

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

口腔微生物組與口腔鱗狀細胞癌復發及五年存活之因果關係

Causal Relationship between Oral Microbiome and Recurrence

and Five-Year Survival in Oral Squamous Cell Carcinoma

研究生：陳品瑄（Chen, Ping-Hsuan）

指導教授：林聖軒（Lin, Sheng-Hsuan）

中華民國一一四年八月

August 2025

口腔微生物組與口腔鱗狀細胞癌復發及五年存活之因果關係

Causal Relationship between Oral Microbiome and Recurrence  
and Five-Year Survival in Oral Squamous Cell Carcinoma

研究生：陳品瑄

指導教授：林聖軒

Student : Chen, Ping-Hsuan

Advisor : Dr. Sheng-Hsuan, Lin



August 2025  
Taiwan, Republic of China  
中華民國 一一四年八月

## 誌謝

首先，謹向我的指導教授林聖軒老師致上最誠摯的謝意。感謝老師在研究過程中提供專業的學術指導，並在細節上給予我寶貴的指教與建議，使我能夠不斷修正並完善我的研究。同時，也感謝傅詩宸老師，在微生物分析的部分給予我許多協助與指導，補助我在生物知識方面的不足。另外也非常感謝黃冠華老師在機器學習的部分給予我很多指導與建議。亦感謝 408 研究室的同學們，在生活中給予我支持與陪伴，特別是我們在研究室一同度過的時光，不論是討論研究、一起吃飯、聊天或玩樂，都讓我的研究所生活更加充實且難忘。此外，我還要特別感謝嘉翎，在討論與分享研究進度中給了我許多幫助。

同時，也要感謝一路陪伴我的朋友們——胡宛欣、胡宛欣的男朋友、林奕亘、鄭鈺靜、張筠笙、喻怡鳳和謝昀庭，謝謝你們在我研究過程中給予的鼓勵與關心，讓我在面對壓力與挑戰時獲得陪伴與溫暖。最後，感謝我的父母與姊姊一路以來的鼓勵與支持，讓我能夠專心完成學業與研究。這份論文不僅是我個人的努力成果，更是所有一路支持我的人共同的成果。謹以此誌，向所有在本研究過程中曾經幫助過我的師長、同儕與親友，表達我最深的感謝。

陳品瑄 謹致於

國立陽明交通大學 統計學研究所

中華民國一一四年八月

## 中文摘要

口腔鱗狀細胞癌（Oral squamous cell carcinoma, OSCC）為全球常見的口腔惡性腫瘤，尤其在東亞地區發病率與死亡率皆居高不下。復發通常發生於治療後的前兩年內，且早期復發與較差的生存預後密切相關。目前針對口腔微生物與 OSCC 復發或存活的關聯研究仍相對有限，既有文獻多未考慮復發時間點、競爭風險或干擾因子控制，對其因果關係的釐清亦不足。本研究納入台灣長庚紀念醫院 181 名 OSCC 患者的口腔沖洗液樣本，進行 16S rRNA 擴增子測序，並透過 Alpha 與 Beta 多樣性分析、相對豐度分析與功能預測探討口腔微生物與復發及五年存活狀態之關聯。在復發分析中，我們將復發依照時間分期，並排除尚未復發即死亡之個案以避免競爭風險干擾。結果顯示，復發組的物種豐富度顯著低於無復發組，微生物組成也存在顯著差異。相對豐度分析指出復發組與特定微生物變化相關，而功能路徑分析則顯示內分泌系統、細胞生長與凋亡、消化系統等代謝路徑有顯著差異。於五年存活分析中，死亡組同樣呈現較低的 Alpha 多樣性，並在脂質代謝與轉錄等功能上與存活組存在顯著不同。此外，我們建構 XGBoost 預測模型，比較臨床參數、微生物資料與兩者結合三種模型的預測效能。結果顯示混合模型在三年內復發的預測中表現最佳（AUC = 0.8495），具備較高的準確率、敏感性與特異性。本研究透過調整干擾因子並考慮中介變項，進一步釐清口腔微生物與 OSCC 復發及存活之因果關係，並證實其在風險預測與臨床應用上的潛力，未來可望發展為個人化治療與預後管理的重要工具。

**關鍵字：** 口腔鱗狀細胞癌、口腔微生物相、復發、五年存活狀態、因果關係、XGBoost

## 英文摘要

Oral squamous cell carcinoma (OSCC) is a common malignant tumor of the oral cavity, with high recurrence rates and poor prognosis, especially in East Asian populations. Most recurrences occur within the first two years after treatment and are closely associated with unfavorable survival outcomes. Although emerging studies suggest a potential association between the oral microbiome and OSCC recurrence or survival, few have considered recurrence timing, competing risks, or confounding factors, leaving the causal relationship unclear. In this study, we collected oral rinse samples from 181 OSCC patients at Chang Gung Memorial Hospital in Taiwan and conducted 16S rRNA amplicon sequencing to examine associations between the oral microbiome and both recurrence and five-year survival outcomes. For recurrence analysis, we stratified patients by recurrence time and excluded cases where patients died before recurrence to minimize competing risk bias. Results showed significantly lower species richness in the recurrence group and distinct microbial community compositions between groups. Differential abundance analysis indicated microbial changes associated with recurrence, and functional prediction revealed significant differences in pathways related to the endocrine system, cell growth and death, and the digestive system. In survival analysis, the deceased group also exhibited lower alpha diversity, along with significant differences in lipid metabolism, transcription, and other functional pathways. Furthermore, we developed XGBoost machine learning models using clinical data, microbiome data, and a combined dataset to predict recurrence. The combined model achieved the best performance for predicting recurrence within three years ( $AUC = 0.8495$ ), with higher accuracy, sensitivity, and specificity. By adjusting for age, sex, and clinical stage, and accounting for mediating factors such as recurrence and complications, this study provides new evidence of a potential causal relationship between the oral microbiome and OSCC outcomes. These findings highlight the clinical utility of incorporating microbial data into personalized risk assessment and prognostic decision-making.

**Keywords:** Oral squamous cell carcinoma, Oral microbiome, Recurrence, Five-year survival, Causal relationship, XGBoost

## 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 口腔鱗狀細胞癌介紹 .....	1
1.1.2 口腔微生物與口腔鱗狀細胞癌之關係 .....	2
1.1.3 口腔微生物與復發過往之研究結果 .....	2
1.1.4 口腔微生物與五年後存活狀態之現有研究結果 .....	3
1.2 研究目的 .....	4
1.3 研究架構 .....	4
第二章、方法與材料 .....	5
2.1 研究對象及資料蒐集方式 .....	5
2.2 口腔微生物資料處理 .....	5
2.3 人口統計特徵方法 .....	6
2.3.1 復發分析 .....	6
2.3.2 五年存活狀態分析 .....	6
2.4 口腔菌資料分析方法 .....	8
2.4.1 Alpha 多樣性及 Beta 多樣性 .....	8
2.4.2 相對豐度差異分析 .....	10
2.4.3 功能路徑豐度差異分析 .....	10
2.5 XGBoost 模型預測 .....	11
第三章、復發 .....	12

3.1 人口統計特徵分析結果.....	12
3.2 Alpha 多樣性及 Beta 多樣性分析結果 .....	14
3.3 相對豐度差異分析結果.....	15
3.4 功能路徑豐度差異分析結果.....	16
3.5 XGBoost 模型預測結果.....	17
第四章、五年存活狀態.....	22
4.1 人口統計特徵分析結果.....	22
4.2 Alpha 多樣性及 Beta 多樣性分析結果 .....	23
4.3 相對豐度差異分析結果.....	24
4.4 功能路徑豐度差異分析結果.....	26
第五章、討論.....	27
5.1 口腔微生物組與復發之關係.....	27
5.1.1 口腔微生物組之結果探討.....	27
5.1.2 功能路徑之結果探討.....	28
5.1.3 XGBoost 模型預測之結果探討.....	28
5.2 口腔微生物組與五年後存活狀態之關係.....	29
5.2.1 口腔微生物組之結果探討.....	29
5.2.2 功能路徑之結果探討.....	30
5.3 研究限制.....	30
5.4 未來展望.....	31
參考文獻.....	32

## 圖目錄

圖（一）、口腔微生物與五年存活狀態之因果關係圖 .....	8
圖（二）、是否在（a）一年內復發（b）兩年內復發（c）三年內復發之兩組患者 Alpha 多樣性指標差異.....	14
圖（三）、是否在（a）一年內復發（b）兩年內復發（c）三年內復發之兩組患者 Beta 多樣性指標差異.....	15
圖（四）、不同復發時間段內兩組受試者相對豐度差異 菌屬層級：（a）一年內復發 （b）兩年內復發（c）三年內復發 菌種層級：（d）一年內復發（e）兩年 內復發（f）三年內復發 .....	16
圖（五）、不同復發時間段內兩組受試者之微生物組功能差異（上：一年內復發，中： 兩年內復發，下：三年內復發） .....	17
圖（六）、OSCC 復發預測模型表現（a）是否復發（b）是否在兩年內復發（c）是否在 三年內復發，由左至右分別為模型一、模型二及模型三.....	20
圖（七）、OSCC 復發預測模型三特徵重要性，由左至右分別為是否復發、是否在兩年內 復發及是否在三年內復發 .....	21
圖（八）、五年後是否存活兩組患者 Alpha 多樣性指標差異（a）校正干擾因子（b）校 正干擾因子與中介因子.....	24
圖（九）、五年後是否存活兩組患者 Beta 多樣性指標差異（a）校正干擾因子（b）校正 干擾因子與中介因子.....	24
圖（十）、五年後存活狀態之兩組受試者之相對豐度差異 菌屬層級：（a）僅校正干擾因 子（b）校正干擾因子及中介因子 菌種層級：（c）僅校正干擾因子（d）校 正干擾因子及中介因子 .....	25
圖（十一）、（上）無分群（中）無復發人群（下）復發人群五年後是否存活兩組受試者 之功能路徑差異 .....	26

## 表目錄

表（一）、Alpha 多樣性指標介紹 .....	9
表（二）、是否復發兩組患者之人口統計特徵比較 .....	13
表（三）、XGBoost 預測結果指標 .....	18
表（四）、模型三之混淆矩陣（Confusion Matrix） .....	19
表（五）、五年後是否存活兩組患者人口統計特徵之比較 .....	22



# 第一章、前言

## 1.1 研究背景

### 1.1.1 口腔鱗狀細胞癌介紹

口腔鱗狀細胞癌（Oral Squamous Cell Carcinoma, OSCC）是全球常見且具高度侵襲性的頭頸部癌症，根據世界衛生組織（WHO）國際癌症研究署（IARC）於 2020 年發布的 GLOBOCAN 報告[1]，全球每年約有超過 37 萬 OSCC 新增病例，並造成近 18 萬人死亡。在亞洲地區，尤其在東亞和東南亞國家中，其發病率和死亡率都位居前列。在台灣，因檳榔咀嚼、吸菸及飲酒等高危險行為普遍，OSCC 的發生率與死亡率居高不下，OSCC 位居男性癌症死亡率的第四名[2]，對公共衛生形成重大挑戰。OSCC 以快速進展及高轉移率為特徵，通常由原發病灶擴散至頸部淋巴結。由於早期症狀不明顯，許多患者確診時已屬晚期，治療難度因此加大。儘管外科手術、放射治療與化學治療等多元療法日漸成熟，OSCC 患者的五年存活率仍僅約 60%[3]，且多項研究結果皆指出，即使進行原發性腫瘤的根治性手術後，復發率仍介於 7 - 47.4%[4-10]。另外，早期復發（定義為治療後 18 個月內）與較差的生存預後高度相關；若患者一旦復發，其中位生存期僅約 6 至 9 個月，若再發生遠端轉移，生存期則進一步縮短至 3 至 4 個月。針對早期 OSCC 患者，為預防頸部淋巴結轉移所進行的「頸部淋巴廓清術」長期以來被視為常見的治療策略，然而，卻可能帶來生活品質與功能面向的嚴重影響[11]。因此，臨床亟需在不犧牲生活品質的前提下，更準確地評估患者之轉移與復發風險，並據此決定預防性手術與後續治療策略，已成為一項重要挑戰。

### 1.1.2 口腔微生物與口腔鱗狀細胞癌之關係

近年來，越來越多研究指出，口腔微生物與 OSCC 的發展及預後之間具有密切關聯[12, 13]。一些初步證據顯示，OSCC 患者的口腔微生物組成可能與復發風險與長期存活率相關；特定菌種與代謝功能途徑的改變，可能透過影響腫瘤微環境或局部免疫反應，進一步干擾腫瘤生長和轉移。然而，相關研究多存在樣本規模不足、資訊不平衡以及干擾因子（例如年齡、性別、臨床分期、生活習慣等）未被充分控制等限制，導致口腔微生物在 OSCC 復發預測與五年後存活狀態方面的因果關係仍有待釐清。在機器學習領域，已有研究嘗試應用隨機森林及若干微生物標誌物來預測 OSCC 是否復發[14]，雖在訓練集展現一定潛力，卻因測試集樣本量偏少且分布不均，模型的準確度與臨床應用價值尚顯不足。後續工作開始嘗試更先進的演算法（如 XGBoost）並擴大樣本數，以期提升在多重指標（如靈敏度、特異性及 AUC）上的表現，同時更全面地整合臨床參數與口腔微生物特徵，以強化預測能力。

### 1.1.3 口腔微生物與復發過往之研究結果

目前，針對口腔微生物與 OSCC 復發之間關聯的研究非常有限，迄今為止僅有一項研究分析了 54 名 OSCC 患者的口腔拭子樣本[14]，並發現復發患者的口腔微生物群與非復發患者存在顯著差異，其中 *Granulicatella*、*Peptostreptococcus*、*Campylobacter*、*Porphyromonas*、*Oribacterium*、*Actinomyces*、*Corynebacterium*、*Capnocytophaga* 和 *Dialister* 等菌屬在復發患者中更為豐富。口腔微生物的功能路徑在不同組別間也有所不同，在經歷復發的患者中，與氨基酸、碳水化合物和脂質代謝相關的代謝功能發生了改變[14]。然而，儘管該項研究已指出口腔微生物與 OSCC 復發之間的關聯，卻仍存在幾個不足。在討論 OSCC 復發風險時，僅考慮復發並不足夠，還必須考慮與死亡風險的競爭關係（competing risks）。該研究未能解決復發時間與死亡風險的競

爭效應問題，也未對不同的復發時間區間進行探討，而是將復發視為一個整體事件，這使得口腔微生物群與 OSCC 復發的具體時間點之間的關聯仍不明確。此外，該研究未能有效控制患者的年齡、性別及臨床分期等潛在干擾因子。例如，口腔微生物群落較差的患者中，許多屬於臨床第四期，這類患者本身復發風險較高，難以確定復發是由口腔微生物變化引起，還是受臨床分期等其他因素影響，因此，無法明確推論口腔微生物與 OSCC 復發之間的因果關係。

除此之外，該研究也嘗試以隨機森林模型結合五種微生物標誌物，開發非侵入性的復發預測工具[14]；然而，其訓練樣本量偏少（僅 54 例），測試集也因極度不均的樣本分布而導致模型評估結果不穩定（47 例中僅 4 例復發），致使模型的準確率和 AUC 難以充分證明對臨床診斷的貢獻。此外，其基準模型 AUC 僅為 0.438，低於隨機預測水平（0.5），顯示基準模型的設計過於薄弱，難以充分證明新模型的優越性，也凸顯了建立更可靠且具臨床應用價值的預測模式仍然是一項挑戰。

#### 1.1.4 口腔微生物與五年後存活狀態之現有研究結果

一項系統性回顧及統合分析，納入了 15 項研究，共計 5,191 個樣本，結果顯示口腔微生物多樣性較低的患者或是口腔微生物組中富含 *Fusobacterium* 和 *P. gingivalis* 的患者，其整體生存期、疾病特異性生存期和無病生存期結果較差 [15]。而其中 Mohamed, Littlekalsøy [16]等人的研究結果顯示出 *Candida* 屬的高唾液攜帶量與較差的 OSCC 患者總生存率（OS）顯著相關。*Malassezia* 屬的高唾液攜帶量則與 OSCC 患者較好的總生存率顯著相關。*Malassezia* 在 OSCC 患者的整體生存率（OS）中被鑑定為一個獨立且有利的預測因子。該研究僅關注單一菌種對疾病的影響，而忽略了微生物群整體結構的分析，這樣的局限性可能會低估微生物群之間相互作用對疾病進展的綜合影響。其次，該研究在分析微生物相時，只局限於相對豐度和 Alpha 多樣性的分析，缺乏 Beta 多樣性和功能預

測分析，這限制了對微生物群體結構差異和代謝功能的理解。此外，該研究未對復發、癌症分期和併發症等干擾因子進行充分調整，導致對微生物群與存活狀態關聯性的評估結果不夠精確。實際上，有許多因素會影響口腔癌患者的預後，包括年齡、性別、癌症分期、復發、治療方式、營養狀況等[17, 18]。研究顯示，口腔癌的復發和晚期分期顯著影響患者的生存率，尤其是在第四期癌症患者中，復發和死亡風險顯著增加[17]。此外，營養狀況和全身性炎症反應也被認為是預後的重要影響因素，低營養預後指數（PNI）與較差的生存結果顯著相關[18]。因此，我們必須在分析中控制這些潛在的干擾因子，才能更準確地評估口腔微生物對五年後存活狀態的影響。

## 1.2 研究目的

基於以上背景，本研究著眼於台灣長庚紀念醫院所收集之 OSCC 病患資料，深入探討口腔微生物的組成及其在 OSCC 復發與五年後存活預測中的角色，並將針對不同的時間區間，同時考量死亡競爭風險與干擾因子，透過綜合的生態學與生物資訊分析方法，評估口腔微生物在 OSCC 預後風險評估上的潛在應用價值。最終目標為協助臨床更精準地辨識高風險患者，並提供可行的輔助診斷依據，進一步提升 OSCC 患者的整體生存率與生活品質。

## 1.3 研究架構

本研究各章節將依序說明資料處理流程、分析方法與研究結果。第二章首先介紹研究對象、樣本來源與微生物資料的前處理方式，並說明人口統計分析、多樣性分析、相對豐度與功能差異分析等方法，最後說明用於復發預測的 XGBoost 模型建構與評估方式。第三章與第四章分別對應本研究的兩大主題：第三章聚焦於復發，依序呈現人口統計特徵、微生物多樣性、菌群組成與功能

變化分析結果，並呈現 XGBoost 模型對復發風險的預測表現。第四章則針對五年後存活狀態進行分析，探討其與口腔微生物多樣性與功能的潛在關聯。第五章將整合研究發現，討論口腔微生物在 OSCC 復發與長期預後中的可能機制與臨床意涵，並說明本研究的限制與未來研究的發展方向。

## 第二章、方法與材料

### 2.1 研究對象及資料蒐集方式

本研究採用從 2014 年 2 月至 2015 年 5 月於台灣長庚紀念醫院收集之 203 份口腔沖洗液樣本，並在五年後通過醫療記錄進行了隨訪。在收集樣本之前，均已獲得所有患者或其家屬之書面知情同意。所有參與者在採集樣本時均未接受抗生素治療。排除非手術患者後，剩下的 181 名患者被用作後續分析。口腔沖洗液樣本在疾病診斷後 1-3 周內採集。所有 OSCC 的診斷均由活檢和病理結果證實。患者用 30 毫升無菌生理鹽水漱口 30 秒，然後吐入 50 毫升無菌試管中。收集樣本後，在 6,000 rpm 轉速下離心 30 分鐘，收集細胞部分並保存在 -80°C 溫度下，直至使用。DNA 之萃取使用 QIAamp DNA Microbiome Kit 從口腔漱口水樣本中分離細菌的總基因組 DNA。移除宿主 DNA 後，細菌 DNA 經過純化，並儲存在 -80°C。DNA 濃度與品質則使用 Qubit 高靈敏度 dsDNA 分析儀進行測定。

### 2.2 口腔微生物資料處理

透過針對 16S rRNA V3-V4 區域的 PCR 擴增，構建了擴增子文庫，並在基因特異性序列上添加了 Illumina 接頭。PCR 產物經過純化後，使用 MiSeq 系統進行測序。測序讀數首先根據樣本條形碼使用 MiSeq Reporter 進行解複用，然後合併成對讀數以獲得較長的讀數，並透過 FLASH (V1.2.11)[19]改進分類方法。

只有合併讀數達到或超過 100,000 的樣本才被保留用於後續分析。接著，使用 QIIME 和 Cutadapt 軟體移除測序引子，並使用 USEARCH 軟體刪除嵌合序列，同時過濾掉 PHRED 分數（Phred quality score）低於 20 的序列。接著利用 Deblur v1.1.1 進行序列的去噪處理[20]，並將序列長度裁剪至 427 個鹼基對（Base pair, bp）。最後基於 scikit-learn 的單純貝氏分類器（Naive Bayes classifier）進行分類，該分類器使用 Silva 138 99% 擴增子序列變體（Amplicon sequence variant, ASV）的全長序列中的 V3-V4 區域進行訓練，以利後續分析。

## 2.3 人口統計特徵方法

### 2.3.1 復發分析

我們根據特定時間區間內的復發狀況，將受試者分為兩組，並以一年、兩年及三年作為時間區隔，進行人口統計特徵的差異比較。原先亦考慮了半年與五年作為分析的時間點。然而，由於半年內發生復發的患者比例過低，難以進行具統計效力的比較，故不予採用。而五年的時間區隔未經復發而死亡的個案比例較高，因此亦未納入分析範疇。本篇研究中欲比較之特徵可分為三類，一為基本資訊及病史，包含：年齡、性別及是否有糖尿病；二為臨床狀況，包含在：總臨床分期、是否傷口感染及術後是否引發吸入性肺炎；三為臨床結果，包含：五年後存活狀態及是否於特定時間區間內死亡。連續型變數採用 Kruskal-Wallis 檢定（Kruskal-Wallis Test）；類別型變數採用卡方檢定（Chi-square Test），並將結果以表格呈現。

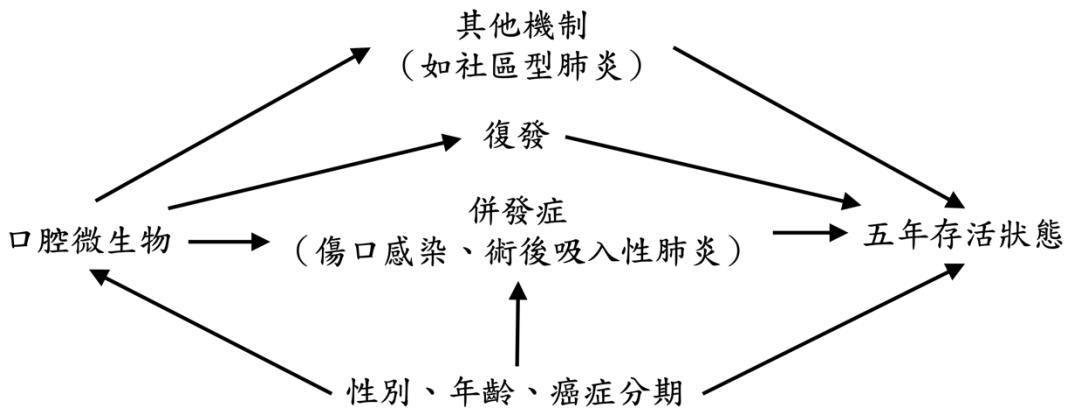
### 2.3.2 五年存活狀態分析

我們以五年後之存活狀態將受試者分為兩組，分別為五年後存活組與五年後死亡組，並比較兩組在人口統計特徵上是否有所差異。本篇研究中欲比較之特徵可分為兩類，一為基本資訊及病史，包含：年齡、性別及是否有糖尿病；

二為臨床狀況，包含在：總臨床分期、是否傷口感染及術後是否引發吸入性肺炎。連續型變數採用 Kruskal-Wallis 檢定（Kruskal-Wallis Test）；類別型變數採用卡方檢定（Chi-square Test），並將結果以表格呈現。

## 因果關係

如圖（一）所示，口腔微生物可能通過多條路徑影響患者五年存活狀態。在本研究中，我們採用三種分析方法來探討口腔微生物與五年存活狀態之間的因果關係，分別是未校正模型、校正干擾因子的模型以及同時校正干擾因子與中介因子的模型。首先，口腔微生物為研究中的主要暴露因子（因），而患者的五年後存活狀態則為結果（果）。在未校正的模型中，我們直接觀察口腔微生物與五年存活狀態之間的關聯，這種方式未考慮到任何其他變數的影響，因而可能受到各種干擾因子的影響。而在校正干擾因子的模型中，我們將年齡、性別及臨床分期視為潛在的干擾因子，因為這些因子都會影響口腔微生物、五年存活狀態及中介因子，因此在校正模型中納入考量。此外，復發與併發症被視為中介因子，因為它們可能是口腔微生物通過影響這些過程間接影響存活結果的路徑。因此，在最後一種模型中，我們同時控制了干擾因子和中介因子，以釐清口腔微生物在不受到干擾因子的影響下，經由其他機制或是直接影響到五年存活狀態的效果。這樣的分析能進一步確認口腔微生物是否會透過與復發和併發症無關的其他機制來影響患者的長期預後，並提供了控制多種潛在影響因素後的清晰解釋。



圖（一）、口腔微生物與五年存活狀態之因果關係圖

## 2.4 口腔菌資料分析方法

### 2.4.1 Alpha 多樣性及 Beta 多樣性

為了探討是否復發以及五年後是否存活與口腔微生物之間的關聯性，本研究首先分析了 Alpha 多樣性。Alpha 多樣性反映了單一環境中的微生物豐富度 (Richness) 和均勻度 (Evenness)。本文選用了四種常見的 Alpha 多樣性指標：Observed、Chao1[21]、Shannon[22] 和 Simpson[23]。Observed 和 Chao1 側重於豐富度，前者屬於實際觀察到的物種數量，後者則更注重低豐度稀有菌種，數值越大代表 Alpha 多樣性越高；Shannon 和 Simpson 則同時考慮豐富度與均勻度，範圍在 0 至 1 之間，數值越大通常代表菌群越均勻、多樣性越高。為了使指標與實際意涵對照更為清晰，本研究將各指標定義整理於表（一）。這些指標是使用 R 套件 phyloseq 計算得出，接著使用 ANOVA 檢定來比較有無復發的兩組人群之間的差異。為了校正潛在的干擾因子，本研究在 ANOVA 檢定中加入干擾因子作為控制變數，確保比較結果能夠排除這些潛在影響因素的干擾。最終，通過箱型圖來展示校正後兩組人群之間的差異。

表（一）、Alpha 多樣性指標介紹

指標	公式	解釋
Observed	$S_{observed}$ ：觀察到的物種數量	數值越高代表觀測到的物種數越多
Chao1	$S_{Chao1} = S_{observed} + \frac{F_1^2}{2F_2}$	數值越高代表觀測到的物種數越多，特別重視稀有菌
Shannon	$H = - \sum_{i=1}^R p_i \ln(p_i)$ $p_i$ 是第 i 個物種的相對豐度	數值越大代表菌相越均勻
Simpson	$D = - \sum_{i=1}^R p_i^2$	數值越接近 1 代表菌相越均勻

第二步分析的是 Beta 多樣性，用來衡量不同環境間物種組成的差異，即比較兩組樣本間口腔微生物組成的相似程度。Beta 多樣性通常通過計算距離矩陣來進行分析，然後使用無母數方法中的 PERMANOVA 檢定[19]來比較群體之間的相似程度。本研究中使用了兩種距離計算方法：一種基於系統發育樹的距離，如加權和非加權的 UniFrac 距離[24]，這些距離可使用 R 套件 phyloseq [25]計算；另一種是廣泛用於生態學研究的 Bray-Curtis 距離[26]，該距離可使用 R 套件 vegan[27]計算。計算距離後，我們使用 PERMANOVA 檢定來檢查有無復發的兩組人在口腔微生物組成上的差異，同時將干擾因子也加入到 PERMANOVA 模型中，確保分析結果更加準確。在 Beta 多樣性的呈現上，常使用微生物生態學中的一種降維方法，稱為主座標分析（Principal Coordinate Analysis，PCoA）。這種分析透過對多維數據進行特徵值和特徵向量排序，找出最重要的元素和結構。PCoA 基於不同的距離測量進行，並選取前兩個主要維度繪製散佈圖，便於觀察結果。

#### 2.4.2 相對豐度差異分析

我們分別針對有無復發的兩組人以及是否在五年後存活的兩組人進行口腔微生物豐度差異分析，以找出在豐度上有顯著差異的微生物。在分析前，為了消除樣本差異和測序過程的影響，我們先對絕對豐度表進行正規化處理。本研究採用了總和比例化（Total-sum scaling）方法，將絕對豐度表轉換為相對豐度表，然後使用 R 套件 glmmTMB 1.1.9-9000 中的 glmmTMB 函數來檢查兩組人之間的豐度差異，該函數使用帶有零膨脹的負二項分佈廣義線性模型（GLM）去進行計算，干擾因子也被納入 GLM 模型中，從而確保模型結果更加準確且能排除外部干擾的影響。為了校正多重比較問題，本研究進一步對 p 值進行 Benjamini-Hochberg（BH）校正[28]，以控制偽發現率（False Discovery Rate）。最後，利用火山圖呈現檢定結果，圖中橫軸為  $\log_2$ （變化倍數），縱軸為 $-\log_{10}$ （p-value）。為解決某些菌屬相對豐度為 0 的問題，我們在計算變化倍數時，對所有相對豐度加上了一個極小值，該極小值定為所有不為 0 的相對豐度中之最小值除上 100。

#### 2.4.3 功能路徑豐度差異分析

我們使用 White's 無母數 t 檢定（White's non-parametric t-test）[29]，這是一種去除常態分佈假設的檢定方法，設定為雙尾檢定，並重複 1000 次。檢定結果的 p-value 經過 Storey FDR 校正[30]後稱為 q-value，以 0.05 為顯著水準。q-value 小於 0.05 表示兩組在該功能路徑的相對豐度有顯著差異。本分析使用 STAMP（Statistical Analysis of Metagenomic Profiles）進行並繪製圖表。

## 2.5 XGBoost 模型預測

本研究採用 XGBoost (Extreme Gradient Boosting) [31]作為主要模型，這是一種以梯度提升決策樹 (Gradient Boosting Decision Trees, GBDT) [32]為基礎的方法，具備高效處理非線性特徵交互的能力，並能支援加權損失函數以解決類別不平衡問題。為了提高模型性能，本研究對 XGBoost 模型的超參數進行了最佳化。使用網格搜索 (Grid Search) 與 5 折交叉驗證 (Cross-validation) 調整以下參數：樹的深度 (max\_depth)：3 到 7，步長為 2；學習率 (learning\_rate)：0.01、0.1 及 0.3；子樣本比例 (subsample)：0.8 及 1.0；迭代數 (nrounds)：100、200 及 500；特徵子集比例 (colsample\_bytree)：0.8 及 1.0。最佳化目標為最大化交叉驗證過程中的 AUC (Area Under the Curve) [33]，以平衡模型在不平衡數據中的靈敏度 (Sensitivity) 與特異性 (Specificity)。

數據集按照 7:3 的比例隨機分為訓練集與測試集，並在切分過程中確保類別比例一致。其中訓練集用於模型構建與參數調整，測試集用於評估模型性能。考慮到少數類別樣本（復發樣本）的稀缺性，使用了 SMOTE (Synthetic Minority Oversampling Technique) [34]方法對訓練集進行少數類別的樣本增強。SMOTE 透過生成新的少數類別樣本來平衡類別分佈，而不是簡單地重複現有樣本。對於訓練集中每個少數類別樣本，SMOTE 首先計算該樣本與其他少數類別樣本之間的歐幾里得距離（使用 K 個最近鄰，默認 K=5）。然後，隨機選擇其中一個最近鄰，根據兩者之間的線性插值生成新樣本。插值公式如下：

$$x_{new} = x_{original} + \lambda \cdot (x_{neighbor} - x_{original})$$

其中， $\lambda$  是介於 0 和 1 之間的隨機數，確保新樣本分佈於原始樣本與其鄰居之間的空間中。這種方法避免了過度重複樣本，並透過生成基於特徵分佈的新樣本，保留了少數類別的多樣性，增強後的數據集可以有效提高分類模型對少數類別的識別能力。

本研究共設計了三種模型來預測口腔癌患者的復發情形：

模型一：基準模型（Baseline model），僅使用性別、年齡、癌症分期及傷口感染這四種基本臨床參數作為預測變數，用於提供基準效能的參照比對。

模型二：口腔微生物模型（Oral microbiome model, Oral），僅使用患者的口腔微生物資料進行預測，利用相對豐度探討口腔微生物對預測復發的影響。

模型三：混合模型（Combined model, Mix），將基準模型的臨床參數與口腔微生物特徵結合，以評估最終的預測效能。

為全面評估模型性能，本研究選擇以下指標：AUC，衡量模型在不同閾值下的分類性能；靈敏度與特異性，檢查模型對正例和負例的分類能力。在測試集上的最終結果包括以上指標，並與基準模型進行比較。

### 第三章、復發

#### 3.1 人口統計特徵分析結果

首先，我們比較了是否復發的兩組受試者的人口統計特徵，其結果統整於表（二）中。因變數眾多，僅列出後續作為干擾因子之變數。根據表中顯示的結果，其中年齡、性別在兩組間沒有顯著差異。癌症分期在有復發和無復發的兩組別出現顯著差異 ( $p\text{-value}<0.001$ )。在後續進一步的分析中，我們將年齡、性別和癌症分期視為潛在的干擾因子。儘管年齡和性別在復發與否的兩組間無顯著差異，這些變數仍可能對復發的結果產生影響，因此也列入干擾因子。

考量到競爭風險問題（復發與死亡風險的競爭），我們特別關注觀察期間內的復發事件，因而未對未復發的死亡個案進一步分析。競爭風險模型中，復發和死亡為互相競爭的事件：如果患者死亡，他們便不會再有復發的機會。在本研究中，我們刪除了未復發但死亡的患者，以避免這些案例對統計分析結果產

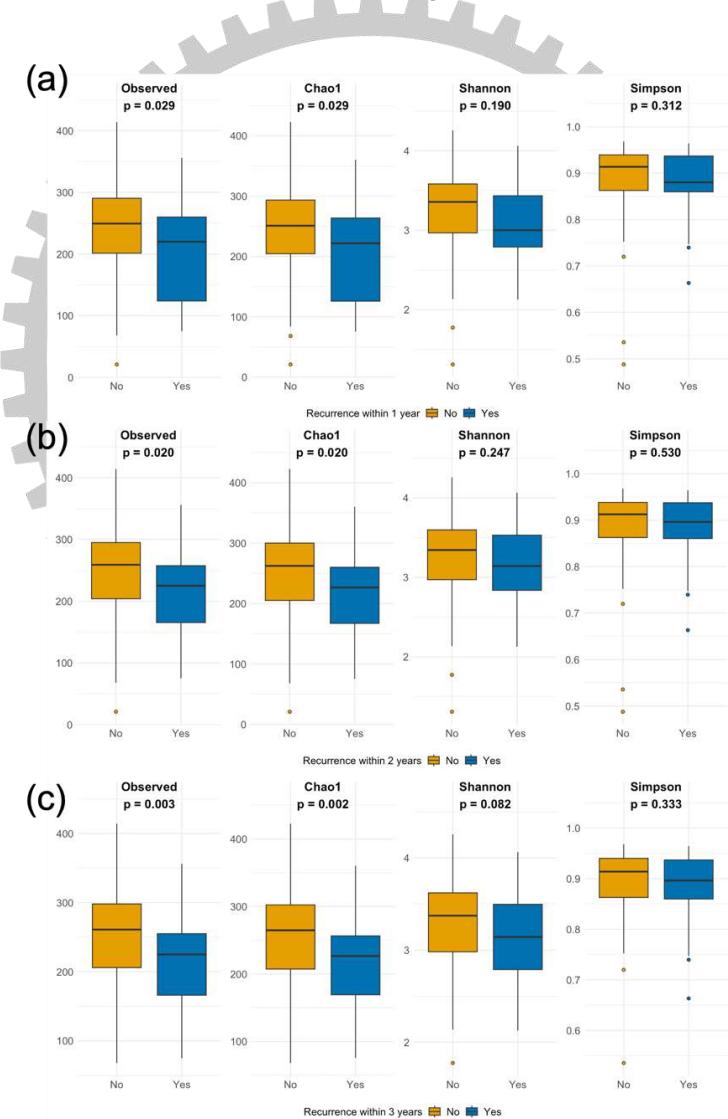
生干擾。由於這些患者既未經歷復發事件，也無法再有復發的機會，他們的存在可能會混淆復發風險的評估。透過刪除這些案例，我們能夠專注於那些在觀察期間內已經復發及未復發且仍然存活的患者，從而更準確地探討復發相關的風險因子。

表（二）、是否復發兩組患者之人口統計特徵比較

一年內復發			
	無 (N=150)	有 (N=25)	p-value
<b>年齡</b>			
平均數 (標準差)	53.68 (10.17)	50.58 (7.69)	0.148
<b>性別</b>			
女性	12 (8.0%)	3 (12.0%)	0.783
男性	138 (92.0%)	22 (88.0%)	
<b>總臨床分期</b>			
I, II, III	93 (62.0%)	7 (28.0%)	0.003
IV	57 (38.0%)	18 (72.0%)	
兩年內復發			
	無 (N=137)	有 (N=35)	p-value
<b>年齡</b>			
平均數 (標準差)	53.60 (10.42)	51.85 (7.94)	0.358
<b>性別</b>			
女性	10 (7.3%)	4 (11.4%)	0.652
男性	127 (92.7%)	31 (88.6%)	
<b>總臨床分期</b>			
I, II, III	87 (63.5%)	12 (34.3%)	0.003
IV	50 (36.5%)	23 (65.7%)	
三年內復發			
	無 (N=127)	有 (N=41)	p-value
<b>年齡</b>			
平均數 (標準差)	53.74 (10.68)	51.73 (7.71)	0.265
<b>性別</b>			
女性	10 (7.9%)	4 (9.8%)	0.957
男性	117 (92.1%)	37 (90.2%)	
<b>總臨床分期</b>			
I, II, III	83 (65.4%)	15 (36.6%)	0.002
IV	44 (34.6%)	26 (63.4%)	

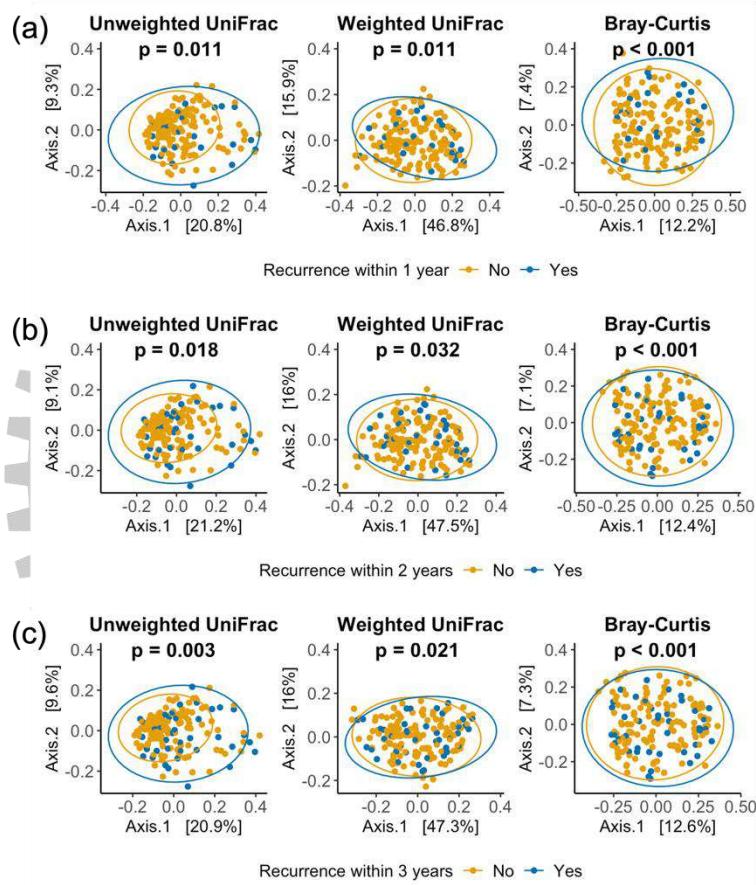
### 3.2 Alpha 多樣性及 Beta 多樣性分析結果

依據復發的時間點分別進行 Alpha 多樣性的檢驗，分為是否在一年內、兩年內及三年內復發，可得到圖（二）的結果。在 Observed 和 Chao1 指標中，皆觀察到了復發組有顯著較低的 Alpha 多樣性，而在 Shannon 和 Simpson 指標中，復發組仍有較低的 Alpha 多樣性，但未達統計上顯著（一年： $p_{observed} = 0.029, p_{Chao1} = 0.029, p_{Shannon} = 0.190, p_{Simpson} = 0.312$ ；兩年： $p_{observed} = 0.020, p_{Chao1} = 0.020, p_{Shannon} = 0.247, p_{Simpson} = 0.530$ ；三年： $p_{observed} = 0.003, p_{Chao1} = 0.002, p_{Shannon} = 0.082, p_{Simpson} = 0.333$ ）。



圖（二）、是否在（a）一年內復發（b）兩年內復發（c）三年內復發之兩組患者 Alpha 多樣性指標差異

結果如圖（三）所示：我們發現基於未加權 UniFrac、加權 UniFrac 和 Bray-Curtis 距離的計算上，在所有情況中是否復發的兩群微生物組成皆有顯著的差異（一年： $p_{Unweighted} = 0.011, p_{Weighted} = 0.011, p_{Bray-Curtis} = 0.0001$ ；兩年： $p_{Unweighted} = 0.018, p_{Weighted} = 0.032, p_{Bray-Curtis} = 0.0002$ ；三年： $p_{Unweighted} = 0.003, p_{Weighted} = 0.021, p_{Bray-Curtis} = 0.0001$ ，置換次數皆為 9999 次）。

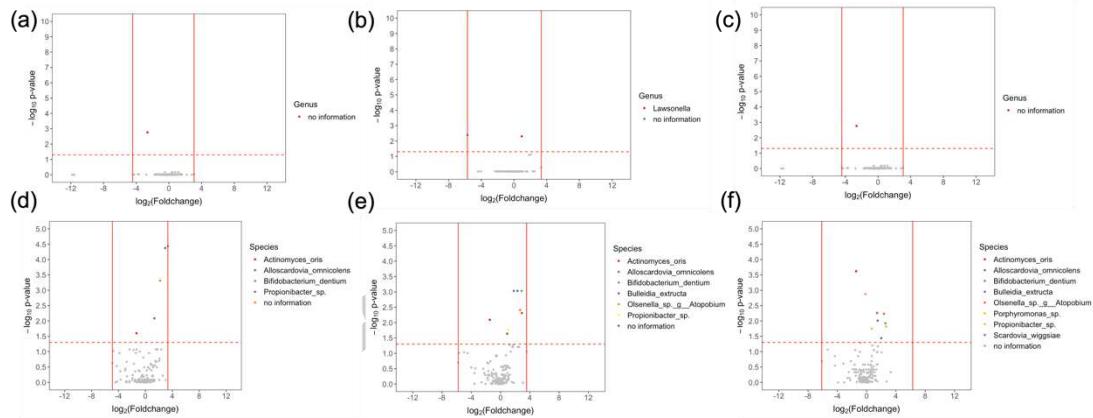


圖（三）、是否在（a）一年內復發（b）兩年內復發（c）三年內復發之兩組患者 Beta 多樣性指標差異

### 3.3 相對豐度差異分析結果

首先，如圖（四）所示，在菌屬層級中有顯著差異的菌屬為 *Lawsonella*，這個菌屬在有復發組中顯著增加。在菌種層級中，相對豐度於三種復發情況中

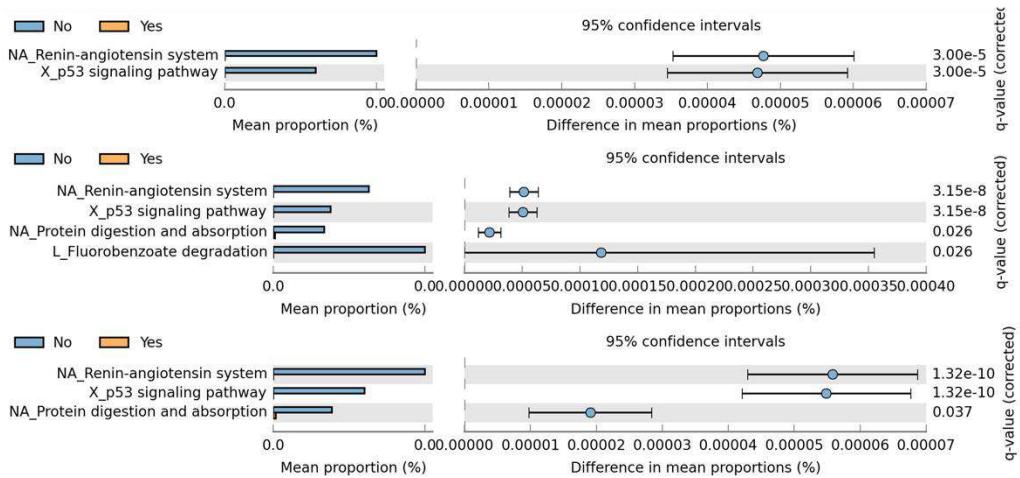
有重複出現的菌種共有五個，包括 *Actinomyces oris*、*Alloscardovia omnicolens*、*Bifidobacterium dentium* 和 *Propionibacter sp.*。此外，在是否於兩年內復發及是否於三年內復發的情況中皆出現顯著的菌種有 *Bulleidia extracta* 和 *Olsenella sp.* 及 *Atoponiu*m。



圖（四）、不同復發時間段內兩組受試者相對豐度差異菌屬層級：(a) 一年內復發 (b) 兩年內復發 (c) 三年內復發 菌種層級：(d) 一年內復發 (e) 兩年內復發 (f) 三年內復發

### 3.4 功能路徑豐度差異分析結果

根據 KEGG 資料庫的 2 級分類，在 148 條功能路徑中，在三種不同復發情況的兩組受試者皆有 2 條相同的功能路徑在無復發組中顯著增加，分別是 Renin-angiotensin system 和 p53 signalling pathway，其中涉及了內分泌系統（Endocrine system）及細胞成長與凋亡（Cell growth and death）。此外，另有 2 條功能路徑在無復發組中顯著增加，為 Protein digestion and absorption 和 Fluorobenzoate degradation，涉及了消化系統（Digestive system）及外源生物降解和代謝（Xenobiotics biodegradation and metabolism）。



圖（五）、不同復發時間段內兩組受試者之微生物組功能差異（上：一年內復發，中：兩年內復發，下：三年內復發）

### 3.5 XGBoost 模型預測結果

模型超參數設置如表（三）。

表（三）、XGBoost 模型參數設置

<b>Recurrent</b>	<b>eta</b>	<b>max_depth</b>	<b>subsample</b>	<b>colsample_bytree</b>	<b>nrounds</b>
Model1	0.1	7	1	1	500
Model2	0.3	5	0.8	0.8	200
Model3	0.1	3	0.8	1	100
<b>Recurrent within 2 years</b>	<b>eta</b>	<b>max_depth</b>	<b>subsample</b>	<b>colsample_bytree</b>	<b>nrounds</b>
Model1	0.3	7	1	1	100
Model2	0.1	7	0.8	1	200
Model3	0.3	7	0.8	0.8	200
<b>Recurrent within 3 years</b>	<b>eta</b>	<b>max_depth</b>	<b>subsample</b>	<b>colsample_bytree</b>	<b>nrounds</b>
Model1	0.1	7	0.8	1	100
Model2	0.3	7	0.8	0.8	100
Model3	0.3	5	0.8	1	100

XGBoost 模型的預測結果，表（四）、表（五）及圖（六）顯示，不同模型在口腔癌復發的預測效能上存在顯著差異。在整體復發的預測中，模型一的 AUC 為 0.5332 (95% CI: 0.3014–0.7561)，準確率為 0.4865，敏感性與特異性分別為 0.5455 與 0.4615。相比之下，僅使用口腔菌資料的模型二顯著提高了

AUC 至 0.6048 (95% CI: 0.4120–0.8327) ，準確率也提升至 0.7297，其特異性達到 0.8462，顯示了口腔微生物資料在預測復發中的重要作用。然而，模型二的敏感性僅為 0.4545，略低於模型一。模型三結合了臨床參數與口腔菌特徵，AUC 進一步提高至 0.6294 (95% CI: 0.4412–0.8175) ，準確率為 0.6486，特異性保持在 0.8077 的高水準，但敏感性下降至 0.2727。

在 2 年內復發的預測中，模型一的 AUC 為 0.4810 (95% CI: 0.1301–0.8318) ，準確率為 0.5676，敏感性與特異性分別為 0.5714 與 0.5667。模型二在 AUC 上達到 0.6048 (95% CI: 0.3445–0.8649) ，準確率顯著提升至 0.7568，特異性達到 0.9000，顯示了微生物資料在區分復發與非復發患者中的優勢，但其敏感性降至 0.1429，表明在檢測復發患者方面存在局限性。模型三的 AUC 最高，為 0.6571 (95% CI: 0.4253–0.8890) ，準確率為 0.7297，特異性為 0.8667，然而敏感性仍偏低，僅為 0.1429。

在 3 年內復發的預測中，模型一的 AUC 為 0.6643 (95% CI: 0.4092–0.8796) ，準確率為 0.4595，敏感性達到 0.8750，但特異性僅為 0.3448。模型二在 AUC 方面提高至 0.6979 (95% CI: 0.4412–0.8175) ，準確率也提升至 0.7838，特異性為 0.8485，顯示微生物資料能顯著提升準確率，但敏感性降低至 0.2500。模型三在 3 年內復發的預測中表現最佳，其 AUC 高達 0.8495 (95% CI: 0.7236–0.9753) ，準確率為 0.8108，特異性達到 0.8710，同時敏感性也提升至 0.5000。

整體而言，研究結果顯示，模型二在特異性方面優於模型一，但敏感性不足，而模型三透過整合臨床與微生物資料，在準確率、AUC 和敏感性上均有提升，特別是在 3 年內復發的預測中展現了顯著優勢。

表（四）、XGBoost 預測結果指標

復發	AUC	準確率	敏感性	特異性
模型一	0.5332 (0.3014-0.7561)	0.4865 (0.3192-0.656)	0.5455	0.4615
模型二	0.6224 (0.4120-0.8327)	0.7297 (0.5588-0.8621)	0.4545	0.8462

模型三	0.6294 (0.4412-0.8175)	0.6486 (0.4746-0.7979)	0.2727	0.8077
<b>兩年內復發</b>	<b>AUC</b>	<b>準確率</b>	<b>敏感性</b>	<b>特異性</b>
模型一	0.4810 (0.1301-0.8318)	0.5676 (0.3949-0.729)	0.5714	0.5667
模型二	0.6048 (0.3445-0.8649)	0.7568 (0.588-0.8823)	0.14286	0.90000
模型三	0.6571 (0.4253-0.8890)	0.7297 (0.5588-0.8621)	0.14286	0.86667
<b>三年內復發</b>	<b>AUC</b>	<b>準確率</b>	<b>敏感性</b>	<b>特異性</b>
模型一	0.6643 (0.4092-0.8796)	0.4595 (0.2949-0.6308)	0.8750	0.3448
模型二	0.6970 (0.4412-0.8175)	0.7838 (0.6179-0.9017)	0.2500	0.8485
模型三	0.8495 (0.7236-0.9753)	0.8108 (0.6484-0.9204)	0.5000	0.8710

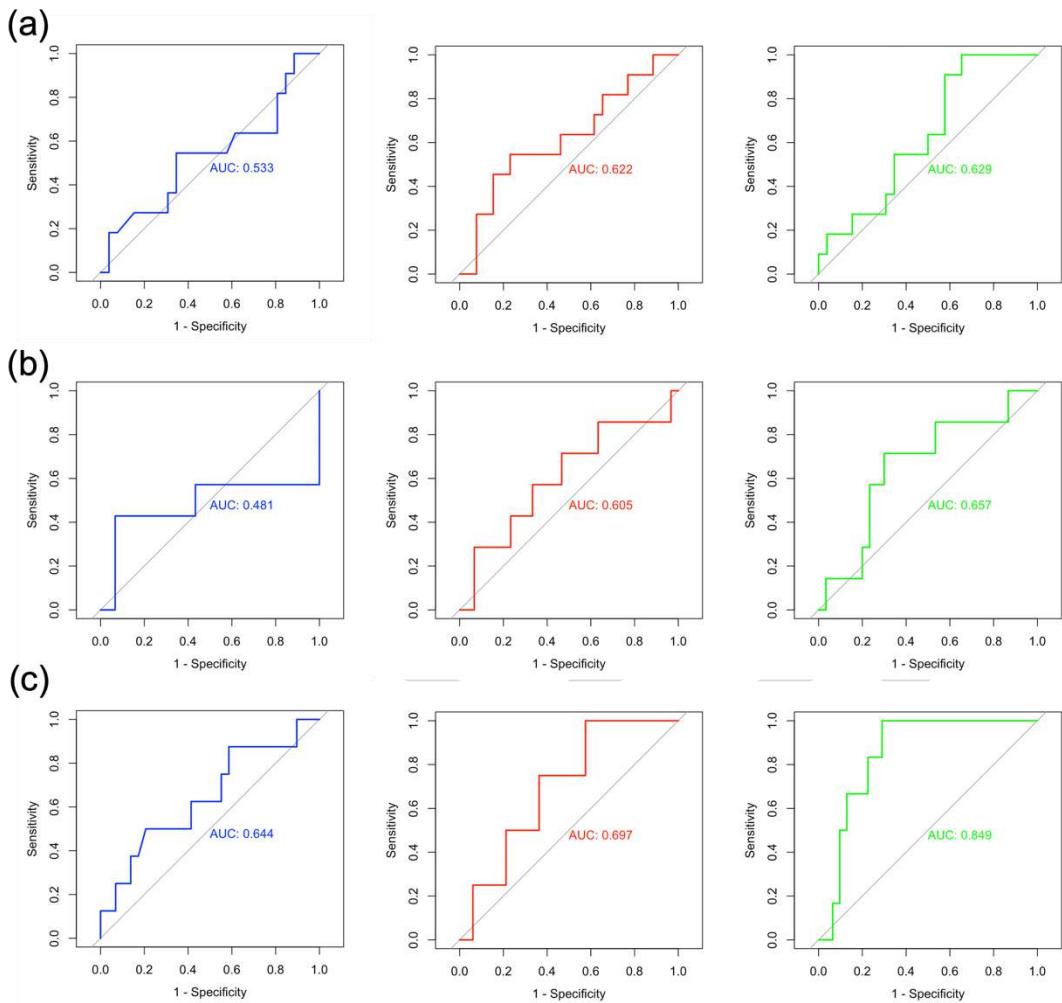
表（五）、模型三之混淆矩陣（Confusion Matrix）

左：是否復發 中：是否於兩年內復發 右：是否於三年內復發

		Actual	
		+	-
Predict	+	3	5
	-	8	21

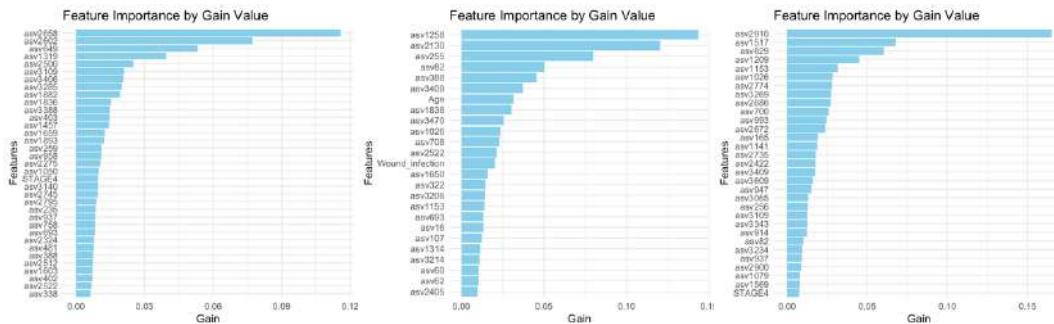
		Actual	
		+	-
Predict	+	1	4
	-	6	26

		Actual	
		+	-
Predict	+	3	4
	-	3	27



圖（六）、OSCC 復發預測模型表現 (a) 是否復發 (b) 是否在兩年內復發 (c) 是否在三年內復發，由左至右分別為模型一、模型二及模型三

除此之外，在本研究中針對三種情況的模型三計算了 cumulative gain，並將累積值達到 0.8 的 ASVs 篩選出來。接著，針對不同模型中篩選出的 ASVs，找出其中的重複項目，作為潛在的生物標誌物。此方法有助於提高生物標誌物的穩定性與可靠性，進一步反映不同模型間的一致性。



圖（七）、OSCC 復發預測模型三特徵重要性，由左至右分別為是否復發、是否在兩年內  
復發及是否在三年內復發

在分析中，透過篩選各模型中累積增益值達到 0.8 的 ASV，如圖（七）所示，發現有多個 ASV 在不同模型間重複出現，包括 asv82、asv388、asv693、asv937、asv1659、asv2522 和 asv3409。這些重複的 ASV 可能在口腔癌復發的預測上具有重要意義，為潛在的生物標誌物提供了研究依據，進一步研究可能揭示其與口腔癌復發的相關性。

## 第四章、五年存活狀態

### 4.1 人口統計特徵分析結果

首先，我們比較了五年後是否存活的兩組患者（存活組 N=114，死亡組 N=58）的人口統計特徵，其結果統整於表（六）中。結果顯示，在年齡、性別兩變數上，兩組間無顯著差異。然而，有幾個變數在兩組間顯示出顯著差異。總臨床分期在存活組和死亡組之間存在顯著差異 ( $p\text{-value}<0.001$ )，存活組的早期（1~3 期）比例顯著高於死亡組，而晚期（4 期）比例相對較低。同樣地，是否於五年內復發在兩組間也存在顯著差異 ( $p\text{-value}<0.001$ )，存活組中未復發的比例顯著高於死亡組。健康狀況方面，傷口感染 ( $p\text{-value}<0.001$ ) 及術後是否引發吸入性肺炎 ( $p\text{-value}=0.009$ ) 在兩組間也有顯著差異，存活組在這些不良術後狀況的比例顯著低於死亡組。根據以上結果，本研究在後續分析中校正了年齡、性別和總臨床分期三項干擾因子，同時將是否復發、傷口感染及術後吸入性肺炎視為中介因子進行調整。存活組早期分期比例較高，校正該變數有助於排除其對預後的影響，復發則是存活的重要預測因素，校正五年內的復發情況可有效排除其對五年後存活結果的影響。傷口感染和術後吸入性肺炎則是術後恢復與長期存活的重要影響因素，校正這些變數能減少術後併發症對結果的影響。這些干擾因子的校正使我們能更準確地探討口腔微生物與五年後存活狀態之間直接的因果關係。

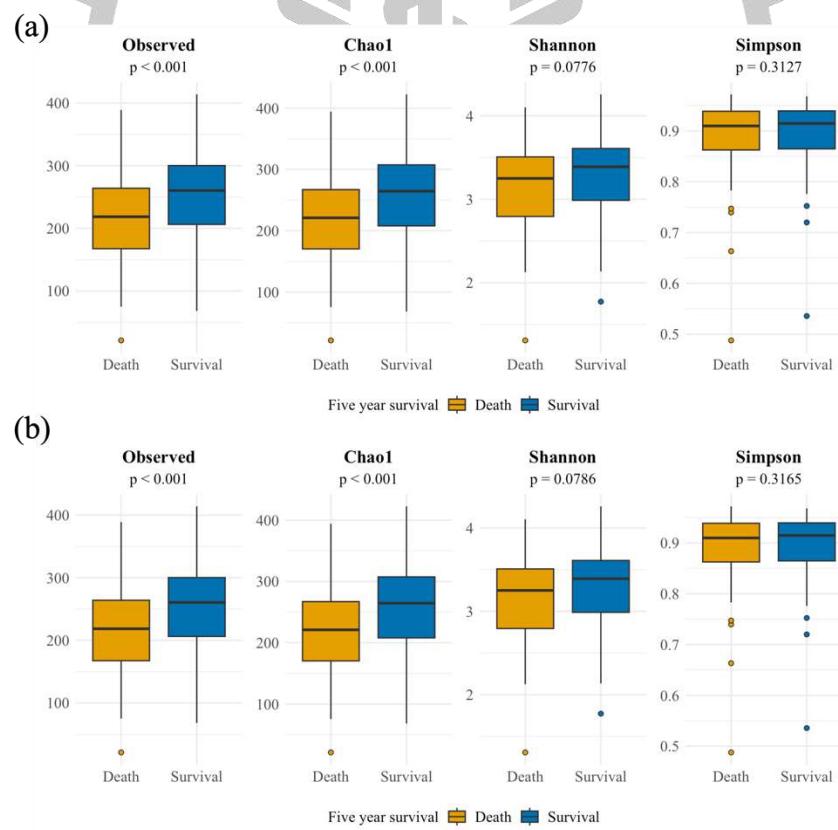
表（六）、五年後是否存活兩組患者人口統計特徵之比較

	死亡 (N=58)	存活 (N=114)	p-value
<b>年齡</b>			
平均值（標準差）	52.69 (8.83)	53.07 (9.96)	0.803
<b>性別</b>			
女性	6 (10.3%)	8 (7.0%)	0.646
男性	52 (89.7%)	106 (93.0%)	
<b>總臨床分期</b>			
I, II, III	20 (34.5%)	76 (66.7%)	38 <0.001
IV	38 (65.5%)	(33.3%)	
<b>是否於五年內復發</b>			
否	20 (34.5%)	104 (91.2%)	<0.001
是	38 (65.5%)	10 (8.8%)	
<b>加護病房住院時長</b>			

平均值 (標準差)	7.09 (3.42)	5.38 (3.25)	0.002
<b>有無傷口感染</b>			
無	22 (37.9%)	78 (68.4%)	<0.001
有	35 (62.1%)	36 (31.6%)	
<b>術後有無引發吸入性肺炎</b>			
無	46 (79.3%)	107 (93.9%)	0.009
有	12 (20.7%)	7 (6.1%)	

## 4.2 Alpha 多樣性及 Beta 多樣性分析結果

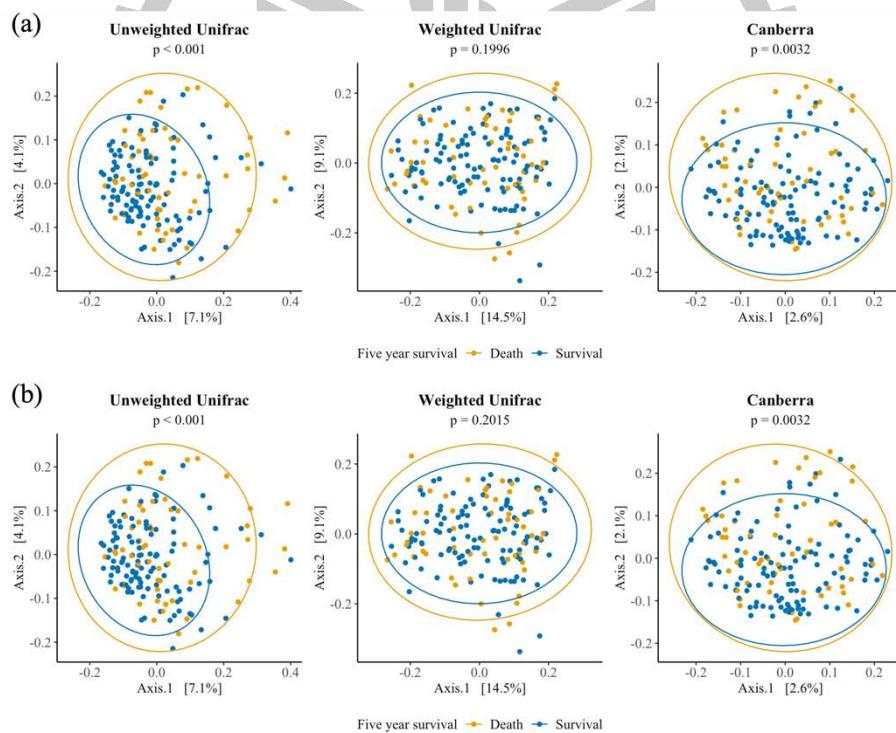
根據 Alpha 多樣性分析結果顯示，無論是否進行校正，Observed 和 Chao1 指標均顯示顯著差異，Shannon 和 Simpson 指標則未達顯著差異，四個指標均顯示五年後存活組的 Alpha 多樣性高於五年後死亡組（校正干擾因子： $p_{observed} < 0.001, p_{chao1} < 0.001, p_{shannon} = 0.0776, p_{simpson} = 0.3127$ ；校正干擾因子與中介因子： $p_{observed} < 0.001, p_{chao1} < 0.001, p_{shannon} = 0.0786, p_{simpson} = 0.3165$ ）。



圖（八）、五年後是否存活兩組患者 Alpha 多樣性指標差異 (a) 校正干擾因子 (b) 校正干擾因子與中介因子

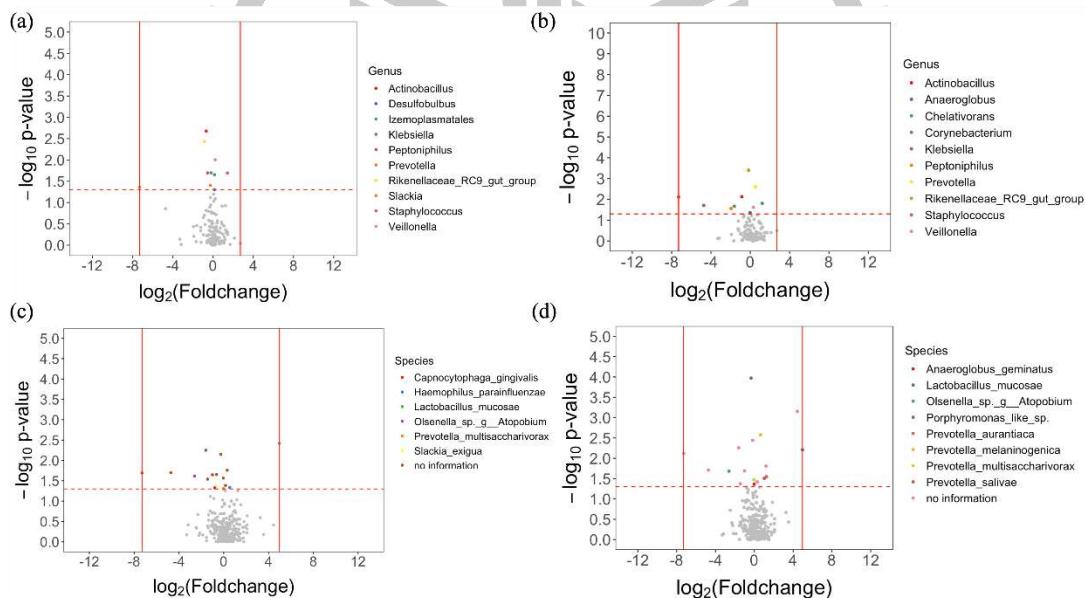
#### 4.3 相對豐度差異分析結果

透過 Beta 多樣性分析方法，可得到圖（九）之結果：我們發現無論是否進行校正基於未加權 UniFrac 與坎貝拉距離的計算上，五年後存活組與五年後死亡組兩群微生物組成有顯著的差異（校正干擾因子： $p_{Unweighted} < 0.001, p_{Canberra} = 0.0032$ ；校正干擾因子與中介因子： $p_{Unweighted} < 0.001, p_{Canberra} = 0.0032$ ，置換次數皆為 9999 次），而基於加權 UniFrac 距離計算則未有顯著差異（校正干擾因子： $p_{Weighted} = 0.1996$ ；校正干擾因子與中介因子： $p_{Weighted} = 0.2015$ ，置換次數皆為 9999 次）。



圖（九）、五年後是否存活兩組患者 Beta 多樣性指標差異 (a) 校正干擾因子 (b) 校正干擾因子與中介因子

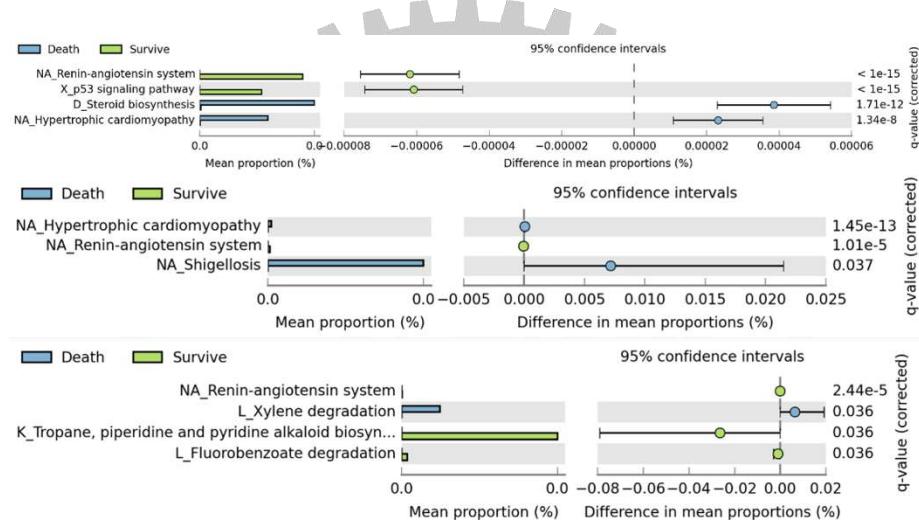
首先，在菌屬層級中，如圖（十）(a)(b) 所示，在僅校正干擾因子和校正干擾因子及中介因子共有的菌屬中，*Actinobacillus*、*Klebsiella*、*Peptoniphilus*、*Prevotella* 和 *Veillonella* 在五年後死亡的組別均顯著增加，而 *Rikenellaceae\_RC9\_gut\_group* 和 *Staphylococcus* 則顯著減少。在額外校正中介因子（併發症和復發）後的圖（十）(b) 中，新增且顯著的菌屬包括 *Anaeroglobus*、*Chelativorans* 和 *Corynebacterium*。在菌種層級中，如圖（十）(c)(d) 所示，在僅校正干擾因子和校正干擾因子及中介因子的模型中共有的菌種有 3 個，分別是 *Lactobacillus mucosae*、*Olsenella\_sp.\_g.\_Atopobium* 和 *Prevotella\_multisaccharivorax*，這些菌種皆在五年後死亡的組別中顯著減少。在額外校正中介因子（併發症和復發）後的圖（十）(d) 中，新增且顯著的菌種包括 *Anaeroglobus geminatus*、*Lactobacillus mucosae*、*Porphyromonas like sp.*、*Prevotella aurantiaca*、*Prevotella melaninogenica*、*Prevotella multisaccharivorax* 和 *Prevotella saliva*。



圖（十）、五年後存活狀態之兩組受試者之相對豐度差異 菌屬層級：(a) 僅校正干擾因子 (b) 校正干擾因子及中介因子 菌種層級：(c) 僅校正干擾因子 (d) 校正干擾因子及中介因子

#### 4.4 功能路徑豐度差異分析結果

根據 KEGG 資料庫的 2 級分類在 148 條功能路徑中，五年後存活狀態不同的兩組受試者有 5 條存在顯著差異，其中涉及了脂質代謝（Lipid metabolism）、內分泌系統（Endocrine system、細胞成長與凋亡（Cell growth and death）、心血管疾病（Cardiovascular disease）及轉錄（Transcription）。根據人口統計特徵結果可知，五年後存活狀態不同的兩組人之復發情況有顯著不同，因此針對有無復發進行分層。其中不論是否復發，腎素-血管收縮素系統（Renin-angiotensin system）的功能路徑皆出現了顯著的差異，其為涉及內分泌系統之功能路徑。



圖（十一）、（上）無分群（中）無復發人群（下）復發人群五年後是否存活兩組受試者之功能路徑差異

## 第五章、討論

### 5.1 口腔微生物組與復發之關係

#### 5.1.1 口腔微生物組之結果探討

研究結果表明，在 Alpha 多樣性的部分，在一年內是否復發、兩年內是否復發和三年內是否復發的情況中，兩個組別之 Observed 與 Chao1 指標都有顯著不同，且復發的組別之指標皆顯著較低，表示復發後之患者有較低的物種豐富度；而 Beta 多樣性也顯示出兩個組別的菌叢組成有顯著的差異，在一年內是否復發、兩年內是否復發和三年內是否復發的情況中，在未加權 UniFrac、加權 UniFrac 和 Bray-Curtis 距離，兩組患者皆有顯著的差異。為了確保分析結果能夠真實反映口腔微生物與 OSCC 復發的因果關係，本研究在 Alpha 與 Beta 多樣性分析中納入年齡、性別及臨床分期這三項變數進行控制。結果顯示，在考量這些可能影響復發風險的臨床因素後，復發組仍表現出顯著較低的 Alpha 多樣性指標，且 Beta 多樣性分析也顯示兩組間菌群組成存在顯著差異。此外，這些發現排除了年長患者或晚期 OSCC 患者本身較高復發風險的影響，進一步強化了口腔微生物與復發之間的因果關係。

在相對豐度分析中，我們同樣控制了性別、年齡和臨床分期對結果的影響。在菌屬層級上有顯著差異的菌屬為 *Lawsonella*，這個菌屬在有復發組別中顯著增加，表示這個菌屬可能與 OSCC 患者的復發有相關聯，這些菌屬在過去和 OSCC 復發的相關研究中並未被發現過，顯示本研究提供了新的數據，有助於進一步理解 OSCC 復發與口腔微生物之間的潛在關係。而在菌種層級中，相對豐度於三種復發情況中皆有顯著差異的菌種共有五個：*Actinomyces oris* 在未復發組別中顯著增加，表示此菌種可能有助於維持較好的口腔微生物平衡，而其餘四個菌種，*Alloscardovia omnicolens*、*Bifidobacterium dentium* 和 *Propionibacter sp* 則在復發組別中顯著增加，表示這些菌種可能與口腔癌的復發有相關。

此外，根據我們的分析，*Bulleidia extracta* 和 *Olsenella sp. g\_Atropionum* 這兩個菌種並未在一年內復發中顯示出顯著的差異，而是直到兩年內復發及三年內復發時才出現顯著增加的情況。這一結果強調了細分復發時間點的重要性，並展示了這些菌種在復發過程中的潛在長期作用。這表明，這些菌種的影響可能與晚期復發更為相關，而在早期復發階段，

可能是其他口腔微生物或因素在主導癌症的進展。隨著時間的推移，特定的菌種，*Bulleidia extracta* 和 *Olsenella sp. g\_Atoponium*，可能開始逐漸發揮其影響，並在較長的復發時間點上出現顯著的差異。這種時間上的差異有助於我們更深入地理解不同菌種在復發過程中發揮的作用，並顯示出僅將復發視為單一事件可能會忽視不同復發時間點的差異。因此，通過對復發時間點的細分分析，我們能夠更好地辨識和理解這些菌種的變化，並進一步釐清它們與口腔癌復發的因果關係。

### 5.1.2 功能路徑之結果探討

在功能路徑差異分析的結果中可發現，在三種不同復發時情況的兩組受試者皆有 2 條相同的功能路徑存在顯著差異，皆是在無復發組中顯著增加，其中涉及了內分泌系統（Endocrine system）及細胞成長與凋亡（Cell growth and death）。而在是否在兩年及三年內復發的情況中，有另外 2 條功能路徑存在顯著差異，其中涉及了消化系統（Digestive system）和外源生物降解和代謝（Xenobiotics biodegradation and metabolism）。

### 5.1.3 XGBoost 模型預測之結果探討

本研究透過口腔微生物相對豐度資料，應用 XGBoost 模型進行口腔癌復發之預測，結果展現了口腔微生物在風險評估上的潛在價值。根據研究結果，單純使用臨床參數的基本模型雖可提供基本參考，卻無法穩定捕捉復雜的病程因素。反觀納入口腔微生物資訊的模型，其特異性與整體準確率均獲得顯著提升，尤其在三年內復發的預測表現最為突出，AUC 高達 0.8495，準確率與敏感度也維持在相對均衡的水平。

在不同時間區段的分析中，口腔微生物結合臨床參數的混合模型普遍能提升預測效能，顯示菌相組成所提供的生物學訊號對於長期預後的評估具有補充作用。這與已有文獻中所提及的微生物與癌症病程之交互影響相呼應。本研究同時採用 SMOTE 方法平衡少數類別資料，並透過網格搜索與交叉驗證優化超參數，使得在復發樣本數量有限的情況下，模型仍能展現出較佳的敏感度與特異性表現。若未來能持續擴充資料量，預期可進一步強化模型的泛化能力與穩定度。

本研究證實口腔微生物資料對於口腔癌復發風險具有重要的早期檢測作用，與臨床參數相互搭配更可提升對高風險個案的辨識度。未來若能持續擴大研究樣本，並在臨床端結合多次追蹤與更豐富的臨床資訊，預期可再度提升預測模型的準確度與實用性，為臨床端提供更具前瞻性的輔助診斷工具，協助醫師及早安排個人化的治療與追蹤策略，最終提升口腔癌患者的整體預後表現。

## 5.2 口腔微生物組與五年後存活狀態之關係

### 5.2.1 口腔微生物組之結果探討

研究結果表明，在 Alpha 多樣性的部分，五年後是否存活的兩個組別之 Observed 與 Chao1 指標都有顯著不同，且五年後死亡組之指標皆顯著較低，表示五年後為死亡狀態之患者有較低的物種豐富度；而 Beta 多樣性也顯示出兩個組別的菌叢組成有顯著的差異，在未加權 UniFrac 和坎貝拉距離，兩組患者皆有顯著的差異。進一步去分析相對豐度的差異，在菌屬層級上，不論是否額外校正中介因子，在五年後死亡的組別中皆出現顯著增加的菌屬有 *Actinobacillus*、*Klebsiella*、*Peptoniphilus*、*Prevotella* 和 *Veillonella*，表示這五個菌屬可能與 OSCC 患者的預後不佳情況有相關聯。根據過往研究顯示，口腔中的 *Klebsiella* 通常含量較低，但在出現牙周病等特殊情況下可能會顯著增加[35]。顯著減少的菌屬則有 *Rikenellaceae\_RC9\_gut\_group* 和 *Staphylococcus*，表示這兩個菌屬可能與 OSCC 患者的較為良好的預後情況有相關聯。在額外校正中介因子（併發症和復發）後出現三個新增的顯著菌屬 *Anaeroglobus*、*Chelativorans* 和 *Corynebacterium*，可能代表口腔微生物在不受這些干擾因子和中介因子影響的情況下，直接與患者的五年存活狀態相關。這些新增的菌屬暗示了，除了復發和併發症這些常見的風險因子外，還可能存在其他途徑或機制參與影響五年存活狀態。在菌種層級中，在五年後死亡的組別中顯著增加的菌種有 *Olsenella sp. g Atopobium*、*Porphyromonas like sp.* 及 *Prevotella multisaccharivorax*，表示這三個菌種可能與 OSCC 患者的預後不佳情況有相關聯。顯著減少的菌種則有 *Anaeroglobus\_geminatus*、*Lactobacillus\_mucosae*、*Prevotella aurantiaca*、*Prevotella melaninogenica*、*Prevotella\_multisaccharivorax* 及 *Prevotella\_salivae*，表示這些菌種可能與

OSCC 患者的較為良好的預後情況有相關聯。此結果與先前在菌屬層級上的發現相吻合，特別是 *Prevotella* 屬中的多個菌種在五年後存活的組別中呈現顯著增加。除此之外，在額外校正中介因子（併發症和復發）後新增且顯著的菌種包括 *Anaeroglobus geminatus*、*Lactobacillus mucosae*、*Porphyromonas like sp.*、*Prevotella aurantiaca*、*Prevotella melaninogenica*、*Prevotella multisaccharivorax* 和 *Prevotella salivae*。這些菌種在校正中介因子後才顯示出顯著的豐度變化，意味著它們可能在不受併發症和復發影響的情況下，直接或間接與患者的長期預後有關聯。

### 5.2.2 功能路徑之結果探討

在功能路徑差異分析的結果中可發現，五年後存活狀態不同的兩組受試者有 5 條存在顯著差異，其中涉及了脂質代謝 (Lipid metabolism)、內分泌系統 (Endocrine system)、細胞成長與凋亡 (Cell growth and death)、心血管疾病 (Cardiovascular disease) 及轉錄 (Transcription)。

### 5.3 研究限制

本研究在資料與方法上仍有三項研究限制值得進一步探討。首先，口腔微生物組成可能受個人生活型態、飲食習慣與外在環境等多重因素影響，若能納入更完整的臨床與生活背景資訊，並於追蹤期間進行多次重複採檢，將有助於掌握微生物組隨時間的變動情形，進一步提升分析準確性。其次，儘管本研究的復發樣本量相較部分早期文獻有所提升，仍難以同時兼顧預測模型之敏感度與特異性，特別是在復發樣本比例偏低的情境下，模型效能可能有所折衷。此外，由於部分患者可能因地理位置、治療選擇或其他個人因素轉至其他醫療機構，我們無法完整追蹤其後續病程，導致部分資料遺漏，亦可能對結果的代表性和推論力產生一定影響。

## 5.4 未來展望

本研究結合臨床變項與口腔微生物組資料，分別探討了口腔微生物組與 OSCC 復發及五年存活狀態之間的因果關係，並透過機器學習模型進行復發風險預測。研究結果顯示，微生物多樣性、組成與功能在不同預後狀態間具有顯著差異，機器學習模型亦展現良好之預測效能。整體而言，雖然本研究樣本數有限，部分分析可能受到統計力不足及潛在干擾因子影響，但結果仍提供了口腔微生物組參與 OSCC 病程的重要假說。未來若能擴充樣本量，並整合更多臨床與生活型態資料進行建模與驗證，將有助於更深入了解口腔微生物與口腔癌復發或長期生存狀態之間的關聯，並強化其於臨床風險評估與預後管理的應用潛力。



## 參考文獻

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA: A Cancer Journal for Clinicians, 2021. **71**(3): p. 209-249.
2. Welfare, M.o.H.a., *The 2023 Statistics on the Causes of Death in Taiwan*. 2024.
3. Argiris, A., et al., *Head and neck cancer*. The Lancet, 2008. **371**(9625): p. 1695-1709.
4. Joosten, M.H.M.A., R. de Bree, and E.M. Van Cann, *Management of the clinically node negative neck in squamous cell carcinoma of the maxilla*. Oral Oncology, 2017. **66**: p. 87-92.
5. Haque, S., et al., *High-risk pathological features at the time of salvage surgery predict poor survival after definitive therapy in patients with head and neck squamous cell carcinoma*. Oral Oncology, 2019. **88**: p. 9-15.
6. González-García, R., et al., *Local recurrences and second primary tumors from squamous cell carcinoma of the oral cavity: A retrospective analytic study of 500 patients*. Head & Neck, 2009. **31**(9): p. 1168-1180.
7. Mücke, T., et al., *Recurrence interval affects survival after local relapse of oral cancer*. Oral Oncology, 2009. **45**(8): p. 687-691.
8. Safi, A.-F., et al., *Analysis of clinicopathological risk factors for locoregional recurrence of oral squamous cell carcinoma – Retrospective analysis of 517 patients*. Journal of Cranio-Maxillofacial Surgery, 2017. **45**(10): p. 1749-1753.
9. Shinn, J.R., et al., *Cumulative incidence of neck recurrence with increasing depth of invasion*. Oral Oncology, 2018. **87**: p. 36-42.
10. Yanamoto, S., et al., *Clinicopathological risk factors for local recurrence in oral squamous cell carcinoma*. International Journal of Oral and Maxillofacial Surgery, 2012. **41**(10): p. 1195-1200.
11. D'Cruz, A.K., et al., *Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer*. New England Journal of Medicine, 2015. **373**(6): p. 521-529.
12. Hashimoto, K., et al., *Feasibility of oral microbiome profiles associated with oral squamous cell carcinoma*. J Oral Microbiol, 2022. **14**(1): p. 2105574.
13. Cai, L., et al., *Integrative analysis reveals associations between oral microbiota dysbiosis and host genetic and epigenetic aberrations in oral cavity squamous cell carcinoma*. npj Biofilms and Microbiomes, 2024. **10**(1): p. 39.
14. Lyu, W.-N., et al., *An Oral Microbial Biomarker for Early Detection of Recurrence of Oral Squamous Cell Carcinoma*. ACS Infectious Diseases, 2023. **9**(9): p. 1783-1792.
15. Li, S., et al., *Prognostic impact of oral microbiome on survival of malignancies: a systematic review and meta-analysis*. Systematic Reviews, 2024. **13**(1): p. 41.
16. Mohamed, N., et al., *Analysis of Salivary Mycobiome in a Cohort of Oral Squamous Cell Carcinoma Patients From Sudan Identifies Higher Salivary Carriage of Malassezia as an Independent and Favorable Predictor of Overall Survival*. Front Cell Infect Microbiol, 2021. **11**: p. 673465.
17. Davaatsend, O., et al., *Factors influencing the 5-year survival rate of oral cancer patients in the Mongolian population: a retrospective cohort study*. Frontiers in Oral Health, 2023. **4**.

18. Abe, A., et al., *Prognostic impact of the prognostic nutritional index in cases of resected oral squamous cell carcinoma: a retrospective study*. BMC Oral Health, 2021. **21**(1): p. 40.
19. Magoč, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. **27**(21): p. 2957-63.
20. Amir, A., et al., *Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns*. mSystems, 2017. **2**(2): p. 10.1128/msystems.00191-16.
21. Chao, A., *Nonparametric Estimation of the Number of Classes in a Population*. Scandinavian Journal of Statistics, 1984. **11**(4): p. 265-270.
22. Shannon, C.E., *A mathematical theory of communication*. The Bell System Technical Journal, 1948. **27**(3): p. 379-423.
23. Simpson, E.H., *Measurement of Diversity*. Nature, 1949. **163**(4148): p. 688-688.
24. Lozupone, C. and R. Knight, *UniFrac: a new phylogenetic method for comparing microbial communities*. Appl Environ Microbiol, 2005. **71**(12): p. 8228-35.
25. McMurdie, P.J. and S. Holmes, *phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data*. PLOS ONE, 2013. **8**(4): p. e61217.
26. Bray, J.R. and J.T. Curtis, *An Ordination of the Upland Forest Communities of Southern Wisconsin*. Ecological Monographs, 1957. **27**(4): p. 325-349.
27. Oksanen, J., et al., *Vegan: Community Ecology Package*. R Package Version 2.2-1, 2015. **2**: p. 1-2.
28. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society: Series B (Methodological), 1995. **57**(1): p. 289-300.
29. White, J.R., N. Nagarajan, and M. Pop, *Statistical methods for detecting differentially abundant features in clinical metagenomic samples*. PLoS Comput Biol, 2009. **5**(4): p. e1000352.
30. Storey, J.D., *A direct approach to false discovery rates*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2002. **64**(3): p. 479-498.
31. Chen, T. and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 785–794.
32. Friedman, J.H., *Greedy function approximation: A gradient boosting machine*. The Annals of Statistics, 2001. **29**(5): p. 1189-1232, 44.
33. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. Radiology, 1982. **143**(1): p. 29-36.
34. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique*. J. Artif. Int. Res., 2002. **16**(1): p. 321-357.
35. Könönen, E., et al., *Prevotella species as oral residents and infectious agents with potential impact on systemic conditions*. J Oral Microbiol, 2022. **14**(1): p. 2079814.