降解和代謝）

3.5 中介效應分析結果

於相對豐富度差異分析中，我們歸納出有無便秘症狀兩組受試者中，有顯著差異的菌屬、菌種及功能路徑分別為 18 個、7 個及 11 條，但該結果並無法證明這些菌屬與菌種為影響便秘症狀造成帕金森氏症的重要中介因子。因此為了了解便秘症狀引發帕金森氏症之機制中的關鍵中介因子，我們進行了中介效應分析，按照上述方法建構了模型一、模型二及模型三，並使用模型中的係數估計腸道微生物變化在該機制的中介效應，結果如表 2 所示。將腸道微生物分別以三個測量單位進行分析，首先，就菌屬層級而言，便秘症狀引發帕金森氏症之機制被腸道微生物變化中介之整體比例為 76.56%（p-value 為 0.0177），其中九個菌屬有顯著的中介效應，分別為 *Anaerostipes*（p-value 為 0.0043）、*Fusicatenibacter*（p-value 為 0.0048）、*Lachnospiraceae_ND3007_group*（p-value 為 0.0076）、*Blautia*（p-value 為 0.0077）、*Ruminococcaceae_UCG_013*（p-value 為 0.0097）、*Coprococcus_3*（p-value 為 0.0155）、*Faecalibacterium*（p-value 為 0.0293）、*Dorea*（p-value 為 0.0322）及 *Lachnospiraceae_UCG_004*（p-value 為 0.033），如表（四）所示。再者，就菌種層級而言，該機制被腸道微生物變化中介之整體比例為 69.8%（p-value 為 0.0309），其中四個菌種有顯著的中介效應，分別為 *Anaerostipes_hadrus*（p-value 為 0.0043）、*Fusicatenibacter_saccharivorans*（p-value 為 0.0048）、*Coprococcus_3_comes*（p-value 為

0.0155)、*Dorea_longicatena* (p-value 為 0.0482)，如表（五）所示。最後，就功能路徑而言，該機制被腸道微生物變化中介之整體比例為 66.27% (p-value 為 0.0445)，而唯一一個有顯著中介效應之路徑為硫辛酸代謝 (p-value 為 0.0276)，如表（六）所示。值得注意的是，上述提到之有顯著中介效應的菌屬、菌種及功能路徑全部都包含於相對豐度差異分析及功能路徑豐度差異分析中顯著的結果。

表（四）：中介效應分析結果（腸道微生物之測量單位：菌屬層級）：未校正之估計值為便秘症狀對帕金森氏症造成的總效應；校正下的估計值為中介效應，其中（A）為多重中介因子之分析結果，（B）則為單一中介因子之分析結果（p-value 由聯合顯著檢定計算而得）。

	估計值 (標準誤)	中介比例 (%)	P-value
未校正	2.3597 (0.7555)		
校正：			
(A)			
所有菌屬	0.5530 (0.0977)	76.56 %	
(B)			
Anaerostipes	2.1552 (0.7549)	8.66 %	0.0043
Fusicatenibacter	2.1843 (0.7596)	7.43 %	0.0048
Lachnospiraceae_ND3007_group	2.1513 (0.7594)	8.83 %	0.0076
Blautia	2.3521 (0.7710)	0.32 %	0.0077
Ruminococcaceae_UCG_013	2.1835 (0.7576)	7.47 %	0.0097
Coprococcus_3	2.2151 (0.7569)	6.13 %	0.0155
Faecalibacterium	2.2783 (0.7581)	3.45 %	0.0293
Dorea	2.2375 (0.7545)	5.17 %	0.0322
Lachnospiraceae_UCG_004	2.2338 (0.7590)	5.33 %	0.0330

表（五）：中介效應分析結果（腸道微生物之測量單位：菌種層級）：未校正之估計值為便秘症狀對帕金森氏症造成的總效應；校正下的估計值為中介效應，其中（A）為多重中介因子之分析結果，（B）則為單一中介因子之分析結果（p-value 由聯合顯著檢定計算而得）。

	估計值 (標準誤)	中介比例 (%)	P-value
未校正	2.3597 (0.7555)		
校正：			
(A)			
所有菌種	0.7127 (0.1064)	69.8 %	
(B)			
Anaerostipes_hadrus	2.1552 (0.7549)	8.66 %	0.0043
Fusicatenibacter_saccharivorans	2.1843 (0.7596)	7.43 %	0.0048
Coprococcus_3_comes	2.2151(0.7569)	6.13 %	0.0155
Dorea_longicatena	2.2280(0.7566)	5.58 %	0.0482

表（六）：中介效應分析（腸道微生物之測量單位：功能路徑）：校正之估計值為便秘症狀對帕金森氏症造成的總效應；校正下的估計值為中介效應，其中（A）為多重中介因子之分析結果，（B）則為單一中介因子之分析結果（p-value 由聯合顯著檢定計算而得）。

	估計值 (標準誤)	中介比例 (%)	P-value
未校正	2.3597 (0.7555)		
校正：			
(A)			
所有功能路徑	0.7960(0.1865)	66.27 %	
(B)			
I_Lipoic_acid_metabolism	2.2630(0.7621)	4.09 %	0.0276

第四章、 討論

4.1 相關文獻回顧

自 2003 年 Braak 等人首次提出腸腦軸假說以來(Braak, Del Tredici et al. 2003, Braak, Rüb et al. 2003)，腸胃道系統已經被推測為帕金森氏症的起源。透過一連串針對帕金森氏症逝世病患的神經病理學研究，Braak 等人提出以下假說：毒素及病原體會經由腸胃道進入人體，進而在腸神經系統中引發發炎及 α -突觸核蛋白 (Alpha-synuclein protein) 聚集。 α -突觸核蛋白是一種突觸前神經元蛋白 (Presynaptic neuronal protein)，可能會以多種方式引發帕金森氏症之發病，因此在遺傳學和神經病理學的角度而言， α -突觸核蛋白與帕金森氏症息息相關(Stefanis 2012)。而在腸神經系統中聚集的 α -突觸核蛋白又會在經由迷走神經 (Vagus nerve) 向上移動至中樞神經系統，導致黑質 (Substantia nigra) 中的多巴胺神經元 (Dopaminergic neurons) 退化(Hawkes, Del Tredici et al. 2007, Keshavarzian, Engen et al. 2020)，其中黑質是中腦的一個神經核團，而其內部的多巴胺神經元會通過黑質紋狀體通路 (Nigrostriatal pathway) 分泌多巴胺 (一種引起大腦活動的神經化學物質) 至大腦中的基底核 (Basal ganglia circuitry)，也就是說黑質是大腦中產生多巴胺的關鍵區域，因此黑質在調節運動神經動作及獎勵功能上扮演個關鍵的角色(Sonne, Reddy et al. 2019)，所以當黑質受損時，可能會導致神經系統疾病，例如：帕金森氏症。Braak 所提出的假說也解釋了為什麼帕金森氏症患者會出現胃腸道症狀，例如：便秘，大多發生在主要的運動症狀及中樞神經系統症狀出現前。儘管有證據可以說明便秘症狀引發帕金森氏症與微生物相關，卻尚無研究提供直接證據證明它們之間的因果關係。

4.2 本篇研究與過往研究之呼應

據我們所知，本文為第一篇利用中介效應分析探討便秘症狀透過腸道微生物變化之中介而引發帕金森氏症之機制的研究。此外，本篇的分析結果也發現有便秘症狀及無便秘症狀兩組之間有 18 個菌屬和 7 個菌種，以及 β 多樣性有顯著差異，表示便秘與腸道微生態失衡有相關，且與過往研究之結果有所呼應(Khalif, Quigley et al. 2005, Ohkusa, Koido et al. 2019, Zhang, Wang et al. 2021)。就腸道微生物之菌屬層級而言，便秘症狀引發帕金森氏症之效應被其中介的比例估計約有 76.56%，充分地提供證據鞏固便秘症狀對帕金森氏症的影響。而我們的結果亦顯示益生菌藥物之療效可能足以預防約 76.56%便秘症狀引發之帕金森氏症的發生。此外，在本篇研究中，與中介機制顯著相關之腸道微生物在

之前的研究中皆被發現與帕金森氏症有密切顯著相關(Hill-Burns, Debelius et al. 2017, Li, Wu et al. 2017, Petrov, Saltykova et al. 2017, Lin, Zheng et al. 2018, Aho, Pereira et al. 2019, Li, Cui et al. 2019, Romano, Savva et al. 2021)，他們同時也是與便秘症狀相關之腸道微生物菌屬（如圖（十）所示）的子集。

4.3 腸道微生物菌屬層級之結果探討

基於先前研究之關聯網路分析（Correlation network analysis）的結果，帕金森氏症相關的腸道微生物可依其菌屬可被分類為三個多微生物群集，分別為伺機性病原菌（Opportunistic pathogens）、短鏈脂肪酸（Short-chain fatty acid, SCFA）生成菌及醣類代謝益生菌（Carbohydrate-metabolizing probiotics）(Dalile, Van Oudenhove et al. 2019)。就如我們於表（四）所看到的，在便秘症狀引發帕金森氏症之機制具顯著中介效應的腸道微生物皆屬於短鏈脂肪酸生成菌(*Anaerostipes*、*Fusicatenibacter*、*Coprococcus*、*Dorea*、*Blautia*、*Faecalibacterium*、*Ruminococcaceae_UCG_013*、*Lachnospiraceae_ND3007_group*及 *Lachnospiraceae_UCG_004*)。短鏈脂肪酸是一組由五個或以下的碳原子組成的飽和脂肪酸，其中包括醋酸鹽（Acetate）、丙酸鹽（Propionate）及丁酸鹽（Butyrate），是胃腸道中微生物發酵作用（Fermentative activity）的主要產物。短鏈脂肪酸過少，尤其是丁酸鹽，會造成腸道通透性增加，進而導致腸漏（Leaky gut）(Dalile, Van Oudenhove et al. 2019)；短鏈脂肪酸也會透過促進短鏈脂肪酸反應基因（SCFA-responsive genes），如：組織蛋白去乙酰酶（Histone deacetylase, HDAC）的微膠細胞（Microglia）表現，調節微膠細胞之成熟性藉以增強血腦障壁（Blood brain barrier）之完整性(Dalile, Van Oudenhove et al. 2019)，因此可得知短鏈脂肪酸與大腦及其相關疾病息息相關。在許多研究中已經證實帕金森氏症患者的短鏈脂肪酸生成菌數過少，此情況亦有發生在其他發炎性疾病，如：炎症性腸病（Inflammatory bowel disease, IBD）、酒精相關病理及代謝症候群（Metabolic syndrome）(Koh, De Vadder et al. 2016)。而在我們的研究中也闡明了便秘症狀透過減少短鏈脂肪酸生成菌進而導致帕金森氏症，此機制也進一步增加腸道通透性，導致內毒素（Endotoxin）或外毒素（Exotoxin）的滲透，誘發帕金森氏症後續的病理變化。此外，阿克曼氏菌屬（*Akkermansia*）、乳酸桿菌屬（*Lactobacillus*）及雙歧桿菌屬（*Bifidobacterium*）皆為抗炎及醣代謝相關之益生菌，並有研究發現在帕金森氏症患者體內，此三種菌屬之豐度會增加(Hill-Burns, Debelius et al. 2017, Petrov, Saltykova et al. 2017, Aho, Pereira et al. 2019, Barichella, Severgnini et al. 2019, Romano, Savva et al. 2021)。

然而，我們並沒有發現這些菌屬有顯著的中介效應，這表示帕金森氏症患者體內增加的醣代謝益生菌在帕金森氏症的發病機制中並非扮演中介因子的角色，較有可能是為了克服腸道微生態失衡而進行的補償性變化。

4.4 功能路徑之結果探討

α -硫辛酸 (Alpha-lipoic acid, ALA) 是單體細胞進行中間代謝 (Intermediary metabolism) 所需的酵素輔因子 (Enzyme cofactors)，可用於分解碳水化合物並產生能量，不僅能在體內自然產生，也可透過蔬菜、水果、肉類、心臟、腎臟和肝臟等食物從外獲得，然而其最重要的特性莫過於抗氧化性 (Spalding and Prigge 2010)，並已經有相關研究證實 α -硫辛酸具有神經保護功能，可防止多巴胺能神經元的損失，減少黑質區域中 α -突觸核蛋白的聚集，進而改善帕金森氏症患者的運動功能障礙 (Li, He et al. 2015)。另外也有一個動物研究顯示， α -硫辛酸可以減少細胞內的活性氧類 (Reactive oxygen species, ROS)，促進多巴胺能神經元之存活，並改善帕金森氏症動物模型的運動缺陷 (Tai, Zheng et al. 2020)。在我們的研究中有發現硫辛酸代謝具有顯著的中介效應，闡明硫辛酸代謝可能在帕金森氏症發病機制中扮演重要的中介因子之角色。

4.5 研究限制

本篇研究中，有三項研究限制值得注意。首先，唯有在所有潛在干擾因子皆被調整的情況下，因果效應才可以不偏地透過回歸係數估計。然而，有些潛在干擾因子，舉例來說，規律的體能活動之頻率 (Mika, Van Treuren et al. 2015, Monda, Villano et al. 2017)，許多研究顯示增加體能活動的頻率將會增加有益微生物種的豐度，影響腸道微生物的組成，進而降低胃腸道疾病的風險，亦能有助於代謝，緩解便秘等症狀；此外，藥用植物（例如：刺毛鰲豆 (*Mucuna pruriens*)）之攝取也是潛在干擾因子，其可能具有緩解便秘的作用，並能更進一步改善腸道菌群失衡，但是原始資料並未提供諸如此類的潛在干擾因子。其次，便秘症狀、腸道微生物群及帕金森氏症可能會相互影響，在本篇研究中，我們是基於便秘症狀為帕金森氏症的病因之假設，但帕金森氏症亦有可能反向地加速便秘症狀的嚴重程度，而這種雙向的因果關係亦存在於帕金森氏症與腸道微生物群之間，由於聚集於中樞神經系統之 α -突觸核蛋白也有可能向下移動至腸胃道系統，並進一步導致腸道微生物變化。同樣地，腸道微生物之變化也有可能導致便秘的發生。然而，上述雙向之因果關係在文獻中被稱為「時變 (Time-varying)」之議題 (Lin, Young et al. 2017,

VanderWeele and Tchetgen Tchetgen 2017, Hernán MA 2020), 這也是一項常見但具有挑戰的議題, 需要進行縱向追蹤研究重複測量腸道微生物及記錄便秘症狀, 這也代表需要建立更複雜的模型, 舉例來說, 在這類型的研究中, 可能會導致便秘症狀的帕金森氏症相關治療藥物, 如: 諾康停 (Entacapone) (Fu, Lee et al. 2022), 也應該被考慮作為中介因子或是時變干擾因子, 此外, 比較有便秘症狀與無便秘症狀兩組受試者之腸道微生物之變化及透過提供益生菌進行動物研究來驗證上述機制也將會是必需進行的分析。最後, 我們的研究結果證實腸道微生物變化解釋了便秘症狀引發帕金森氏症之機制的 76.56%, 但剩餘的 23.44% 仍不清楚, 而其中未知的機制可能涉及外毒素聚合等作用, 這一部分可能需要進一步的研究來證實。

4.6 未來展望

未來, 我們希望可以同樣透過中介效應分析, 去探討腸道微生物在帕金森氏症與其他因素之間所扮演的角色。舉例而言, 過去有實驗發現帕金森氏症與咖啡攝取量有關, 咖啡因為目前最廣泛攝取之精神活性物質 (Psychoactive substances), 亦可以發揮神經保護作用, 而攝取咖啡因會造成人體腸道微生物變化進而減少或延緩帕金森氏症之發展 (Ren and Chen 2020), 但是腸道微生物變化在上述關係中是否扮演重要的中介因子, 這部份需要再進一步的進行分析並探討; 此外, 也有另外的研究發現帕金森氏症患者的身體質量指數顯著低於正常人 (van der Marck, Dicke et al. 2012), 因此我們懷疑是否為帕金森氏症引發腸道微生物變化進而導致其患者身體質量指數過低, 不過同樣地, 腸道微生物變化在這之中所扮演的角色也需要進行中介效應分析才能得知。所以我們期望中介效應分析能夠闡明腸道微生物在許多帕金森氏症的相關機制中所扮演的角色, 並為往後帕金森氏症的臨床治療帶來更多貢獻。

第五章、 結論

本篇為第一篇藉由中介效應分析探討便秘症狀、腸道微生物變化及帕金森氏症之間的因果關係之研究。我們的研究結果證實腸道微生物之變化在便秘症狀相關之帕金森氏症的發病機致中扮演重要的角色，且主要是透過減少短鏈脂肪酸生成菌而導致。因此，未來可能可以透過藥物治療或是飲食調整來治療及預防便秘症狀相關之帕金森氏症，尤其是短鏈脂肪酸生成菌相關之益生菌。

陽明交大
NYCU

參考文獻

- Adams-Carr, K. L., et al. (2016). "Constipation preceding Parkinson's disease: a systematic review and meta-analysis." Journal of Neurology, Neurosurgery & Psychiatry **87**(7): 710-716.
- Aho, V. T., et al. (2019). "Gut microbiota in Parkinson's disease: temporal stability and relations to disease progression." EBioMedicine **44**: 691-707.
- Aitchison, J. (1982). "The statistical analysis of compositional data." Journal of the Royal Statistical Society: Series B (Methodological) **44**(2): 139-160.
- Anderson, M. J. (2005). "Permutational multivariate analysis of variance." Department of Statistics, University of Auckland, Auckland **26**: 32-46.
- Arnold, J. W., et al. (2016). "Emerging technologies for gut microbiome research." Trends in microbiology **24**(11): 887-901.
- Barichella, M., et al. (2019). "Unraveling gut microbiota in Parkinson's disease and atypical parkinsonism." Movement disorders **34**(3): 396-405.
- Baron, R. M. and D. A. Kenny (1986). "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." Journal of personality and social psychology **51**(6): 1173.
- Boca, S. M., et al. (2014). "Testing multiple biological mediators simultaneously." Bioinformatics **30**(2): 214-220.
- Bolger, A. M., et al. (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.
- Braak, H., et al. (2003). "Staging of brain pathology related to sporadic Parkinson's disease." Neurobiology of aging **24**(2): 197-211.
- Braak, H., et al. (2003). "Idiopathic Parkinson's disease: possible routes by which vulnerable

neuronal types may be subject to neuroinvasion by an unknown pathogen." Journal of neural transmission **110**(5): 517-536.

Callahan, B. J., et al. (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nature methods **13**(7): 581-583.

Carabotti, M., et al. (2015). "The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems." Annals of gastroenterology: quarterly publication of the Hellenic Society of Gastroenterology **28**(2): 203.

Chao, A. (1984). "Nonparametric estimation of the number of classes in a population." Scandinavian Journal of statistics: 265-270.

Dalile, B., et al. (2019). "The role of short-chain fatty acids in microbiota–gut–brain communication." Nature reviews Gastroenterology & hepatology **16**(8): 461-478.

De Lau, L. M. and M. M. Breteler (2006). "Epidemiology of Parkinson's disease." The Lancet Neurology **5**(6): 525-535.

Douglas, G. M., et al. (2020). "PICRUSt2 for prediction of metagenome functions." Nature Biotechnology **38**(6): 685-688.

Fu, S.-C., et al. (2022). "A Pilot Study Exploring the Association of Entacapone, Gut Microbiota, and the Subsequent Side Effects in Patients With Parkinson's Disease." Frontiers in Cellular and Infection Microbiology **12**.

Hamza, T. H., et al. (2010). "Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease." Nature genetics **42**(9): 781-785.

Hawkes, C. H., et al. (2007). "Parkinson's disease: a dual-hit hypothesis." Neuropathology and applied neurobiology **33**(6): 599-614.

Hernán MA, R. J. (2020). Causal Inference: What If. Boca Raton, Chapman & Hall/CRC.

Hill-Burns, E. M., et al. (2017). "Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome." Movement disorders **32**(5): 739-749.

Hoehn, M. M. and M. D. Yahr (1998). "Parkinsonism: onset, progression, and mortality." Neurology **50**(2): 318-318.

Huang, Y.-T. (2018). "Joint significance tests for mediation effects of socioeconomic adversity on adiposity via epigenetics." The Annals of Applied Statistics **12**(3): 1535-1557.

Huang, Y., et al. (2021). "The role of intestinal dysbiosis in Parkinson's disease." Frontiers in Cellular and Infection Microbiology **11**.

Imai, K., et al. (2010). "Identification, inference and sensitivity analysis for causal mediation effects." Statistical science **25**(1): 51-71.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic acids research **28**(1): 27-30.

Keshavarzian, A., et al. (2020). "The gut microbiome in Parkinson's disease: A culprit or a bystander?" Progress in Brain Research **252**: 357-450.

Khalif, I., et al. (2005). "Alterations in the colonic flora and intestinal permeability and evidence of immune activation in chronic constipation." Digestive and Liver Disease **37**(11): 838-849.

Koh, A., et al. (2016). "From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites." Cell **165**(6): 1332-1345.

Kucera, M. and B. A. Malmgren (1998). "Logratio transformation of compositional data: a resolution of the constant sum constraint." Marine Micropaleontology **34**(1-2): 117-120.

Lance, G. N. and W. T. Williams (1967). "Mixed-Data Classificatory Programs I -

Agglomerative Systems." Australian Computer Journal **1**(1): 15-20.

Lembo, A. and M. Camilleri (2003). "Chronic constipation." New England Journal of Medicine **349**(14): 1360-1368.

Li, C., et al. (2019). "Gut microbiota differs between Parkinson's disease patients and healthy controls in northeast China." Frontiers in molecular neuroscience: 171.

Li, W., et al. (2017). "Structural changes of gut microbiota in Parkinson's disease and its correlation with clinical features." Science China Life Sciences **60**(11): 1223-1233.

Li, Y.-H., et al. (2015). "Lipoic acid protects dopaminergic neurons in LPS-induced Parkinson's disease model." Metabolic brain disease **30**(5): 1217-1226.

Lin, A., et al. (2018). "Gut microbiota in patients with Parkinson's disease in southern China." Parkinsonism & related disorders **53**: 82-88.

Lin, C.-H., et al. (2014). "Risk of Parkinson's disease following severe constipation: a nationwide population-based cohort study." Parkinsonism & related disorders **20**(12): 1371-1375.

Lin, H. and S. D. Peddada (2020). "Analysis of microbial compositions: a review of normalization and differential abundance analysis." NPJ biofilms and microbiomes **6**(1): 1-13.

Lin, S.-H., et al. (2017). "Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders." Epidemiology (Cambridge, Mass.) **28**(2): 266.

Lin, X., et al. (2021). "Constipation induced gut microbiota dysbiosis exacerbates experimental autoimmune encephalomyelitis in C57BL/6 mice." Journal of Translational Medicine **19**(1): 1-16.

Lozupone, C., et al. (2011). "UniFrac: an effective distance metric for microbial community

comparison." The ISME journal **5**(2): 169-172.

Lozupone, C. A., et al. (2012). "Diversity, stability and resilience of the human gut microbiota." Nature **489**(7415): 220-230.

MacKinnon, D. P. (2012). Introduction to statistical mediation analysis, Routledge.

Magurran, A. E. (1988). Ecological diversity and its measurement, Princeton university press.

Mandal, S., et al. (2015). "Analysis of composition of microbiomes: a novel method for studying microbial composition." Microbial ecology in health and disease **26**(1): 27663.

McMurdie, P. J. and S. Holmes (2013). "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one **8**(4): e61217.

Mika, A., et al. (2015). "Exercise is more effective at altering gut microbial composition and producing stable changes in lean mass in juvenile versus adult male F344 rats." PloS one **10**(5): e0125889.

Minato, T., et al. (2017). "Progression of Parkinson's disease is associated with gut dysbiosis: two-year follow-up study." PloS one **12**(11): e0187307.

Monda, V., et al. (2017). "Exercise Modifies the Gut Microbiota with Positive Health Effects." Oxidative Medicine and Cellular Longevity **2017**: 3831972.

Mullis, K., et al. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press.

Ohkusa, T., et al. (2019). "Gut microbiota and chronic constipation: a review and update." Frontiers in medicine **6**: 19.

Palmer, M. W. (2004). "Ordination Methods-an overview." Botany Department, Oklahoma

State University, Stillwater, Oklahoma **74078**.

Petrov, V., et al. (2017). "Analysis of gut microbiota in patients with Parkinson's disease." Bulletin of experimental biology and medicine **162**(6): 734-737.

Picillo, M., et al. (2021). "The PRIAMO study: age-and sex-related relationship between prodromal constipation and disease phenotype in early Parkinson's disease." Journal of neurology **268**(2): 448-454.

Rai, S. N., et al. (2020). "Mucuna pruriens in Parkinson's and in some other diseases: recent advancement and future prospective." 3 Biotech **10**(12): 1-11.

Rai, S. N. and P. Singh (2020). "Advancement in the modelling and therapeutics of Parkinson's disease." Journal of Chemical Neuroanatomy **104**: 101752.

Rai, S. N., et al. (2021). "Promising drug targets and associated therapeutic interventions in Parkinson's disease." Neural Regeneration Research **16**(9): 1730.

Ren, X. and J.-F. Chen (2020). "Caffeine and Parkinson's disease: multiple benefits and emerging mechanisms." Frontiers in Neuroscience: 1334.

Romano, S., et al. (2021). "Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation." npj Parkinson's Disease **7**(1): 1-13.

Rosen, M. J., et al. (2012). "Denoising PCR-amplified metagenome data." BMC bioinformatics **13**(1): 1-16.

Rosenzweig, M. L. (1995). Species diversity in space and time, Cambridge University Press.

Saleh, A. M. E., et al. (2019). Theory of ridge regression estimation with applications, John Wiley & Sons.

Savica, R., et al. (2009). "Medical records documentation of constipation preceding Parkinson

disease: a case-control study." Neurology **73**(21): 1752-1758.

Sonne, J., et al. (2019). "Neuroanatomy, substantia nigra."

Spalding, M. D. and S. T. Prigge (2010). "Lipoic acid metabolism in microbial pathogens." Microbiology and Molecular Biology Reviews **74**(2): 200-228.

Stefanis, L. (2012). " α -Synuclein in Parkinson's disease." Cold Spring Harbor perspectives in medicine **2**(2): a009399.

Stirpe, P., et al. (2016). "Constipation: an emerging risk factor for Parkinson's disease?" European Journal of Neurology **23**(11): 1606-1613.

Tai, S., et al. (2020). "Alpha-Lipoic Acid Mediates Clearance of Iron Accumulation by Regulating Iron Metabolism in a Parkinson's Disease Model Induced by 6-OHDA." Frontiers in Neuroscience: 612.

van der Marck, M. A., et al. (2012). "Body mass index in Parkinson's disease: a meta-analysis." Parkinsonism & related disorders **18**(3): 263-267.

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

VanderWeele, T. J. and E. J. Tchetgen Tchetgen (2017). "Mediation analysis with time varying exposures and mediators." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **79**(3): 917-938.

VanderWeele, T. J. and S. Vansteelandt (2009). "Conceptual issues concerning mediation, interventions and composition." Statistics and its Interface **2**(4): 457-468.

Weiss, S., et al. (2017). "Normalization and microbial differential abundance strategies depend upon data characteristics." Microbiome **5**(1): 1-18.

White, J. R., et al. (2009). "Statistical methods for detecting differentially abundant features in clinical metagenomic samples." PLoS computational biology **5**(4): e1000352.

Whittaker, R. H. (1972). "Evolution and measurement of species diversity." Taxon **21**(2-3): 213-251.

Xia, F., et al. (2013). "A logistic normal multinomial regression model for microbiome compositional data analysis." Biometrics **69**(4): 1053-1063.

Xu, Z. and R. Knight (2015). "Dietary effects on human gut microbiome diversity." British Journal of Nutrition **113**(S1): S1-S5.

Yatsunenko, T., et al. (2012). "Human gut microbiome viewed across age and geography." Nature **486**(7402): 222-227.

Zhang, S., et al. (2021). "Role of gut microbiota in functional constipation." Gastroenterology Report **9**(5): 392-401.

黃靖鈞, et al. (2018). "帕金森氏症之中醫辨證論治的回顧性文獻探討." 中醫藥研究論叢 **21**(1): 71-83.

附錄（一）：資料集

本篇論文所採用的資料，經過第 2.3.1 小節中所承述之 16S rRNA 資料處理流程後，用以做後續分析之資料如下，歡迎下載取用：

<https://drive.google.com/file/d/185K1Rm1JfsiaP8kTC3WTDL7awgU5uTsR/view?usp=sharing>

陽明交大
NYCU

附錄（二）：程式碼—Alpha 多樣性

本篇研究採用擴增子序列變異估計四種常見的 Alpha 多樣性指標，分別為 Observed、Chao1、Shannon 及 Simpson。其計算方式、檢定方法及箱型圖繪製之程式碼如下：

```
library(phyloseq)
library(SpadeR)
library(ape)
library(vegan)
library(magrittr)
library(ggplot2)
library(dplyr)
library(omics)
library(xlsx)
library(sqldf)
library(magrittr)

load("Hill330.rda")
df = data.frame(ps_rar@sam_data)
index=which(df$ss_constipation%in% c("Yes", "No"))

psLE1 = phyloseq(otu_table(ps_rar@otu_table[index,], taxa_are_rows=FALSE),
                 sample_data(df[index,]),
                 tax_table(ps_rar@tax_table))

#data frame
R = estimate_richness(psLE1, measures=c("Observed", "Chao1", "ACE", "Shannon", "Simpson", "InvSimpson", "Fisher"))
df = data.frame(psLE1@sam_data)
table(df$ss_constipation)
df$ss_constipation <- ifelse(df$ss_constipation=="Yes", "有", "無")
R_sub1 = cbind(R, df$ss_constipation) %>% .[index,]

#calculate alpha diversity
alpha_diversity = data.frame()
for(i in 1:9){
  alpha_diversity[1,i] = summary(aov( R_sub1[,i] ~
R_sub1[,10]))[[1]]$`Pr(>F)`[1]
}
colnames(alpha_diversity) = colnames(R)
view(alpha_diversity)

#plot alpha diversity
library(extrafont)
```

```

loadfonts(device="win")
windowsFonts(A = windowsFont("Times New Roman"))

plot_richness(psLE1, x="ss_constipation", measures=c("Observed", "Chao1",
"Shannon", "Simpson")) +
  geom_boxplot(aes(fill=ss_constipation), outlier.shape = 21) +
  theme_minimal()+
  theme(axis.text=element_text(family = "A", size=12),
        axis.title=element_text(family = "A", size=12),
        legend.text = element_text(family = "A", size=12),
        text = element_text(family = "A", size = 12),
        strip.text = element_text(family = "A", size = 12)
  )+
  xlab("便秘症狀") +
  ylab("數值") +
  scale_colour_manual(values=c("#E69F00", "#0072B2"))+
  scale_fill_manual(values=c("#E69F00", "#0072B2"))+
  labs(fill="便秘症狀")

#ANOVA
one_ANOVA = function(){
  m = c();
  for (i in c(1:6)){
    t0 = summary(aov(R_sub[,i] ~ R_sub$case_control))[[1]]$`Pr(>F)`[1]
    t = c(t0);
    m = cbind(m,t);
  }
  colnames(m) = c("Observed" , "Chao1" , "se.chao1" #, "ACE", "se.ACE"
, "Shannon", "Simpson","ichao1_list")#, "InvSimpson",
"Fisher")
  rownames(m) = c("PD_Control")
  return(m);
}

```

附錄（三）：程式碼—Beta 多樣性

本篇研究採用三種距離矩陣計算方法，分別為加權 UniFrac 距離、非加權 UniFrac 距離及坎貝拉距離。其計算方式、檢定方法及主座標分析圖繪製之程式碼如下：

```
library(phyloseq)
library(SpadeR)
library(ape)
library(vegan)
library(magrittr)
library(ggplot2)
library(dplyr)
library(omics)
library(xlsx)
library(sqldf)
library(magrittr)

load("Hill330.rda")
df = data.frame(ps_rar@sam_data)
index=which(df$ss_constipation%in% c("Yes", "No"))

psLE1 = phyloseq(otu_table(ps_rar@otu_table[index,], taxa_are_rows=FALSE),
                 sample_data(df[index,]),
                 tax_table(ps_rar@tax_table))
psLEr <- transform_sample_counts(psLE1, function(otu) otu/sum(otu))

#calculate beta
dft = data.frame(psLEr@sam_data)
dft$ss_constipation <- ifelse(dft$ss_constipation=="Yes","有","無")
set.seed(123)
random_tree = rtree(ntaxa(psLEr), rooted=TRUE, tip.label=taxa_names(psLEr))
psLEr1 = merge_phyloseq(psLEr, random_tree)

beta = function(psLEr){
  unifrac.dist=UniFrac(psLEr, weighted=TRUE, normalized=TRUE, parallel=FALSE, fast=TRUE)
  ununifrac.dist=UniFrac(psLEr, weighted=FALSE, normalized=TRUE, parallel=FALSE, fast=TRUE)
  canberra.dist = vegdist(psLEr@otu_table, method="canberra")

  dist_method = c("unifrac.dist", "ununifrac.dist", "canberra.dist")
  testm = NULL
```

```

for(i in 1:length(dist_method)){
  set.seed(123)
  test = adonis(as.dist(as.matrix(get(dist_method[i]))))~dft$case_control,data=dft,permutations = 9999)
  testc = c(round(test$aov.tab$`Pr(>F)`[1],4),
round(test$aov.tab$R2[1],4), round(length(test$f.perms),0))
  testm = cbind(testm, testc)
}

rownames(testm) = c("p-value", "R2","permutation")
colnames(testm) = c("Weighted Unifrac","Unweighted Unifrac", "Canberra")
return(testm)
}

set.seed(123)
adonis(as.dist(as.matrix(get("unifrac.dist"))))~dft$case_control,data=dft,permutations = 9999)$aov.tab

set.seed(123)
adonis(as.dist(as.matrix(ununifrac.dist))~dft$entacapone,data=dft,permutations = 9999)$aov.tab

set.seed(123)
adonis(as.dist(as.matrix(canberra.dist))~dft$entacapone,data=dft,permutations = 9999)$aov.tab

#plot beta diversity: PCoA
pcoa.ununifrac = ordinate(psLER1, method="PCoA", distance="unifrac",
weighted=FALSE)
pcoa.canberra = ordinate(psLER1, method="PCoA", distance="canberra")
pcoa.unifrac = ordinate(psLER1, method="PCoA", distance="unifrac",
weighted=TRUE)
p1 = plot_ordination(psLER1 , pcoa.unifrac ,color="ss_constipation", title="weighted Unifrac PCoA") +
  geom_point(size = 1) +
  scale_colour_manual(values=c("#E69F00", "#0072B2")) +
  theme_classic() +
  theme(axis.text=element_text(family = "A", size=12),
        axis.title=element_text(family = "A", size=12),
        legend.text = element_text(family = "A", size=12),
        text = element_text(family = "A", size = 12),
        strip.text = element_text(family = "A", size = 12)
  )+
  labs(colour = '便秘症狀',title = "UniFrac 主座標分析:加權 UniFrac 距離")+
  stat_ellipse(type = "t")

```

p1

```
p2 = plot_ordination(psLEr1, pcoa.ununifrac ,color="ss_constipation", title="unweighted Unifrac PCoA") +  
  geom_point(size = 1) +  
  scale_colour_manual(values=c("#E69F00", "#0072B2")) +  
  theme_classic() +  
  theme(axis.text=element_text(family = "A", size=12),  
        axis.title=element_text(family = "A", size=12),  
        legend.text = element_text(family = "A", size=12),  
        text = element_text(family = "A", size = 12),  
        strip.text = element_text(family = "A", size = 12))  
)+  
labs(colour = '便秘症狀',title = "主座標分析:非加權 UniFrac 距離")+  
stat_ellipse(type = "t")
```

p2

```
p3 = plot_ordination(psLEr1 , pcoa.canberra ,color="ss_constipation", title="Canberra PCoA") +  
  geom_point(size = 1) +  
  scale_colour_manual(values=c("#E69F00", "#0072B2")) +  
  theme_classic() +  
  theme(axis.text=element_text(family = "A", size=12),  
        axis.title=element_text(family = "A", size=12),  
        legend.text = element_text(family = "A", size=12),  
        text = element_text(family = "A", size = 12),  
        strip.text = element_text(family = "A", size = 12))  
)+  
labs(colour = '便秘症狀',title = "主座標分析:坎貝拉距離")+  
stat_ellipse(type = "t")
```

p3

附錄（四）：程式碼－火山圖

相對豐度差異分析之結果之火山圖的程式碼如下所示，以菌屬為例：

```
library(phyloseq)
library(ape)
library(vegan)
library(magrittr)
library(ggplot2)
library(dplyr)
library(omics)
library(dplyr)
library(stringr)
library(RColorBrewer)
library(pals)
load("Hill330.rda")

ps<-ps_rar

new_otu1 = t(apply(ps@otu_table,1,function(x) x/sum(x)))
ps1 <- phyloseq(otu_table(new_otu1, taxa_are_rows=FALSE),
               sample_data(ps@sam_data),
               tax_table(ps@tax_table))
asv<-data.frame(ps1@otu_table)
few <- c()
for(i in 1:ncol(asv)){
  k=length(which(asv[,i]>0))
  if(k<33){
    few <- c(few,i)
  }
}
asv_new <- asv[,-few]

taxa<-data.frame(ps1@tax_table)
Genus<-taxa[-few,6]
Species<-paste0(taxa[-few,6],"_",taxa[-few,7])
metadata<-data.frame(ps1@sam_data)
Group<-metadata$ss_constipation
sample<-rownames(metadata)
asv_new<-asv_new[rownames(asv_new)%in%sample,]
new_data<-cbind(Group,asv_new)
new_data <- new_data[new_data$Group!="NA",]
new_data$Group <- as.factor(new_data$Group)

k<-aggregate(new_data[2:ncol(new_data)], list(new_data$Group), mean)
```

```

all_zero<-c()
for (i in 1:dim(k)[2]){
  if (k[1,i]==0&& k[2,i]==0){
    all_zero<-c(all_zero,i)
  }
}
k<-k[,-c(1,all_zero)]
min<-min(asv_new[asv_new>0])/100
result<-data.frame()

for (i in 1:(dim(new_data)[2]-1)){
  if(k[2,i]==0|k[1,i]==0){
    FoldChange<-(k[2,i]+min)/(k[1,i]+min)
  }else{
    FoldChange<-(k[2,i])/(k[1,i])
  }

  result[i,1]<-FoldChange
}

for (i in 1:(dim(new_data)[2]-1)){
  criteria<-ifelse(k[1,i]==0|k[2,i]==0,"1","0")
  result[i,2]<-criteria
}

for (i in 1:(dim(new_data)[2]-1) ){
  formula<- as.formula(paste(colnames(new_data)[i+1],"~","Group"))
  #deal with ties
  p<-exactRankTests::wilcox.exact(formula,data=new_data)$p.value
  result[i,3]<-p
}

result$Bacterium<-colnames(new_data)[2:ncol(new_data)]
result$Genus<-Genus
result$Species<-Species
colnames(result)<-c("FoldChange","Criteria","pvalue","Bacterium","Ge-
nus","Species")

result$sig_g<-ifelse(result$pvalue<0.05,result$Genus,NA)
result$col<-ifelse(result$Genus=="Anaerostipes",1,ifelse(result$Ge-
nus=="Bacteroides",2,ifelse(result$Genus=="Blautia",3,ifelse(result$Ge-
nus=="Christensenellaceae_R-

```

```

7_group",4,ifelse(result$Genus=="Coprococcus_3",5,ifelse(result$Ge-
nus=="Dorea",6,ifelse(result$Genus=="Faecalibacterium",7,ifelse(result$Ge-
nus=="Fusicatenibacter",8,ifelse(result$Genus=="Lachnoclostrid-
ium",9,ifelse(result$Genus=="Lachnospiraceae_ND3007_group",10,ifelse(re-
sult$Genus=="Lachnospiraceae_UCG-004",11,ifelse(result$Genus=="Negativiba-
cillus",12,ifelse(result$Genus=="Oscillibacter",13,ifelse(result$Ge-
nus=="Parasutterella",14,ifelse(result$Genus=="Phascolarctobacte-
rium",15,ifelse(result$Genus=="Prevotella",16,ifelse(result$Genus=="Rumino-
coccaceae_UCG-010",17,ifelse(result$Genus=="Ruminococcaceae_UCG-
013",18,NA)))))))))
result[result$pvalue<0.05&is.na(result$Genus),c(7,9)] <- "無菌屬資訊"
result$col <- as.factor(result$col)
result$col<-factor(result$col,levels=c("1","2","3","4","5","6","7","8",
"9","10","11","12","13","14","15","16","17","18","無菌屬資訊"),
labels=c("Anaerostipes","Bacteroides","Blautia","Christensenellaceae_R-
7_group","Coprococcus_3","Dorea","Faecalibacterium","Fusicatenibac-
ter","Lachnoclostridium","Lachnospiraceae_ND3007_group","Lachnospi-
raceae_UCG-004","Negativibacillus","Oscillibacter","Parasutterella","Phas-
colarctobacterium","Prevotella","Ruminococcaceae_UCG-010","Ruminococca-
ceae_UCG-013","no information"))
mycol1<-brewer.pal(11,"RdYlBu")
mycol2<-brewer.pal(11,"BrBG")
mycol<-c(mycol1,mycol2[c(1:3,8:11)],"purple1")

result%>%
  mutate(color=ifelse(abs(log2(FoldChange))<=1,"gray","blue"))%>%
  ggplot(aes(x=log2(FoldChange), y=-log10(pvalue),color=sig_g),size=1) +
  theme(panel.background = element_blank(),panel.border=element_rect(colour
= "black",fill=NA))+
  geom_point() +
  scale_color_manual(values=c("Anaerostipes"=mycol[1],"Bacteroides"=my-
col[2],"Blautia"=mycol[3],"Christensenellaceae_R-7_group"=mycol[4],"Copro-
coccus_3"=mycol[5],"Dorea"=mycol[6],"Faecalibacterium"=mycol[7],"Fusicaten-
ibacter"=mycol[8],"Lachnoclostridium"=mycol[9],"Lachnospi-
raceae_ND3007_group"=mycol[10],"Lachnospiraceae_UCG-004"=mycol[11],"Nega-
tivibacillus"=mycol[12],"Oscillibacter"=mycol[13],"Parasutterella"=my-
col[14],"Phascolarctobacterium"=mycol[15],"Prevotella"=mycol[16],"Rumino-
coccaceae_UCG-010"=mycol[17],"Ruminococcaceae_UCG-013"=mycol[18],"no infor-
mation"=mycol[19]),na.value="gray",labels=c("Anaerostipes","Bac-
teroides","Blautia","Christensenellaceae_R-7_group","Coprococ-
cus_3","Dorea","Faecalibacterium","Fusicatenibacter","Lachnoclostrid-
ium","Lachnospiraceae_ND3007_group","Lachnospiraceae_UCG-004","Negativiba-
cillus","Oscillibacter","Parasutterella","Phascolarctobacte-
rium","Prevotella","Ruminococcaceae_UCG-010","Ruminococcaceae_UCG-013","無
菌屬資訊"))+

```

```

xlab("log2 變化倍數") +
ylab(expression(-log10*paste(" ", "p-value")))) +
labs(colour="菌屬")+
scale_y_continuous(breaks = seq(from = 0, to = 3, by = 0.5),limits =
c(0,3)) +
scale_x_continuous(breaks = seq(from = -10, to = 10, by = 5),limits = c(-
10,10)) +
geom_hline(yintercept = -log10(0.05),color='red',linetype = "dashed")+
geom_vline(xintercept = max(log2(result[result$Crite-
ria==0,1])),color='red')+
geom_vline(xintercept = -max(log2(result[result$Crite-
ria==0,1])),color='red')+
guides(color=guide_legend(nrow=19)) +
theme(legend.position = "right",
      axis.title = element_text(family = "A", size = rel(2)),
      axis.text = element_text(family = "A", size = rel(1.5)),
      legend.text = element_text(family = "A", size=10),
      legend.title = element_text(family = "A", size=13),
      legend.margin=margin(17, 2, -25, 2),
      plot.margin = margin(0.6,0.6 , 0.6, 0.3, "cm"),
      axis.title.x = element_text(family = "A", vjust = -0.8),
      axis.title.y = element_text(family = "A", vjust = 1.3))

```

NYCU

附錄（五）：程式碼—中介效應分析

中介效應分析之程式碼如下所示，以菌屬之資料處理為例：

```
library(phyloseq)
library(msm)
library(readxl)
library(compositions)
library(stringr)
library(kableExtra)
library(glmnet)
library(tidyverse)
library(microbiome)

load("Hill330.rda")
ps<-ps_rar
new_otu1 = t(apply(ps@otu_table,1,function(x) x/sum(x)))
ps1 <- phyloseq(otu_table(new_otu1, taxa_are_rows=FALSE),
               sample_data(ps@sam_data),
               tax_table(ps@tax_table))
asv<-data.frame(ps1@otu_table)
taxa <- as.data.frame(ps1@tax_table)
metadata_new <- data.frame(ps1@sam_data)
metadata_new<-metadata_new[,-c(213:217)]
metadata_new[metadata_new=="NA"]<-NA

med_pvalue = function(real, std){
  z <- (real-0)/(std)
  return(2 * pnorm( -abs(z)))
}

all_in_reg <- function(data,n){
  data_new1 <- data[!is.na(data$X)&!is.na(data$fruits_or_vegetables),]
  y <- as.factor(data_new1[,1])
  x <- as.matrix(data_new1[,2:ncol(data_new1)])
  get_beta <- function() {
    train_index <- sample(1:length(y), 270, replace = FALSE)
    x_train <- x[train_index,]
    x_test <- x[-train_index,]
    y_train <- y[train_index]
    cv.lasso <- cv.glmnet(x_train, y_train, alpha = 0, family = binomial(link="logit"))
    lasso_mod <-
      glmnet(x_train, y_train, family = binomial(link="logit"), alpha = 0,
            lambda = cv.lasso$lambda.min)
```

```

    beta_result <-
      coef(lasso_mod) %>% t() %>% as.matrix() %>% as.data.frame()
    return(beta_result)
  }
  beta_list <- list()
  for (i in 1:n) {
    beta_list[[i]] <- get_beta()
  }
  beta_df <-
    beta_list %>%
    bind_rows()
  result <- beta_df %>%
    gather() %>%
    group_by(key) %>%
    summarise(
      beta = mean(value),
      sd = sd(value)
    )
  return(result)
}

reg_coefs_all_new = function(data){
  bootstrap <- data.frame()
  M1 = glm(Y ~ X+age+sex+fruits_or_vegetables, data=data,family=bino-
mial(link="logit"))
  beta1 = summary(M1)$coefficients[2,1]
  OR1 = exp(beta1)
  stdbeta1 = summary(M1)$coefficients[2,2]
  stdOR1 = sqrt(OR1*stdbeta1^2*OR1)

  M3 = all_in_reg(data,1000)
  beta3 = M3[M3$key=="X",]$beta
  stdbeta3 = M3[M3$key=="X",]$sd
  OR3 = exp(beta3)

  PE_all_in = beta1-beta3
  stdPE_all_in = sqrt(stdbeta1^2+stdbeta3^2)
  pPE_all_in = med_pvalue(PE_all_in, stdPE_all_in)

  Proportion_Eliminated = PE_all_in/beta1

  results=matrix(
c(paste0(round(beta1,5), "(" ,round(stdbeta1,5), ")"),paste0(round(beta3,5), "("
",round(stdbeta3,5), ")"),"",paste0(round(PE_all_in,5), "(" ,round(stdPE_all_i
n,5), ";" , "[" ,round(PE_all_in-
```

```

qnorm(.975,0,1)*stdPE_all_in,4),",",round(PE_all_in+qnorm(.975,0,1)*stdPE_a
ll_in,4),"]",")"),"",paste(round(Proportion_Elimi-
nated,4)*100,"%"),"",round(pPE_all_in,5),"", "", "", "", "", "", "", "", "",
w = 2,ncol = 9)
  colnames(results) = c('Estimate(SE)', 'PE(SE,[95%CI])', 'Proportion Elim-
inated(%)', 'P-value', "b1_hat", 'P-value of b1', "c2_hat", "P-value of
c2", "max(P-value of b1,P-value of c2)")

  return(list(bootstrap=M3,results=results))
}

```

陽明交大
NYCU

```

reg_coefs_all_new1 = function(data){
  M1 =
  glm(Y ~ X+age+sex+fruits_or_vegetables, data=data,family=binomial(link="logit"))
  beta1 = summary(M1)$coefficients[2,1]
  OR1 = exp(beta1)
  stdbeta1 = summary(M1)$coefficients[2,2]
  stdOR1 = sqrt(OR1*stdbeta1^2*OR1)

  result_M <- data.frame()
  for(i in 6:ncol(data)){
    formula <-
    as.formula(paste("Y", " ~ ", "X+age+sex+fruits_or_vegetables+", colnames(data)[i]))
    M <- glm(formula,data = data,family = binomial(link="logit"))
    beta2 = summary(M)$coefficients[2,1]
    OR2 = exp(beta2)
    stdbeta2 = summary(M)$coefficients[2,2]
    stdOR2 = sqrt(OR2*stdbeta2^2*OR2)
    PE = beta1-beta2
    stdPE = sqrt(stdbeta1^2+stdbeta2^2)
    pPE= med_pvalue(PE, stdPE)
    Proportion_Eliminated = (beta1-beta2)/beta1

    formula1 <- as.formula(paste(colnames(data)[i], "~", "X+age+sex+fruits_or_vegetables"))
    M1 <- lm(formula1,data=data)
    beta3 <- summary(M1)$coefficients[2,1]
    pbeta3 <- summary(M1)$coefficients[2,4]

    beta4 <- summary(M)$coefficients[6,1]
    pbeta4 <- summary(M)$coefficients[6,4]

    max <- max(pbeta3,pbeta4)

    result_M[i-5,1] <- paste0(round(beta2,4), "(", round(stdbeta2,4), ")")
    result_M[i-5,2] <-
    paste0(round(PE,4), "(", round(stdPE,4), ";", round(PE-
    qnorm(.975,0,1)*stdPE,4), ",", round(PE+qnorm(.975,0,1)*stdPE,4), ")", ")")
    result_M[i-5,3] <- paste(round(Proportion_Eliminated,4)*100,"%")
    result_M[i-5,4] <- round(pPE, 4)
    result_M[i-5,5] <- round(beta3, 4)
    result_M[i-5,6] <- round(pbeta3, 4)
    result_M[i-5,7] <- round(beta4, 4)
    result_M[i-5,8] <- round(pbeta4, 4)
    result_M[i-5,9] <- round(max, 4)
  }
}

```

```

    }
    colnames(result_M) <- c('Estimate(SE)', 'PE(SE, [95%CI])', 'Proportion Elim-
inated(%)', 'P-value', 'b1_hat', 'P-value of b1', 'c2_hat', 'P-value of
c2', 'max(P-value of b1, P-value of c2)')
    rownames(result_M) <- colnames(data)[6:ncol(data)]
    result_M <- result_M[order(result_M$`max(P-value of b1, P-value of c2)`),]
    return(result_M)
}

library(stringr)
genus <- data.frame(ps1@tax_table)[,6]
new_otu = data.frame(genus = genus, t(ps1@otu_table))
new_otu = new_otu[which(!is.na(new_otu[,1])),]

few <- c()
for(i in 1:nrow(new_otu)){
  k=length(which(new_otu[i,-1]>0))
  if(k<33){
    few <- c(few,i)
  }
}

new_otu1 <- new_otu[-few,]
new_otu2 = aggregate(new_otu1[,2:331], by = list(genus = new_otu1$genus),
FUN=sum)
new_otu2 = setNames(data.frame(t(new_otu2[, -1])), new_otu2[,1])
k <- min(new_otu2[new_otu2>0])/100
new_otu2 <- new_otu2+k
asv_new_log <- as.data.frame(log(new_otu2))

metadata_select <- metadata_new%>%select(case_control,ss_constipa-
tion,age,fruits_or_vegetables,sex)%>%as.data.frame()
data_new <- cbind(metadata_select,asv_new_log)
data_new$fruits_or_vegetables <- ifelse(data_new$fruits_or_vegetables=="At
least once a day",1,ifelse(is.na(data_new$fruits_or_vegetables),NA,0))
colnames(data_new)[1:2] <- c("Y","X")
k <- colnames(data_new)
kk <- stringr::str_replace_all(k,"-","_")
kk <- stringr::str_replace_all(kk,"/","_")
colnames(data_new) <- kk
data_new$Y <- ifelse(data_new$Y=="Control",0,1) #PD_CTRL
data_new$X <- ifelse(data_new$X=="No",0,1) #constipation
data_new$sex <- ifelse(data_new$sex=="female",0,1) #sex
data_new$Y <- as.factor(data_new$Y)

result_g <- reg_coefs_all_new(data_new)

```

```

result_g_M <- reg_coefs_all_new1(data_new)
Genus_1000 <- result_g$bootstrap
result_g <- result_g$results
resultg_M_new <- rbind(result_g,result_g_M)
rownames(resultg_M_new)[1:2] <- c("Unadjusted(TE)","All_in")

kable(resultg_M_new, booktabs = TRUE,align=c(rep('c',times=39))) %>%
pack_rows(
  index = c("Unadjusted" = 1, "Adjusted for" = 108)
)%>%
  kable_styling(font_size = 12)
Genus_1000$`p-value` <- round(med_pvalue(Genus_1000$beta,Genus_1000$sd),4)
Genus_1000 <- Genus_1000[Genus_1000$beta!=0,]
Genus_1000 <- Genus_1000[order(Genus_1000$`p-value`,decreasing = FALSE),]

```

陽明交大
NYCU