

國立臺灣大學公共衛生學院流行病學與預防醫學研究所



碩士論文

Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Master Thesis

利用分類樹進行胰臟癌病患的存活分群

Utilizing Classification and Regression Tree for
Prognostic Grouping of Patients with Pancreatic Cancer

張家甄

Chia-Chen Chang

指導教授:盧子彬 博士

Advisor: Tzu-Pin Lu, Ph.D.

中華民國 109 年 7 月

July 2020



國立臺灣大學碩士學位論文
口試委員會審定書

論文中文題目 利用分類樹進行胰臟癌病患的存活分群

論文英文題目 Utilizing Classification and Regression Tree
for Prognostic Grouping of Patients with Pancreatic
Cancer

本論文係 張家甄 君(學號 r07849011)在國立臺灣大學
流行病學與預防醫學研究所完成之碩士學位論文，於民國
109 年 7 月 7 日承下列考試委員審查通過及口試及格，特此
證明。

口試委員：

張子杼

(簽名)

(指導教授)

李文宗

江漢如

田智文

施耀量



致謝

進入台大流預所的時光飛逝，轉眼兩年就過去了。首先最感謝的是盧子彬老師，不厭其煩地教導我，當我研究時遇到瓶頸，老師總能提點我去思考研究的下一步怎走，讓我受益良多，真的非常謝謝老師。也非常謝謝台大醫院外科的吳健暉醫師，在胰臟癌的專業上給予許多建議。同時也非常感謝在研究上幫助我許多的景榮學長、癌登的江濬如老師及明勳，每當我資料定義不明白時，他們都能讓我更瞭解癌登及資料的定義。

再來謝謝同研究室的學長姊、同學及學弟妹，在 meeting 中提供意見跟想法。特別感謝涵晴及亦文學姊，在我研究最忙碌時陪我吃飯聊聊天，可以放鬆心情，能夠更認識妳們真好。以及婉羽、凱元及滄煜，我們一起討論、參加會議及準備論文期間互相打氣。過了這兩年，很慶幸當初有進來台大流預所，所上的老師及同學都非常友善，也學到了許多專業領域的知識及應用，讓我順利且充實的度過了這兩年。

最後，謝謝我的男朋友日昇，在我碩二下這繁忙的日子中，陪我討論論文及練習口試，也非常感謝我的家人，從我求學的過程一路支持我，一步步都讓我自己做選擇，使我無後顧之憂地完成我的學業，謝謝你們。


張家甄 謹致

中文摘要



背景：不論在全球或台灣，胰臟癌皆是死亡率極高的癌症，組織型態上以腺癌 (Pancreatic ductal adenocarcinoma, PDAC) 為主要亞型，胰臟癌五年存活率僅約 6%，其主因為發現罹癌時多為晚期，且胰臟位於人體後腹膜腔，手術治療困難，目前僅兩成病患發現時可以接受手術治療，其餘的八成皆須仰賴化療。本研究欲使用臨床變數及組織型態研究胰臟癌病人存活之間的異質性，主要欲回答在美國癌症聯合委員會(AJCC)第八版中佔有最多數的第三期病人(84%)的存活情形是否具有組內差異，及胰臟癌病患存活是否可以不同組織型態進一步分類。由於精確的病患分類是增進醫療效果的重要步驟，本計畫預期針對此二目標進行研究，進而提升治療效果。

方法：本研究使用台灣癌症登記資料庫研究，研究納入該資料庫中長表自 2013 年 1 月 1 日至 2017 年 12 月 31 日期間被診斷胰臟癌的病患之期別資訊，以及納入年報檔自 1995 年 1 月 1 日至 2017 年 12 月 31 日之數據以探討組織型態之存活差異。西方驗證資料方面則使用 2000 年 1 月 1 日至 2016 年 12 月 31 日之美國癌症登記資料 (The Surveillance, Epidemiology, and End Results, SEER)，以胰臟癌的初診日為研究基準日，再探討臨床變項例如期別、腫瘤分級分化(grade)等之預測效益。在臨床變項之存活分析上，本研究使用對數秩檢驗檢定組別之差異及 Cox 比例風險模式評估計算其風險比值，而進一步將存活差異進行分類時則採用決策樹進行分群。



結果：胰臟癌有手術的病患存活狀況比沒手術的病患顯著改善(p 值小於 0.0001)，因此是否進行手術為胰臟癌病患存活最重要的分層因子，進而才會是腫瘤本身特性之影響。在接受手術的胰臟癌第三期病患中，本研究透過決策樹挑選出兩個重要性最高的變數，為淋巴結擴散程度及腫瘤分級分化。研究結果經調整干擾因子後，針對淋巴結擴散程度，N1 的死亡風險為 N0 的 1.52 倍(95%信賴區間為[1.18-1.94], p 值小於 0.0001)，N2 的死亡風險為 N0 的 2.30 倍(95%信賴區間為[1.71-3.08], p 值小於 0.0001)；在腫瘤分級分化下，分化中度的死亡風險為分化良好的 1.70 倍(95%信賴區間為[1.03-2.79], p 值為 0.004)，分化不良組的死亡風險為分化良好的 3.80 倍(95%信賴區間為[2.25-6.39], p 值小於 0.0001)。透過交互作用項進一步進行病患之分類，依照病人存活狀況，我們使用決策樹可分類為四個組別，預後較高組的平均存活時間為 23.5 個月，而中高組、中低組及最低組的中位存活時間分別為 18.4 個月、14.5 個月及 9 個月，四組間皆有顯著差異(p 值小於 0.0001)。胰臟癌病患的組織型態上，經由分組後可將病患之存活曲線可分為四個的存活區塊，第一組為最好的組織亞型，其四年存活率為 90%，而第二組、第三組及四組的中位存活時間分別為 41.0 個月、13.5 個月及 5.4 個月。而第二組、第三組及第四組的死亡風險分別為第一組的 5.52 倍、9.66 倍及 20.32 倍，對數秩檢定結果為四組有顯著差異(p 值小於 0.0001)，本分群方式驗證在美國癌症登記資料庫亦有相同的結果。

結論：在本研究中，淋巴結擴散程度及腫瘤分級分化皆為第三期胰臟癌病患之重要存活預測因子，未來可提供臨床上治療決策之輔助，同時由於此結果能適用美國的胰臟癌病患，顯示此結果具有外推性，不僅限於台灣病患。組織型態之結果

以固狀偽乳頭狀腫瘤預後狀況最佳，其餘組織型態則各有差異，代表在未來治療上應將組織型態納入治療考量。



關鍵字：胰臟癌，台灣癌症登記資料庫，美國癌症登記資料庫，組織型態，存活率

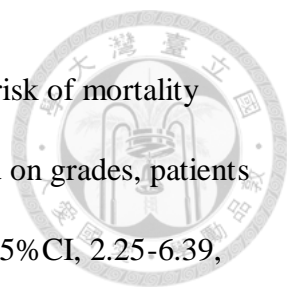
Abstract



Background: Pancreatic cancer is one of the most malignant cancer types with poor survival outcomes. Currently, physicians use the American Joint Committee on Cancer (AJCC) staging system to classify patients into different subgroups with distinct outcomes. However, the AJCC staging system classifies the majority of pancreatic cancer patients into stage III and dramatic heterogeneity is observed among the patients. The aim of this study is to provide precise classification systems by utilizing the clinical variables and histological subtypes of pancreatic cancer in order to help clinicians to make a better treatment plan.

Method: We retrieved the data from the Taiwan Cancer Registry (TCR) and the US-based Surveillance, Epidemiology, and End Results (SEER). Stage-specific and histological overall survival were analyzed by the Kaplan-Meier survival curves and log-rank tests. Cox proportional hazard regression model was applied to determine the predictive effects of clinical variables. Patients were further divided into subgroups with distinct survival outcomes by using the classification and regression trees (CART). All results were validated by the SEER data.

Results: Pancreatic cancer patients who underwent surgery showed significantly better survival outcomes than those patients without surgery, suggesting that whether a patient receives surgery is the most important factor. Subsequently, two clinical variables including lymph nodes involvement and cancer grade showed significant associations.



Regarding lymph nodes involvement, patients with N2 had a higher risk of mortality (HR=2.30, 95%CI, 1.71-3.08, P<.0001) than patients with N0. Based on grades, patients with poorly-differentiated had a higher risk of mortality (HR=3.80, 95%CI, 2.25-6.39, P<.0001) than patients with well-differentiated. In accordance with the survival outcomes of patients, we could divide them into different groups by interaction terms with CART algorithm. The median survival times of the four groups were 23.5, 18.4, 14.5, and 9.0 months for ‘High’, ‘Moderately High’, ‘Moderately Low’, and ‘Low’ group, respectively (P<.0001). Histological subtypes were analyzed by previously described pipelines. Eight histological subtypes of pancreatic cancer patients could be classified into four different groups with distinct survival outcomes from high (Group1) to low (Group4). The 4-year survival rates of Group1 was 90%. The median survival of Group2, Group3, and Group4 were 41.0, 13.5, and 5.4 months (P<.0001). Similar results were observed by using the SEER data.

Conclusion: Among the stage III pancreatic cancer patients, our results showed that lymph nodes involvement and cancer grades are important predicting factors for the survival outcomes. Further classification of pancreatic cancer patients based on their histological information is another important predictor of the survival outcomes, and the differences among these 4 groups are dramatically significant. In conclusion, taking the clinical and histological information into consideration may help physicians to make better therapeutic strategies.

Keywords: Pancreatic cancer, Histology, Taiwan Cancer Registry, SEER, Survival

目錄



口試委員審定書	II
致謝.....	III
中文摘要.....	IV
ABSTRACT	VII
第一章 導論.....	1
1.1 研究背景.....	1
1.2 研究動機.....	2
1.3 研究目的.....	3
第二章 材料與方法.....	3
2.1 樣本	3
2.1.1 資料來源.....	3
2.1.2 資料特性.....	4
2.1.3 樣本篩選.....	5
2.1.4 期別轉換方式.....	6
2.2 分析方法.....	7
2.2.1 對數秩檢驗(Log-rank test).....	7
2.2.2 Cox 比例風險模型(Cox proportional hazard regression model)	7
2.2.3 臨床變項分析.....	8
2.2.4 決策樹 (Decision Tree)	8
第三章 結果.....	10



3.1 資料篩選.....	10
3.2 臨床變數結果.....	10
3.3 探討影響胰臟癌病人的預後因子及分類.....	11
3.3.1 決策樹(CART)挑變數的結果.....	11
3.3.2 透過 Cox 風險比例模式找出重要的預後因子.....	11
3.3.3 腫瘤分級分化跟淋巴結擴散程度之關聯分析.....	12
3.3.4 探討兩變數存活曲線之加成性.....	12
3.3.5 使用分層分析變數的加成效果.....	13
3.4.6 透過決策樹進行預後因子的分群.....	13
3.4.7 輔助性化療在四組間的影響.....	14
3.4 外部資料驗證結果: 美國癌症登記資料.....	15
3.4.1 透過 Cox 風險比例模式.....	15
3.4.2 驗證台灣癌症登記分群結果.....	15
3.4.3 輔助性化療在四組間的影響.....	16
3.5 不同組織型態的結果.....	16
3.5.1 透過決策樹進行組織型態的分群.....	17
3.5.2 探討分群後的組內差異.....	17
3.5.3 外部驗證資料結果: 美國癌症登記資料.....	18
第四章 結論與討論.....	19
4.1 主要發現.....	19
4.2 研究討論.....	20
4.2.1 腫瘤分級分化在其他病理期別中的影響.....	20

4.2.2 組織型態的影響.....	20
4.2.3 研究限制.....	21
4.3 臨床與公共衛生的意義.....	21
參考資料.....	22



圖目錄



圖 1 胰臟癌台灣癌登資料庫長表篩選流程圖	25
圖 2 胰臟癌台灣癌登資料庫年報檔篩選流程圖	26
圖 3 胰臟癌美國癌症登記資料庫(2010-2015)篩選流程圖	27
圖 4 胰臟癌美國癌症登記資料庫(2000-2016)篩選流程圖	28
圖 5 台灣癌症登記資料庫胰臟癌(A)所有病人(B)手術變數的 Kaplan-Meier 分析圖	29
圖 6 台灣癌症登記資料庫(A)AJCC7 及(B)AJCC8 分期別的 Kaplan-Meier 分析圖	30
圖 7 台灣癌症登記資料庫之透過決策樹接受手術第三期之重要預後因子	31
圖 8 台灣癌症登記資料庫之第三期病人的兩變數交互作用的 Kaplan-Meier 分析圖	32
圖 9 台灣癌症登記資料庫(A)固定 N0 下(B)固定 N1 下(C)固定 N2 下的 Kaplan- Meier 分析圖.....	33
圖 10 台灣癌症登記資料庫(A)固定分化良好下(B)固定分化中度下(C)固定分化不 良下的 Kaplan-Meier 分析圖	34
圖 11 台灣癌症登記資料庫之使用兩變數交互作用分群結果	35
圖 12 台灣癌症登記資料庫四組的 Kaplan-Meier 分析圖.....	36
圖 13 台灣癌症登記資料庫預後(A)最高組(B)中高組(C)中低組(D)最差組切兩年化 療變數的 Kaplan-Meier 分析圖	37
圖 14 美國癌症登記資料庫第三期病人的(A)兩變數交互作用(B)使用台灣癌登分 群方法的 Kaplan-Meier 分析圖	38

圖 15 美國癌症登記資料庫預後(A)最高組(B)中高組(C)中低組(D)最差組切兩年化療變數的 Kaplan-Meier 分析圖	39
圖 16 台灣癌症登記資料庫組織型態亞型的 Kaplan-Meier 分析圖	40
圖 17 台灣癌症登記資料庫之八個組織亞型(A)決策樹分群結果(B) (B)使用台灣癌症分群方法的 Kaplan-Meier 分析圖	41
圖 18 台灣癌症登記資料庫分組後(A)Group 3 (B)Group 4 中組內差異的 Kaplan-Meier 分析圖	42
圖 19 美國癌症登記資料庫(A)組織型態(B)分四個組別的 Kaplan-Meier 分析圖 ..	43
圖 20 美國癌症登記資料庫分組後(A)Group 3(B)Group 4 中組內差異的 Kaplan-Meier 分析圖	44
圖 21 台灣癌症登記資料庫之腫瘤分級分化在第四期病人的 Kaplan-Meier 分析圖	45

表目錄



表 1 資料描述清單.....	46
表 2 AJCC7 和 AJCC8 之 TNM 的定義	47
表 3 AJCC7 和 AJCC8 版的制定期別定義.....	48
表 4 台灣癌症登記資料中病人之變數敘述統計	49
表 5 台灣癌症有手術第三期病人 cox regression 結果.....	50
表 6 腫瘤分級分化和淋巴結擴散程度數量表	51
表 7 分級和淋巴結擴散程度之交互作用對數秩檢定結果.....	52
表 8 分層控制在(A)N0 (B)N1 (C)N2 下 腫瘤分級分化之對數秩檢驗結果	53
表 9 分層控制在(A)分化良好(B)分化中度(C)分化不良下 淋巴結侵犯程度之對數 秩檢驗結果	54
表 10 美國癌症登記資料庫接受手術第三期病人 cox 回歸模型結果.....	55
表 11 台灣癌症登記年報檔的組織型態分佈	56



第一章 導論

1.1 研究背景

胰臟癌(pancreatic cancer)是現今社會相當關注的癌症，雖然在全球的發生率不高，但是為致死率很高的癌症，五年存活率約 6% [1]，過去研究顯示其五年存活率之範圍介於 2%~9% [2, 3]。即使是早期的患者，五年的存活率也僅有 30%~40%，且一旦經確診後，癌細胞的轉移機率將高達 50%，奪命速度相當迅速。根據 2015 年研究顯示，胰臟癌在全球已造成約 411,600 人死亡 [4]，且為美國第四大死因 [5]，在全球的統計，目前僅 20% 病患發現時可以接受手術治療，其餘的八成皆須仰賴化療 [6]。根據 2016 年臺灣癌症登記報告 [7]，當年次初次診斷為胰臟癌患者共 2,202 人，發生率在所有癌中排名中男性為第 13 位、女性為第 13 位，同年則有 1,996 人死於胰臟癌，死亡率的排名於男性為第 8 位、女性為第 5 位。胰臟不是前十大發生率高的癌症，但是卻在十大癌症死亡率的第八名。在 2016 年臺灣 2,202 位胰臟癌患者的首次療程上，不少病人求診時因腫瘤移轉已無法開刀切除，接受手術治療的人數為 529 人 (24%)，其餘 1673 人 (76%) 沒有接受手術治療。化療為最常見的輔助性治療療程，在所有病人中，接受化學治療的人數為 1,081 人 (49%)，其餘 1121 人沒有接受化療。胰臟癌依據腫瘤型態可分為外分泌腫瘤 (Exocrine tumor)，神經內分泌癌 (Neuro-Endocrine Tumor, NET) 兩種，其中外分泌腫瘤以胰腺癌 (Pancreatic ductal adenocarcinoma, PDAC) 為最多數，胰臟癌病患中約九成是常見的胰腺癌，胰臟癌被診斷通常有 25% 的人活過一年，而五年存



活率約只有 5%[8]，而神經內分泌癌的預後較好，五年存活率為 65%[9]，代表胰臟癌病患間的組織型態具有影響存活預後之重要能力。

胰臟癌目前最常使用的分期系統是由美國癌症聯合委員會(American Joint Committee on cancer, AJCC)與國際抗癌聯盟(Union for International Cancer Control, UICC)共同制定，AJCC-UICC 分期系統將胰臟癌根據 TNM 系統分為四期，「T」代表腫瘤的大小，「N」代表淋巴結擴散程度，「M」則代表遠端轉移[10]，目前此系統將胰臟癌病患主要分為四期，知道病患期別後，能判斷其預後狀況及治療計畫，而針對不同的分期的胰臟癌病人其治療策略亦有所不同。

1.2 研究動機

胰臟癌為極惡性的腫瘤，存活率非常低，而本研究欲探究胰臟癌的病人存活是否因某些臨床變數影響而具有異質性存在。目前利用臨床變數進行胰臟癌病患的分期指引為 AJCC，全球於 2018 年 1 月起開始使用 AJCC 第八版分期系統[11]，在胰腺癌的第八版分期裡，第三期病理期別為多數(84%)，在此期別中是否存在病患存活時間上的異質性值得探討，因此本研究預期藉由台灣癌症登記資料庫找出第三期病人之存活因子，並架構存活分類模型。此外胰臟癌本身的組織型態是否具有病患存活時間之差異亦值得後續探討，過往研究指出美國癌症登記資料中，神經內分泌癌的預後存活好過外分泌腫瘤，而外分泌腫瘤中，黏液性癌(mucinous carcinoma)的預後存活率較腺癌(adenocarcinoma)佳[12, 13]，其中以及胰管內乳突狀黏液性癌(Intraductal papillary-mucinous carcinoma)[14]預後最佳。而胰臟實質性假乳突狀癌(Solid pseudopapillary carcinoma)是一種罕見但低度惡性的腫



瘤，預後在胰臟癌中良好[13, 15]。因此，本研究第二目標為使用台灣癌症登記資料庫進行組織型態之存活差異研究，確認組織型態是否會影響胰臟癌預後的存活率。

1.3 研究目的


由臨床變項判定病人預後之後，臨床醫師將給予不同的治療選擇，所以精確的病患分類是增進醫療效果的重要步驟。因此，本研究針對第三期胰臟癌病患的臨床變項以及癌症的組織型態進行研究，期待能對胰臟癌病患進行更精確的分類，進而提升治療效果。此外除了找出重要因子之外，次要目的為確認台灣癌症登記處(東方資料)之分類結果和美國癌症登記處資料(西方資料)在本研究變數中是否存在種族的差異性。

第二章 材料與方法

2.1 樣本

2.1.1 資料來源

此研究的胰臟癌病患資料來自「台灣癌症登記中心(Taiwan cancer registry, TCR)」，在台灣癌症登記資料中[16, 17]，每位患者的資料均有性別、出生日期、戶籍地及診斷資料(診斷日期、腫瘤部位、腫瘤的組織病理學和腫瘤分級分



化)，其中包括有長表(long form, LF)使用 2013 至 2017 年的申報病患之有期別的資料及串連死因統計檔之病患的存活狀態及死因紀錄；年報檔使用 1995 至 2017 年的申報病患資料，而年報檔彙整時沒有申報期別的資訊。外部資料上，則使用 2000 至 2016 年之美國癌症登記資料[18](The Surveillance, Epidemiology, and End Results, SEER)的胰臟癌病患作為驗證，美國癌症登記資料庫目前收案的範圍為美國的 21 個地區，約涵蓋美國 30%的人口數。在第一個主題中，為了採用與台灣癌症登記資料庫相同期別版本的數據(AJCC 第七版)，使用 2010 至 2015 年的資料進行驗證，而在第二個主題組織型態中，在考量樣本數限制下，則使用 2000 至 2016 年的資料進行組織型態的驗證。本研究使用的資料集搜集年份、人數及最終使用人數如表 1 所示，台灣癌症登記資料庫的長表有 5,685 人，而年報檔中搜集的胰臟癌病患共有 33,364 人，美國癌症登記資料庫共有 67,916 人。

2.1.2 資料特性

在台灣癌症登記資料庫，本研究使用的連續變數為診斷年齡及腫瘤大小(mm)，而類別變數為性別、腫瘤分級分化、TNM 系統(包括腫瘤大小組別、淋巴結擴散程度、是否遠端轉移)及輔助性化療、放療進行臨床變數的探討，針對腫瘤分級被定義為四個等級，grade1 定義為分化良好(Well differentiated)，而 grade2、grade3 及 grade4 分別為分化中度(Moderately differentiated)、分化不良(Poorly differentiated)及未分化(Undifferentiated)。



2.1.3 樣本篩選

本研究所使用的兩筆資料中，台灣癌症登記資料長表為主要資料，年報檔因少了病患期別資訊，因此年報檔僅使用在探討組織型態之分析，而美國癌症登記資料庫為外部驗證資料，探討期別時，取和台灣癌症登記期別系統之同版本(AJCC 第七版)年份為 2010 年至 2015 年之分析，定義胰臟癌之外分泌腫瘤的代碼: (1)診斷部位為胰臟: C250-C259; (2) Histology ICD-O-3 codes 為外分泌腫瘤(Hist 不為 '8240' '8246' '8249' '8154' '8151' '8013' '8153' '8150' '8152' '8155' '8156' '8041')。

台灣癌症資料庫之第一個主題探討期別的篩選流程圖如圖 1，使用癌症登記長表進行分析，樣本納入條件為: 1)組織型態為外分泌腫瘤且排除預後良好的組織型態分別為胰臟實性假乳突狀癌[13]及胰管內乳突狀黏液性癌[14];2)非為原位癌之病患;3)為期別已知;4)超過 18 歲以上;5)無轉院之病患;6)如有接受過切除手術但未接受術前輔助性治療;7)追蹤連續期間必須超過 50 個月，若超過 50 個月之後才死亡之病患，其結果變項則視為存活。

第二個主題探討組織型態的篩選流程圖如圖 2，使用的樣本為年報檔，樣本納入條件為: 1)組織型態為外分泌腫瘤;2)非為原位癌之病患;3)超過 18 歲以上;4)追蹤連續期間必須超過 50 個月，若超過 50 個月之後才死亡之病患，其結果變項則視為存活。

美國癌症登記資料庫之第一個主題驗證資料篩選流程圖如圖 3，納入條件為: 1)和台灣癌症登記處相同之 AJCC 第七版的期別分類系統(2010 年至 2015 年);2)組織型態為外分泌腫瘤且排除預後良好的組織型態(胰臟實性假乳突狀癌及胰管內乳



突狀黏液性癌);3)非為原位癌之病患 4)超過 18 歲以上;5)死亡狀態已知;6)如有接受過切除手術但未接受術前輔助性治療;7)追蹤連續期間必須超過 50 個月，若超過 50 個月之後才死亡之病患，其結果變項則視為存活。

在驗證資料中，第二個主題的篩選流程圖如圖 4，樣本納入條件為: 1)組織型態為外分泌腫瘤;2)非為原位癌之病患;3)超過 18 歲以上;4)追蹤連續期間必須超過 50 個月，若超過 50 個月之後才死亡之病患，其結果變項則視為存活。

2.1.4 期別轉換方式

目前 AJCC 分期系統最新版本為第八版，而原始癌登資料的分期系統是 AJCC 第七版，因此本研究將期別依照第八版之 TNM 定義(表 2)進行轉換。首先，針對病理 T (Pathologic T)欄位，第八版的 T1 和第七版的 T1 為相同定義，要獲得第八版的 T2 和 T3 則需進行轉版，先將第七版的 T2 的人選出，再依照第八版的定義使用腫瘤大小(Tumor Size)欄位判斷為 T2(腫瘤大小大於 20mm 且腫瘤大小不超過 40mm)或 T3(腫瘤大小大於 40mm)。在第八版的 T4 中，則依照定義將原第七版的 T3 及 T4 都歸入此類別。對於病理 N (Pathologic N)的欄位轉換，第八版的 N0 和第七版的 N0 為相同的定義無須轉換，第八版 N1 和 N2 則需進行轉換，先將第七版的 N1 選出，再依照第八版的定義使用區域淋巴結侵犯數目 (Regional Lymph Nodes Positive)欄位判斷為 N1(淋巴侵犯數目介於 1 顆至 3 顆)及 N2(淋巴侵犯數目超過 4 顆)。而第八版病理 M(Pathologic M)和第七版的 M 是相同的定義，最後將轉換完成的第八版 TNM 依照表 3 進行期別轉換。



2.2 分析方法

首先使用對數秩檢驗(Log-rank test)及 Cox 比例風險模型(Cox proportional hazards regression model)探討胰臟癌腺癌病患期別及臨床變項與病患生存的相關性，視覺化圖形部分採用 Kaplan-Meier 曲線呈現，而中位存活時間(Median survival time)定義在 Kaplan-Meier 估計量到 50% 對應的時間，在研究病理期別第三期病人之預測因子時，使用決策樹(Decision Tree)找出影響存活之預測因子，而研究不同組織型態中，再對臨床變項進行分析，個別方法及說明如下：

2.2.1 對數秩檢驗(Log-rank test)


在進行存活分析時，能透過對數秩檢驗來比較組別之間的存活曲線是否有顯著差異[19]，此方法為單變量分析檢驗，其背後假設為一種無母數的檢定方法，本研究使用 R 軟體之 survival 套件[20]完成此一檢定。

2.2.2 Cox 比例風險模型(Cox proportional hazard regression model)

此存活分析方法為 Cox 在 1972 年提出[21]，目前是常使用在醫學研究上建立存活預測的統計模型，主要用來研究病人的存活時間與一個或多個預測變量之間的關係，而本研究使用 R 軟體之“survival” 套件[20] 進行估計

$$\begin{aligned}h(t) &= h_0(t) \times \exp(b_1x_1 + b_2x_2 + \dots + b_px_p) \\ &= h_0(t)HR(x)\end{aligned}$$

其中 t 代表存活時間， $h(t)$ 代表在 t 時間點給定 p 個共變數值下的風險函數， $h_0(t)$ 代表在起始點(baseline)所估計的風險值， b_1, b_2, \dots, b_p 分別代表 p 個共變量所



估計出的係數值，「HR」是「Hazard Ratio」，表示在某個時間點之下會發生事件的風險比，因此 $HR(x)$ 就是表示在給定 X 值的情況之下會發生事件的風險比，所謂的 X 值指的就是自變項的數值。此方法可用來估計各種風險因素(risk factors)或臨床治療(clinical treatments) 與存活之間的風險比(hazard ratio)。

2.2.3 臨床變項分析

在癌登資料中，為了看在不同組織型態中期別、淋巴結侵犯程度(lymph nodes involvement)及腫瘤分級分化之變項是否會影響到探討組織型態之存活函數，我們進行了臨床變項顯著性的分析。在類別變項上，採用較保守的費雪精確檢定 (Fisher's exact test)進行，顯著性皆以 P 值 0.05 為基準，此部分分析將透過 R 統計軟體執行。

2.2.4 決策樹 (Decision Tree)

決策樹是用來處理分類問題的樹狀結構，為監督式機器學習模型(Supervised machine learning)，決策樹選出分類能力最好的屬性做為樹的內部節點，將內部節點的所有不同資料產生出對應的分支，遞迴重複上面的過程直到滿足終止條件 [22]，本研究使用的是 CART (Classification and Regression Tree)演算法，為 Friedman 等人於 1980 年代提出[23]，CART 是一種產生二元樹的技術，以基尼係數(Gini)做為選擇屬性的依據。

假設資料集合 S 包含 n 個類別，基尼係數 $Gini(S)$ 定義為

$$Gini(S) = 1 - \sum_{j=1}^n p_j^2$$



p_j 為在 S 中的值組屬於類別 j 的機率。

利用屬性 A 分割資料集合 S 為 S_1 與 S_2 (二元分割)。則根據此一分割要件的吉尼係數 $Gini_A(S)$ 為

$$Gini_A(S) = \frac{|S_1|}{|S|} Gini(S_1) + \frac{|S_2|}{|S|} Gini(S_2)$$

其中， S_1 與 S_2 是針對欄位 A 內的不同數值所構成的兩組資料子集合。

$$\text{不純度的降低值為: } \Delta Gini(A) = Gini(S) - Gini_A(S)$$

挑選擁有最大不純度的降低值、或吉尼係數 $Gini_A(S)$ 最小的屬性作為分割屬性，找出分割屬性後，再進行樹狀結構的分類，而分類的變數是存活結果，CART 將存活時間進行指數的轉換(exponential scaling)，再將轉換結果放入 Poisson 模型中 [24]，即可使用 CART 決策術進行分類。本研究使用 R 軟體的 rpart 套件[25]完成。



第三章 結果

3.1 資料篩選

在台灣癌症登記長表(2013-2017)的胰臟癌病人資料中，原始樣本共 5,685 個登記資料，經過篩選條件後(圖 1)，共計 3,361 位病人進行後續分析，再篩選掉病理期別未知及腫瘤大小遺失後，其中 855 筆為有進行根除性之手術病人。另一方面，在台灣癌症登記年報檔(1995-2017)的胰臟癌病人資料中，經過篩選後(圖 2)共有 31,898 位病人進行組織亞型之分析。此外，本研究之外部驗證資料為美國癌症登記資料庫，經過同樣篩選條件(圖 3)後，在 2010-2015 年度中共有 26,784 個樣本，再篩選掉期別未知及腫瘤大小遺失後，其 4,552 位有進行根除性之手術病人，可進行期別之分析，而組織亞型的資料則共有 67,370 位病人可進行分析。

3.2 臨床變數結果

如圖 5B，有無接受手術的病人存活在 Kaplan- Meier 存活曲線中有顯著的差異 (對數秩檢定 p 值小於 0.0001)，可看到手術為影響存活之重要的分層因子 [26]。在接受根除性手術之 855 位病人中，進一步的輔助性治療為化療居多，其中有 62.69% 的病人會再接受化療(表 4)。

我們將有手術的病病人之病理期別進行 AJCC 第八版的分期(表 2 及表 3)，從原本登記檔的 AJCC 第七版改成 AJCC 的第八版，數量結果如圖 6 所示，結果可以看到第七版分類上病人在第二期別佔最大宗(82%)，而第八版的分類結果，則以



第三期病人佔最多數(84%)，因此我們認為有手術的第三期病人中有存活情況可能有異質性存在，將在 3.3 節進行變數及存活差異的討論。

3.3 探討影響胰臟癌病人的預後因子及分類

3.3.1 決策樹(CART)挑變數的結果

再有 719 位手術第三期的病人中，去除掉變數的遺失值後，剩下 701 位有手術第三期病人，圖 7 為決策樹分類結果。由圖 7 可見，演算法依據變數之重要性挑選了腫瘤分級分化做為第一個分支，將腫瘤分化良好及分化中度合為一組，共有 552 位病人，而腫瘤分化不良及未分化組有 149 位病人。其次，被挑選為第二重要的變數為淋巴結侵犯程度，根據決策樹的分析結果，後續分析上我們將使用腫瘤分級分化和淋巴結侵犯程度重要預後因子。

3.3.2 透過 Cox 風險比例模式找出重要的預後因子

由於未分化的樣本為小樣本($n=6$)，且在對數秩檢定中，分化不良和未分化為不顯著(p 值為 0.847)，因此將分化不良及未分化合併為同一組，為分化不良組(Grade3)。我們使用 cox 風險迴歸模型尋找胰臟癌手術病人的預後因子(表 5)，經調整其他干擾因子後，不同腫瘤分級分化中，「中度」的死亡風險為「良好」的 1.70 倍(95%信賴區間為[1.03-2.79], p 值為 0.004)，「不良」的死亡分險為「良好」的 3.80 倍(95%信賴區間為[2.25-6.39], p 值小於 0.0001)；淋巴結擴散程度中，N1 的死亡風險為 N0 的 1.52 倍(95%信賴區間為[1.18-1.94], p 值小於 0.0001)，N2 的死亡分險為 N0 的 2.30 倍(95%信賴區間為[1.71-3.08], p 值小於



0.0001)。在分類樹(3.3.1)的分析中亦得到相似的結果，這些數據皆說明腫瘤分級分化和淋巴結擴散程度為顯著的預後因子，後續我們探討兩因子之間的關聯。

3.3.3 腫瘤分級分化跟淋巴結擴散程度之關聯分析

將台灣癌症登記處腫瘤分級分化和淋巴結擴散程度進行對數秩檢定，結果顯示合併腫瘤分級分化為三組後，分級分化在統計上顯著差異(p 值皆小於 0.05)。在淋巴結擴散程度中，三種類彼此間皆有顯著差異(p 值皆小於 0.05)。

接著，我們使用卡方獨立性檢定來檢測兩變數間的獨立性(表 6)，結果為顯著相關(p 值小於 0.0001)，分級分化及淋巴結擴散程度有相當程度的關聯，確認其關聯性後，為了區別病人間存活異質性，我們將兩變數作為本研究分類的預後因子，使用兩變數做交互作用項的探討。

3.3.4 探討兩變數存活曲線之加成性

腫瘤分級分化和淋巴結擴散程度為重要預後因子，為了找出與病人存活間的異質性，我們以交互作用細分兩變數之間的加成性，分析上可產生共九種加成性結果(圖 8)，N0-Grade1 與 N2-Grade3 之存活曲線分別為最佳與最差，在腫瘤分級分化及淋巴結擴散加成性下，能分出組別間的差異性(對數秩檢定 p 值小於 0.0001)。其餘存活曲線間互有交錯，然而 Kaplan- Meier 分析圖僅提供視覺化的結果，因此我們進一步做對數秩檢定，觀察彼此組間的差異(表 7)，結果組間顯著差異的趨勢相當不一致，無法作為分群依據，下一節我們將以分層分析來檢定預後因子間的影響。




3.3.5 使用分層分析變數的加成效果

我們分別固定兩預後因子來探討分層加成效果，首先，固定淋巴結擴散程度，探討其分層後變數的影響(圖 9A, B)，淋巴結擴散程度低時，腫瘤分級分化的三個組別有統計上顯著差異(如表 8A, B)，淋巴結擴散程度嚴重下，腫瘤分級分化的影響沒有統計上顯著差異(表 8C, 所有 p 值皆大於 0.05)；再者如圖 10 所示，固定腫瘤分級分化(表 9B)，在腫瘤分化程度中度時，淋巴結擴散程度的影響有統計上顯著的差異，而腫瘤分化良好及不良的組別中，淋巴結擴散程度沒有統計上顯著的差異(表 9A, C, 所有 p 值皆大於 0.05)，而在分化良好的組別中，推測可能是因為樣本數較少的緣故導致沒有顯著差異。在預後因子程度輕微下，另一預後因子的影響有顯著差異；反之則不然，病患存活率皆不高，程度嚴重時另一因子影響不顯著，本結果確定預後因子間具有加成性，接下來我們會使用決策樹分群九種交互作用的存活狀況。

3.4.6 透過決策樹進行預後因子的分群

在 3.4.5 節中，由於對數秩檢定無法將九條存活曲線進行分群，我們改以決策樹進行分群，使用預後因子的交互作用為分群變數。701 位病人共分為七個節點(node)、四個組別(圖 11)：預後最高組(N0-Grade1, N0- Grade2, N1- Grade1)、預後中高組(N1-Grade2, N2-Grade1)、預後中低組(N0-Grade3, N2-Grade2)及預後最低組(N1-Grade3, N2-Grade3)。以 Kaplan-Meier 存活曲線呈現分組後的存活狀況(圖 12)可看到，預後最高組的病人有 200 位(佔 28%)，中位存活時間為 23.5 個月(95%信賴區間為 21.0 個月至無窮大)；而預後最低組的中位存活時間最短，此組



的病人有 108 位(佔 15%)，中位存活時間為 9.0 個月(95%信賴區間為 7.9 至 11.4 個月)，預後最低組的死亡風險為預後最高組的 4.13 倍(95%信賴區間為[3.01-5.68], p 值小於 0.0001)；而預後中高組的中位存活時間為 18.4 個月(95%信賴區間為 15.8 至 21.2 個月)，本組的死亡風險為預後最高組的 1.63 倍(95%信賴區間為 [1.23-2.15], p 值為 0.0009)；而預後中低組的中位存活時間為 14.5 個月(95%信賴區間為 11.2 至 16.9 個月)，此組的死亡風險為預後最高組的 2.46 倍(95%信賴區間為[1.81-3.37], p 值小於 0.0001)，我們使用對數秩檢定確定分組後是否存在組間差異，四組間有顯著差異(p 值小於 0.0001)，分群結果顯示此兩項變數能顯著地將病人存活狀況區分，進而協助病患之精準分類。

3.4.7 輔助性化療在四組間的影響

化療為胰臟癌常見的術後輔助性治療，我們欲確認化療是否對四個組別有顯著影響。實務上，化療對病患的影響約為兩年，因此我們將兩年設為切點，分析化療與存活是否有統計上顯著的差異。在預後最高組、中低及最低組中(圖 13A, C, D)，接受化療的死亡風險為沒化療的 0.40 倍(95%信賴區間為[0.24-0.69], p 值為 0.0009)、0.61 倍(95%信賴區間為[0.36-1.04], p 值為 0.07)及 0.40 倍(95%信賴區間為[0.24-0.65], p 值為 0.0002)，而在預後中高組中，化療與否沒有統計上顯著不同(圖 13B)，推測是因為化療藥物效果不足。若將預後中高組的時間點切為一年時，化療與否就具有統計上顯著的差異(p 值為 0.026)，在存活狀況不同的組別中，接受化療為重要的預後因子，可在臨床上規劃輔助性化療的治療方針，進而提升治療效果。



3.4 外部資料驗證結果：美國癌症登記資料庫


為了確保本研究之結論具有外推性，我們也針對美國癌症登記資料庫進行驗證，確認本結果是否可套用於西方資料。

3.4.1 透過 Cox 風險比例模式

同樣將有手術之 3,495 位第三期胰臟癌病人作為驗證樣本，移除掉變數的遺失值後，共有 3,360 位病人。cox 風險比例模式結果中(表 10)，在腫瘤分級分化中，「中度」的死亡風險為腫瘤分化「良好」的 1.54 倍(95%信賴區間為[1.32-1.79], p 值小於 0.0001)，「不良」的死亡分險為「良好」的 1.93 倍(95%信賴區間為[1.65-2.24], p 值小於 0.0001)；在淋巴結擴散程度中，N1 的死亡風險為 N0 的 1.44 倍(95%信賴區間為[1.29-1.59], p 值小於 0.0001)，N2 的死亡分險為 N0 的 1.84 倍(95%信賴區間為[1.65-2.05], p 值小於 0.0001)，因此可證明腫瘤分級分化和淋巴結擴散程度在美國癌症登記資料庫中亦為顯著的預後因子。

3.4.2 驗證台灣癌症登記分群結果

以 Kaplan-Meier 存活曲線呈現分組後的存活狀況(圖 14A)，預後最高組的中位存活時間最長，此組的病人有 705 位(佔 21%)，中位存活時間為 26 個月(95%信賴區間為 24 至 29 個月)；而預後最低組中位存活時間最短，此組的病人有 1008 位(佔 30%)，中位存活時間為 14 個月(95%信賴區間為 13 至 15 個月)，預後最低組的死亡風險為預後最好組的 2.15 倍(95%信賴區間為[1.91-2.42], p 值小於 0.0001)；而預後中高組的中位存活時間為 21 個月(95%信賴區間為 20 至 22 個



月)，本組的死亡風險為預後最佳組的 1.58 倍(95%信賴區間為[1.39-1.79], p 值小於 0.0001)；而預後中低組的中位存活時間為 18 個月(95%信賴區間為 16 至 20 個月)，本組的死亡風險為預後最佳組的 1.65 倍(95%信賴區間為[1.46-1.88], p 值小於 0.0001)，接著我們使用對數秩檢定驗證分組後是否存在組間差異，結果為四組間有顯著差異(p 值小於 0.0001)，結果與先前一致。

3.4.3 輔助性化療在四組間的影響

將時間點切兩年探討四組間化療的影響(圖 15)，四個組別中結果都為接受化療與否有達統計上顯著的差異。在預後最高的組別中，接受化療的死亡風險為沒化療的 0.62 倍(95%信賴區間[0.48-0.81], p 值為 0.0003)，在預後中高組，接受化療的死亡風險為沒化療的 0.41 倍(95%信賴區間[0.33- 0.51] p 值小於 0.0001)；在預後中低組，接受化療的死亡風險為沒化療的 0.49 倍(95%信賴區間[0.40- 0.60], p 值小於 0.0001)；在預後最差組中，接受化療的死亡風險為沒化療的 0.46 倍(95%信賴區間[0.39- 0.55], p 值小於 0.0001)，總結來說，本研究方法驗證兩國之資料結果為一致，輔助性化療在四組中皆有顯著影響，因此我們確認了本分群結果可以套用於在西方資料中，具有外推性。

3.5 不同組織型態的結果

胰臟癌的組織型態數量取自於台灣癌症登記處年報檔(表 11)，我們欲探討樣本數足夠且較稀有的組織型態。結果顯示胰臟實性假乳突狀癌(8452)之存活曲線在八個組織型態中為最佳(圖 16)，而胰管內乳突狀黏液性癌(8453)為存活曲線最佳的組織型態，圖下接近之四條存活曲線的組織型態分別為未分化型之胰臟癌



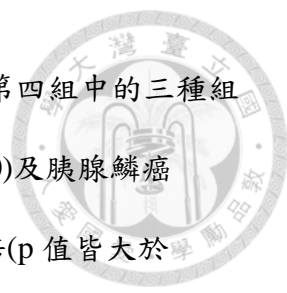
(8020)、胰腺鱗癌(8560)及鱗狀細胞胰臟癌(8070)則為最差的，我們在 3.5.1 會使用決策樹將八個組織型態進行分群。

3.5.1 透過決策樹進行組織型態的分群

我們透過決策樹，對 1,357 位病人之八個組織型態進行分群，分為七個節點、五個組別，由預後佳至差分為第一組至第四組(圖 17A)，第一組(8452)的病人有 163 位(佔 12%)，四年存活率為 90%；第二組(8453)的病人有 226 位(佔 16%)，中位存活時間為 41 個月(95%信賴區間為 26.9 個月至無窮大)，第二組的死亡風險為第一組的 5.52 倍 (95%信賴區間為[2.84-10.76], p 值小於 0.0001)；而第三組(8470, 8480, 8550)的病人有 599 位(佔 44%)，中位存活時間為 13.5 個月(95%信賴區間為 11.87 至 15.47 個月)，第三組的死亡風險為第一組的 9.66 倍(95%信賴區間為[5.08-18.39], p 值小於 0.0001)；第四組(8020, 8560,8070)的病人有 369(佔 28%)，中位存活時間為 5.4 個月(95%信賴區間為 4.8 個月至 6.1 個月)，第四組的死亡風險為第一組的 20.32 倍(95%信賴區間為[10.59-38.97], p 值小於 0.0001)。對數秩檢定結果顯示四組的組間有顯著差異(圖 17B, p 值小於 0.0001)，不同組織型態會影響預後結果，但並非所有組織型態預後結果都很差。

3.5.2 探討分群後的組內差異

為了確認組內是否有差異，我們將組內大於一種組織型態(第三組，第四組)進行檢驗，對數秩檢定結果顯示第三組中之三種組織型態具有顯著差異(圖 18A, p 值為 0.0061)。cox 風險比例模式中，黏液性胰腺癌(8480)的死亡風險為黏液性囊腺癌(8470)的 1.04 倍(p 值為 0.71)，胰臟腺泡細胞癌(8550)的死亡風險為黏液性



囊腺癌(8470)的 0.89 倍(p 值為 0.44)，死亡風險比值趨近 1；第四組中的三種組織型態有顯著差異(圖 18B, p 值為 0.007)，鱗狀細胞胰臟癌(8070)及胰腺鱗癌(8560)的死亡風險為未分化型之胰臟癌(8020)的 1.02 倍及 0.81 倍(p 值皆大於 0.05)，未達統計上顯著差異，依據 cox 風險比例模式的結果，第三組及第四組組內差異並不明顯。

3.5.3 外部驗證資料結果：美國癌症登記資料庫

我們以本研究分群方式套用於美國癌症登記資料庫(2,874 位病人)驗證是否具有外推性(圖 19A)，其趨勢與台灣癌症登記資料庫相同，四組組間具有顯著差異(圖 19B, p 值小於 0.0001)。在組內差異中，第三組之三個組織型態具有顯著差異(圖 20A, p 值小於 0.0001)，在 cox 風險比例模式中，黏液性腺癌(8480)的死亡風險為黏液性囊腺癌(8470)的 1.58 倍(p 值小於 0.0001)，胰臟腺泡細胞癌(8550)的死亡風險為黏液性囊腺癌(8470)的 0.93 倍(p 值為 0.61)。我們推論第三組中的組織型態可能有種族間的差異。第四組中的三種組織型態有顯著差異(圖 20B, p 值為 0.014)，鱗狀細胞胰臟癌(8070)及胰腺鱗癌(8560)的死亡風險為未分化型之胰臟癌(8020)的 1.09 倍及 1.25 倍(p 值皆大於 0.05)，未達統計上顯著差異，表示第四組組內差異並不明顯。結果顯示除第三組外，本分群方式具有外推性。



第四章 結論與討論

4.1 主要發現

影響存活的重要預後因子為腫瘤分級分化和淋巴結擴散程度，並可依存活狀況的差異進行分群，胰臟癌本身的組織型態亦是具有病患存活比例之差異，本研究可協助臨床醫師擬訂不同的治療選擇，進而提升病人的治療效果。

在胰臟癌第三期中已接受手術的病人，其存活狀況中存在異質性。先前已有研究淋巴結擴散程度對於期別分類是重要的因子，能將病人存活曲線分類[27]，本研究發現腫瘤分級分化也會影響病人的存活狀況。分層分析後，我們發現僅在腫瘤分級分化良好及中度下，不同淋巴結擴散程度會影響病人之存活，兩預後因子具有加成性。接著使用決策樹分為四個組別(預後最高、中高、中低及最低組)，四組間皆有顯著差異，本研究結果亦能驗證於美國癌症登記資料庫，具有外推性。

組織型態分析中，胰臟實性假乳突狀癌及胰管內乳突狀黏液性癌這兩種組織型態是胰臟癌中預後較佳的[14, 28, 29]，而胰腺鱗癌之預後比胰腺癌還差[30]。組織型態分群中，四組組間皆具有顯著差異。第三組中，臺美資料分析後的結果較為不同，美國資料中，不同組織型態間之存活狀況具有顯著差異，我們推論可能有種族間的差異造成分析結果不同。

胰臟癌雖然為存活率極低的癌症，但其病人的存活狀況確實具有異質性。本研究提出同時使用兩預後因子可進一步歸納出四組預後結果，可提供臨床上作輔

助性治療的藥物決策。此外，組織型態在病人中亦有顯著差異[31]，我們藉由回顧性研究資料分析台灣與美國胰臟癌病人中可能具有種族間的差異。



4.2 研究討論

4.2.1 腫瘤分級分化在其他病理期別中的影響

以 COX 風險比例模型分析樣本數較少的期別，觀察腫瘤分級分化是否為重要預後因子。發現在第一期及第二期病人中，腫瘤分級分化等級對胰臟癌病人的存活並沒有顯著差異。在第四期病人中，腫瘤分級分化為主要影響存活的預後因子，「中度」的死亡風險為「良好」的 5.49 倍(95%信賴區間為[1.10-27.36], p 值為 0.037)，「不良」的死亡風險為「良好」的 10.43 倍(95%信賴區間為[1.79-60.65], p 值為 0.009)，對數秩檢定結果 p 值為 0.09(圖 19)，由於第四期樣本數較少的關係，造成信賴區間差距較大。我們認為在較嚴重期別中，腫瘤分級分化的程度會影響病人存活間差異，因此在臨床上可納入此變數評估病人的狀況。

4.2.2 組織型態的影響

在第三組中，美國癌症登記資料庫具有組內顯著的差異，與台灣資料結果不同，可能因種族差異而造成組織型態不同或是樣本數較少而無法精準反映實際狀況。另外，組織型態的存活差異可能源自於腫瘤本身或是基因序列差異導致，未來可對這部分更進一步的探討。此外，仍有部分組織型態過於稀少且樣本數不足，無法納入本研究進行分析。



4.2.3 研究限制

本研究使用台灣癌症登記處的長表資料，但是長表申報是從 2013 年開始，各家醫院才陸續申報胰臟癌病人，因此本研究分群使用的樣本數相對於美國癌症登記資料庫是較少的。此外資料集本身的變數收錄不足(例如：基因的變項)，雖然優點是台灣癌症族群(population-based)資料庫，但可能會有未知的共變項(missing covariates)，未來可納入這些變項進行研究。研究設計的限制為回顧性研究資料的搜集的目的並不是為了本研究的設計，因此研究設計可能存在偏差(bias)，這個偏差可能是已知但無法在本研究處理(例如：未知的共變項)。


4.3 臨床與公共衛生的意義

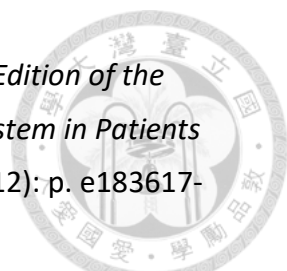
本研究使用台灣癌症登記資料庫，為台灣具有代表性且最大的癌症資料庫，研究結果找出腫瘤分級分化為影響存活的重要因子，可納入考慮病患預後狀況，協助醫師判斷輔助性化療的用藥等級，期望可以應用到其他亞洲國家的病人。本研究結果亦證實分群狀況具有外推性。組織型態可能因種族不同而不一致，未來仍需要進一步的研究，期望能在全球的治療計畫上提供幫助。

參考資料



1. McGuigan, A., et al., *Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes*. World journal of gastroenterology, 2018. **24**(43): p. 4846-4861.
2. Ilic, M. and I. Ilic, *Epidemiology of pancreatic cancer*. World journal of gastroenterology, 2016. **22**(44): p. 9694-9705.
3. Luo, J., et al., *The incidence and survival rate of population-based pancreatic cancer patients: Shanghai Cancer Registry 2004-2009*. PLoS One, 2013. **8**(10): p. e76052.
4. Wang, H., et al., *Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015*. The Lancet, 2016. **388**(10053): p. 1459-1544.
5. Hariharan, D., A. Saied, and H.M. Kocher, *Analysis of mortality rates for pancreatic cancer across the world*. HPB : the official journal of the International Hepato Pancreato Biliary Association, 2008. **10**(1): p. 58-62.
6. Bond-Smith, G., et al., *Pancreatic adenocarcinoma*. BMJ : British Medical Journal, 2012. **344**: p. e2476.
7. Health Promotion Administration, Ministry of Health and Welfare, Taiwan. *Taiwan Cancer Registry Annual Report of 2016*.
8. *World Cancer Report 2014*. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015
9. American Cancer Society. *Cancer Facts & Figures 2010*. Atlanta: American Cancer Society; 2010.
10. Cascinu, S., et al., *Pancreatic cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology, 2010. **21**: p. v55-v58.
11. Amin MB, E.S., Greene F, et al, *AJCC Cancer Staging Manual, 8th edition*. 2017: New York: Springer.
12. Fesinmeyer, M.D., et al., *Differences in survival by histologic type of pancreatic cancer*. Cancer Epidemiol Biomarkers Prev, 2005. **14**(7): p. 1766-73.

- 
13. Shuja, A. and K.A. Alkimawi, *Solid pseudopapillary tumor: a rare neoplasm of the pancreas*. Gastroenterol Rep (Oxf), 2014. **2**(2): p. 145-9.
 14. Machado, N.O., H. Al Qadhi, and K. Al Wahibi, *Intraductal Papillary Mucinous Neoplasm of Pancreas*. North American journal of medical sciences, 2015. **7**(5): p. 160-175.
 15. Yagci, A., et al., *Diagnosis and treatment of solid pseudopapillary tumor of the pancreas: experience of one single institution from Turkey*. World Journal of Surgical Oncology, 2013. **11**(1): p. 308.
 16. Chiang, C.-J., Y.-W. Wang, and W.-C. Lee, *Taiwan's Nationwide Cancer Registry System of 40 years: Past, present, and future*. Journal of the Formosan Medical Association, 2019. **118**(5): p. 856-858.
 17. Chiang, C.-J., et al., *Quality assessment and improvement of nationwide cancer registration system in Taiwan: a review*. Japanese Journal of Clinical Oncology, 2015. **45**(3): p. 291-296.
 18. *SEER*Stat Database: NPCR and SEER Incidence—U.S. Cancer Statistics Public Use Database*, Nov 2019 submission (2000–2016)
 19. Mantel, N., *Evaluation of survival data and two new rank order statistics arising in its consideration*. Cancer Chemother Rep, 1966. **50**(3): p. 163-70.
 20. Therneau, T.M., *A Package for Survival Analysis in R*. 2020.
 21. Cox, D.R., *Regression Models and Life-Tables*, in *Breakthroughs in Statistics: Methodology and Distribution*, S. Kotz and N.L. Johnson, Editors. 1992, Springer New York: New York, NY. p. 527-541.
 22. Chen, S.-J. *Decision Tree Learning course slides*. 2005.
 23. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. , *Classification and Regression Trees*. 1984: Monterey, CA: Wadsworth and Brooks.
 24. LeBlanc, M. and J. Crowley, *Relative risk trees for censored survival data*. Biometrics, 1992. **48**(2): p. 411-25.
 25. Terry Therneau and Beth Atkinson, *rpart: Recursive Partitioning and Regression Trees*, 2019. R package.
 26. Bilimoria, K.Y., et al., *Validation of the 6th edition AJCC Pancreatic Cancer Staging System: report from the National Cancer Database*. Cancer, 2007. **110**(4): p. 738-44.

- 
27. Van Roessel, S., et al., *International Validation of the Eighth Edition of the American Joint Committee on Cancer (AJCC) TNM Staging System in Patients With Resected Pancreatic Cancer*. *JAMA Surgery*, 2018. **153**(12): p. e183617-e183617.
 28. Izumo, W., et al., *Evaluation of preoperative prognostic factors in patients with resectable pancreatic ductal adenocarcinoma*. *Scand J Gastroenterol*, 2019. **54**(6): p. 780-786.
 29. Law, J.K., et al., *A Systematic Review of Solid-Pseudopapillary Neoplasms: Are These Rare Lesions?* *Pancreas*, 2014. **43**(3).
 30. Kardon, D.E., et al., *Adenosquamous carcinoma of the pancreas: a clinicopathologic series of 25 cases*. *Mod Pathol*, 2001. **14**(5): p. 443-51.
 31. Hackeng, W.M., et al., *Surgical and molecular pathology of pancreatic neoplasms*. *Diagn Pathol*, 2016. **11**(1): p. 47.

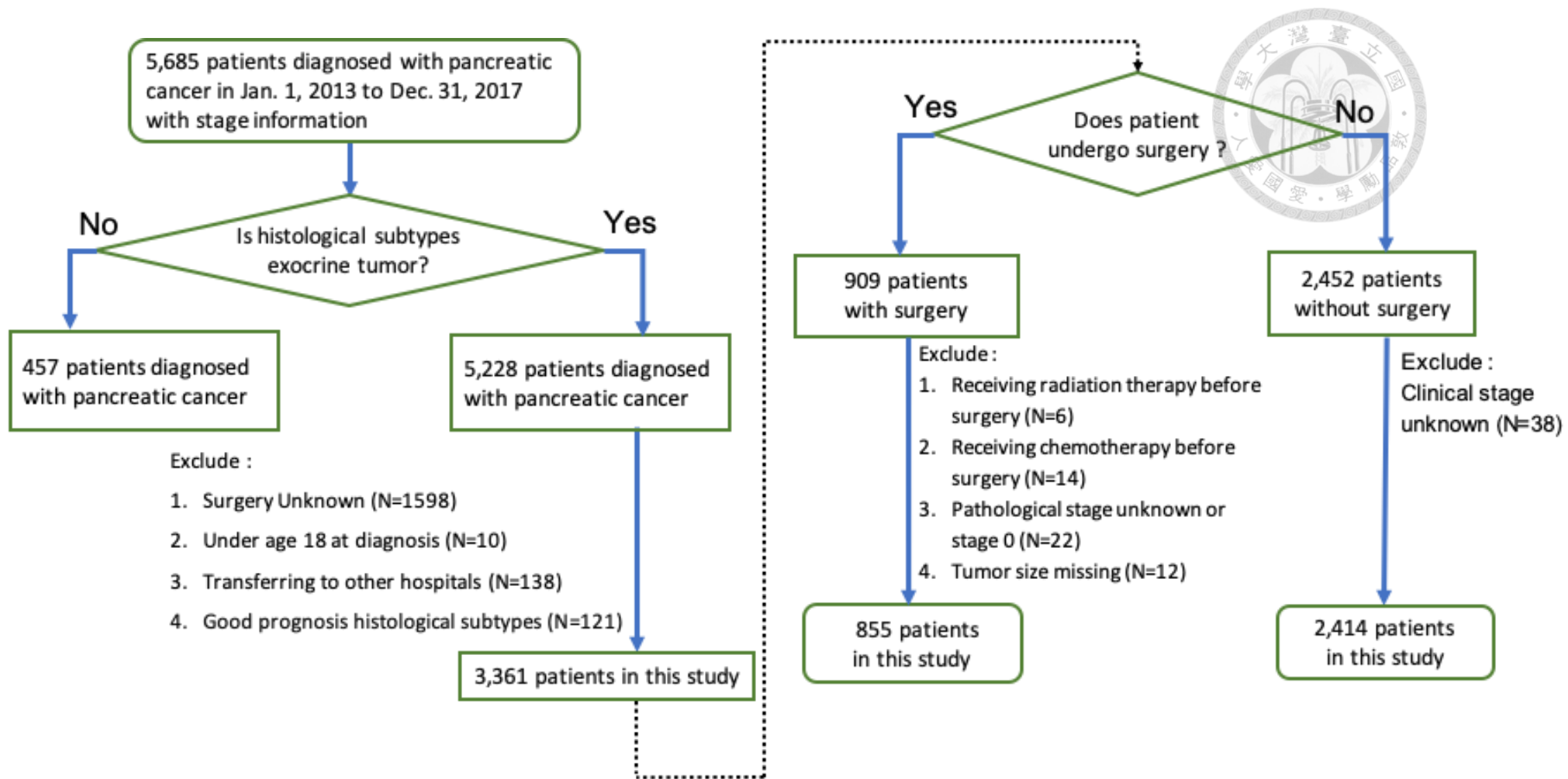


圖 1 胰臟癌台灣癌登資料庫長表篩選流程圖

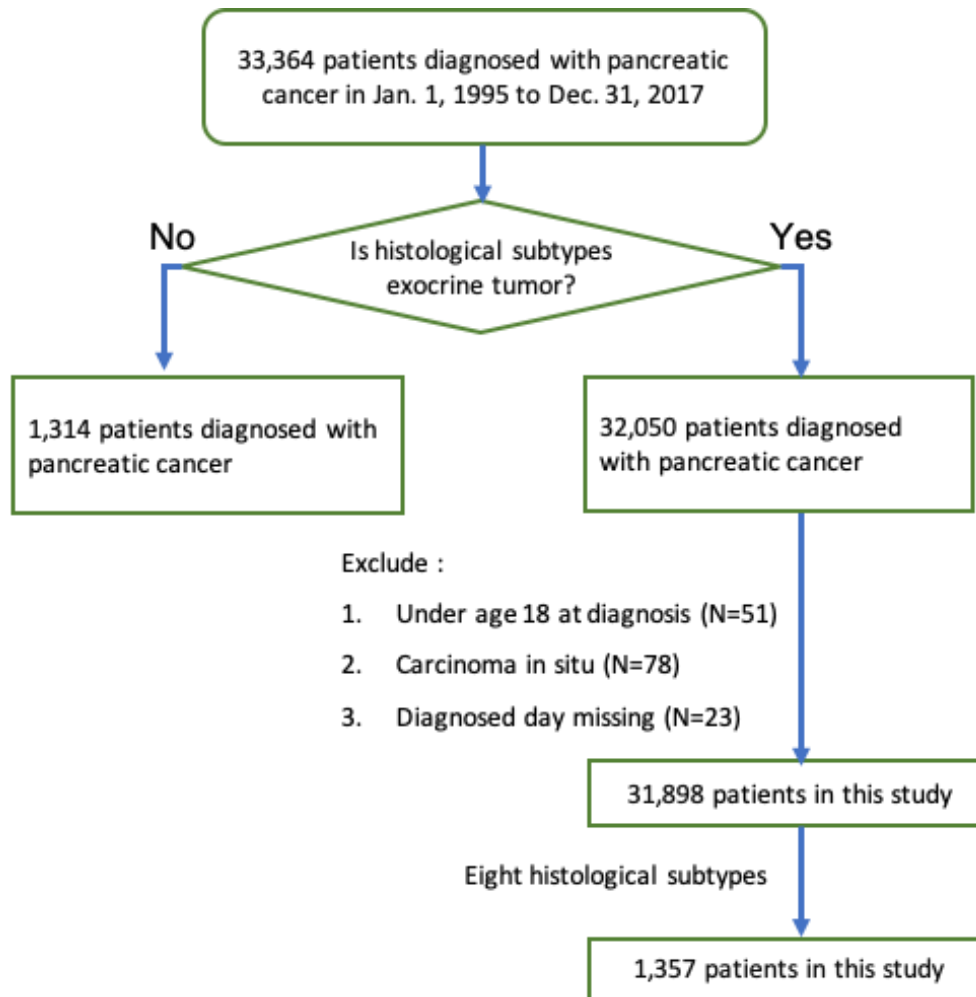


圖 2 胰臟癌台灣癌登資料庫年報檔篩選流程圖

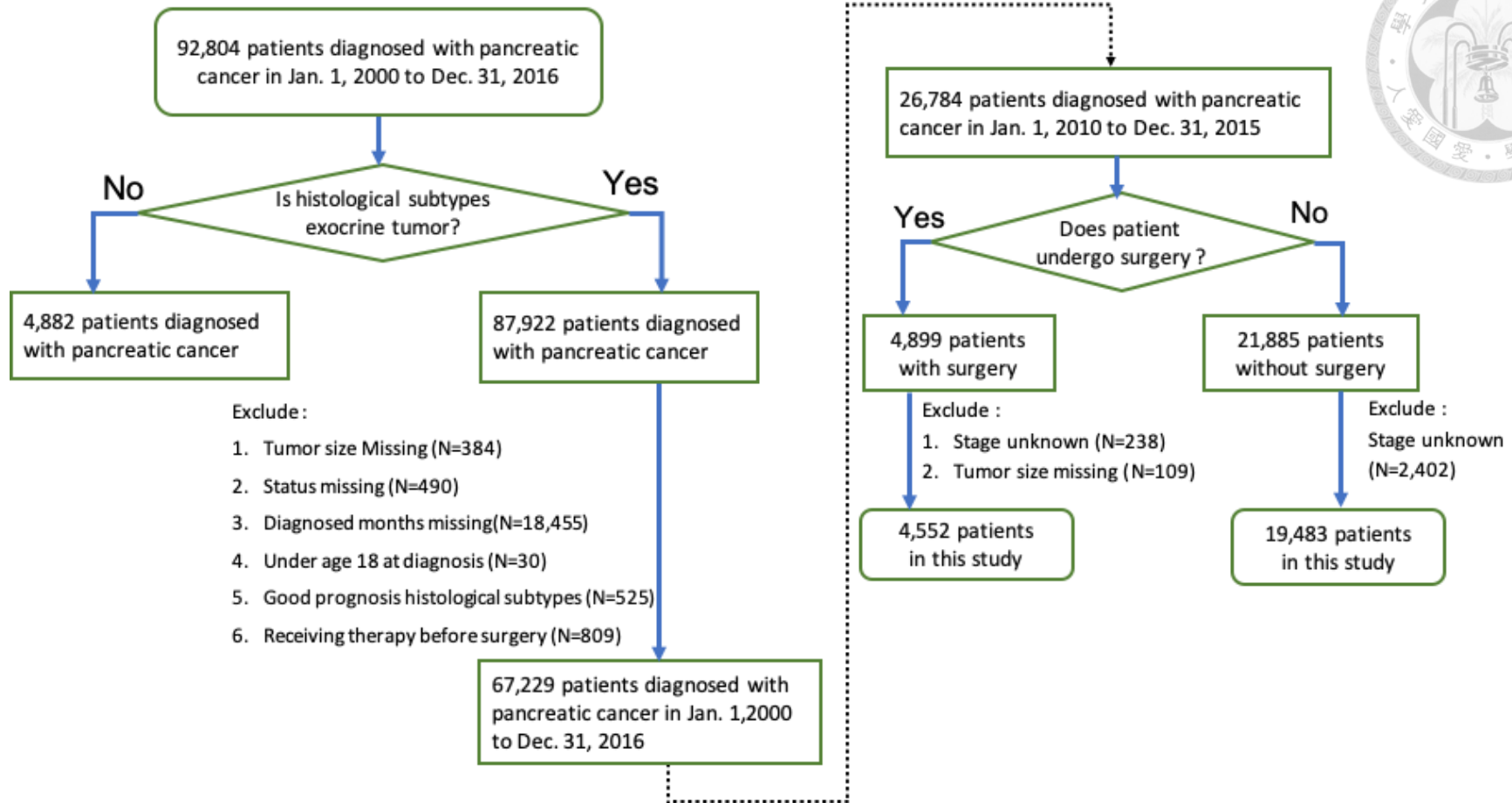


圖 3 胰臟癌美國癌症登記資料庫(2010-2015)篩選流程圖

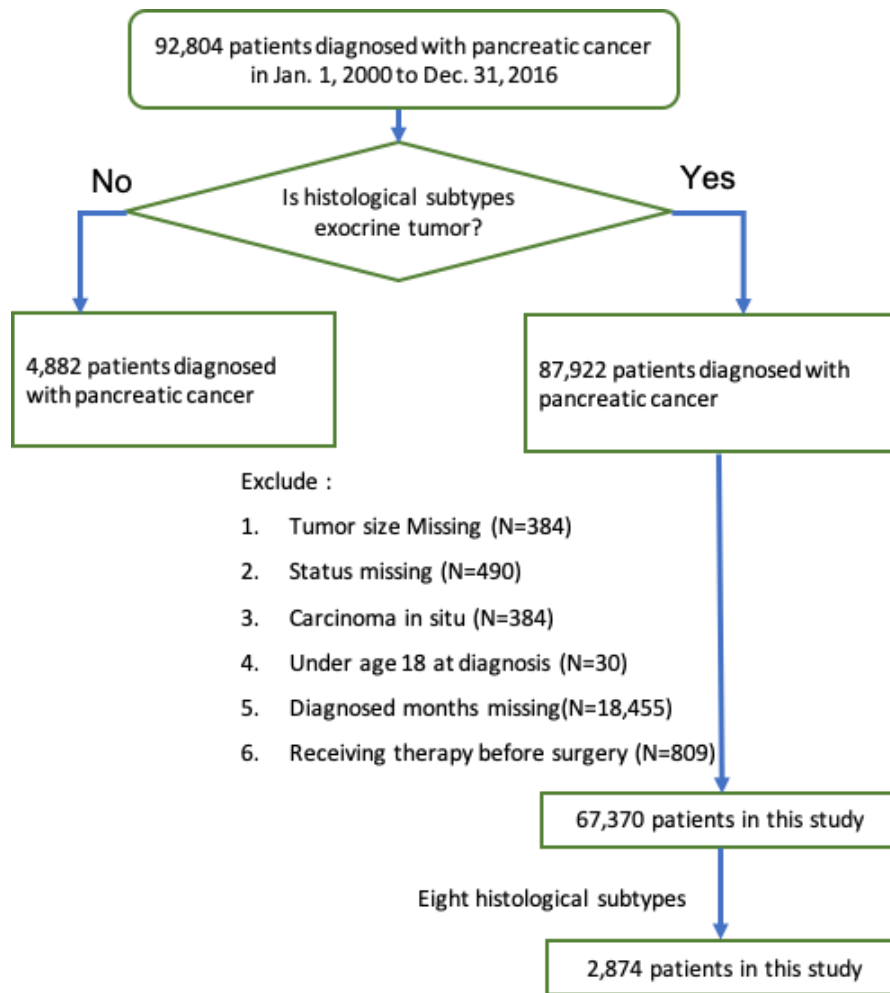
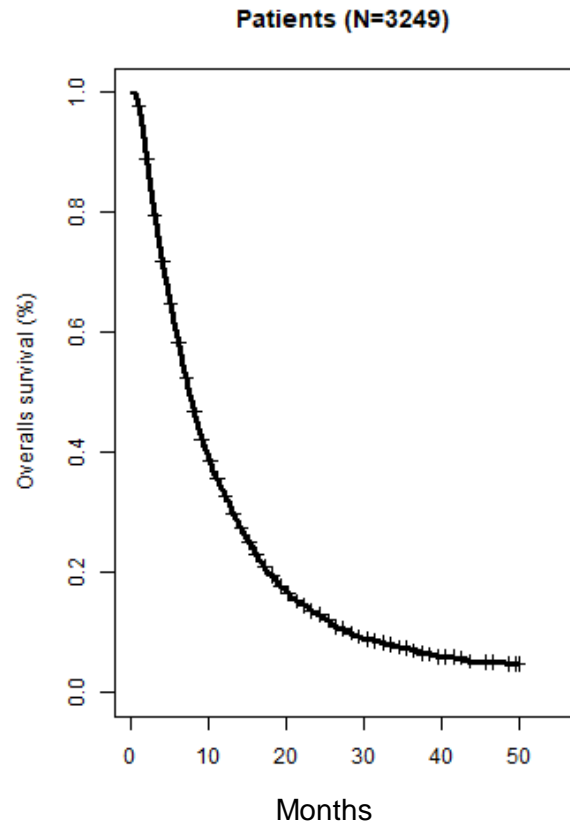


圖 4 胰臟癌美國癌症登記資料庫(2000-2016)篩選流程圖



(A)



(B)

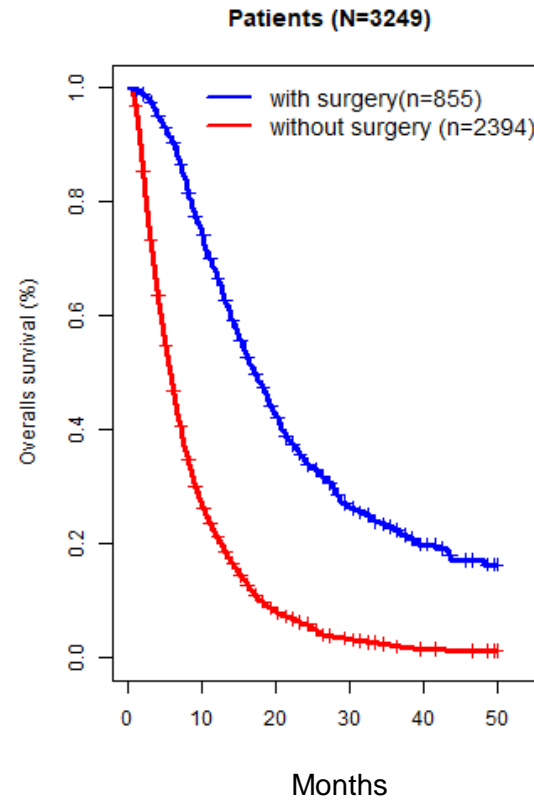


圖 5 台灣癌症登記資料庫胰臟癌(A)所有病人(B)手術變數的 Kaplan-Meier 分析圖

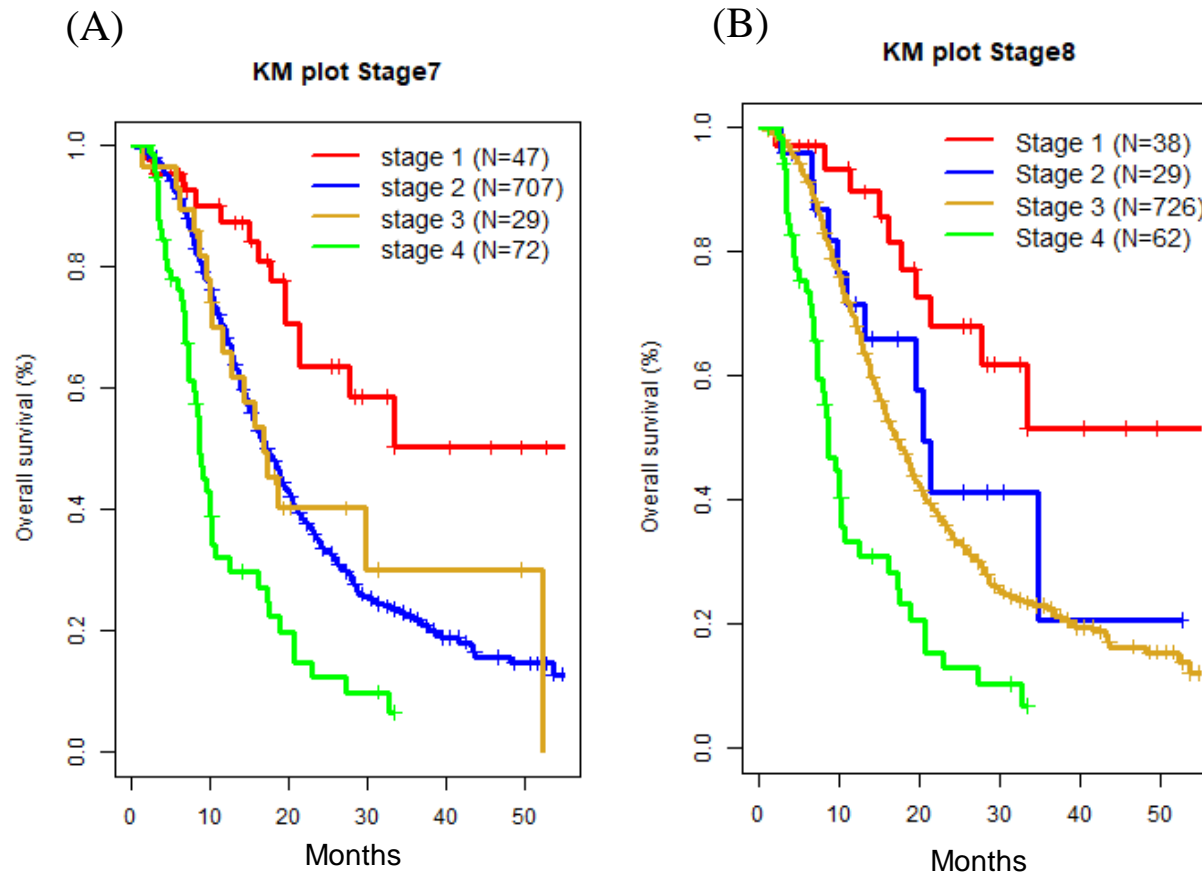


圖 6 台灣癌症登記資料庫(A)AJCC7 及(B)AJCC8 分期別的 Kaplan-Meier 分析圖

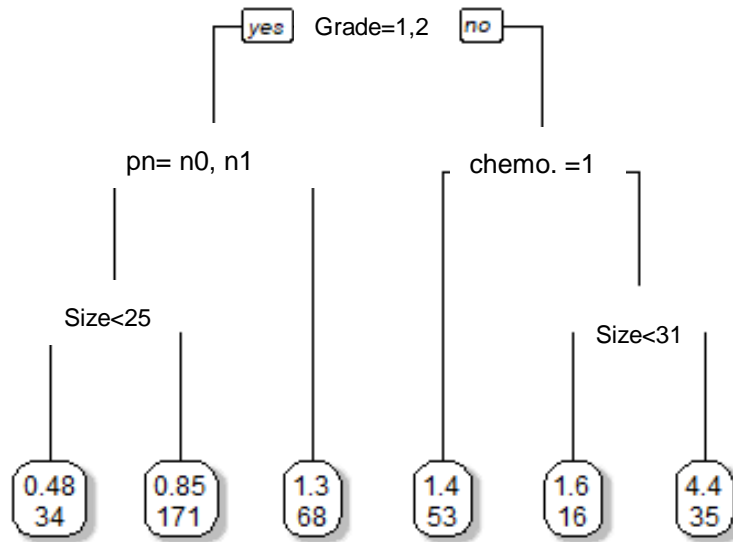


圖 7 台灣癌症登記資料庫之透過決策樹接受手術第三期之重要預後因子

註：Pn 為淋巴結擴散程度



lymph nodes involvement and grade interaction

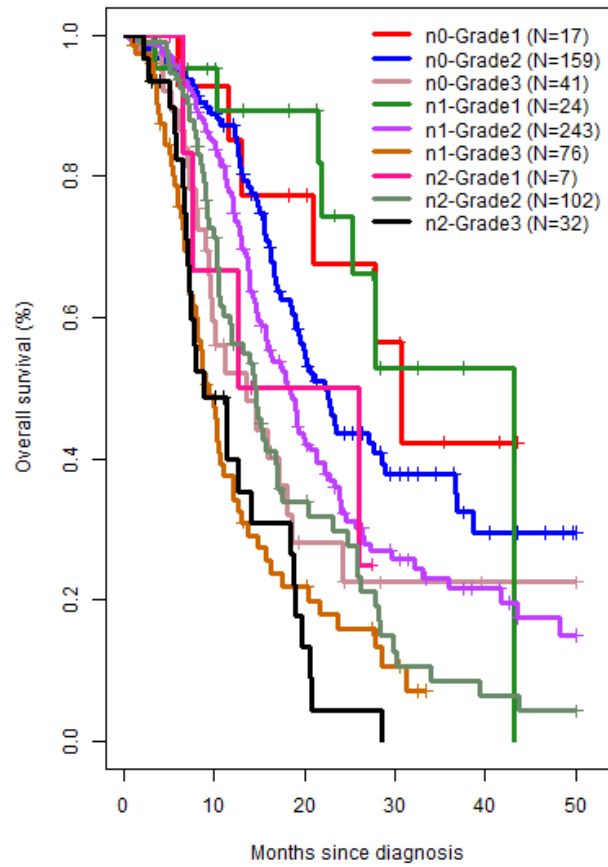


圖 8 台灣癌症登記資料庫之第三期病人的兩變數交互作用的 Kaplan-Meier 分析圖

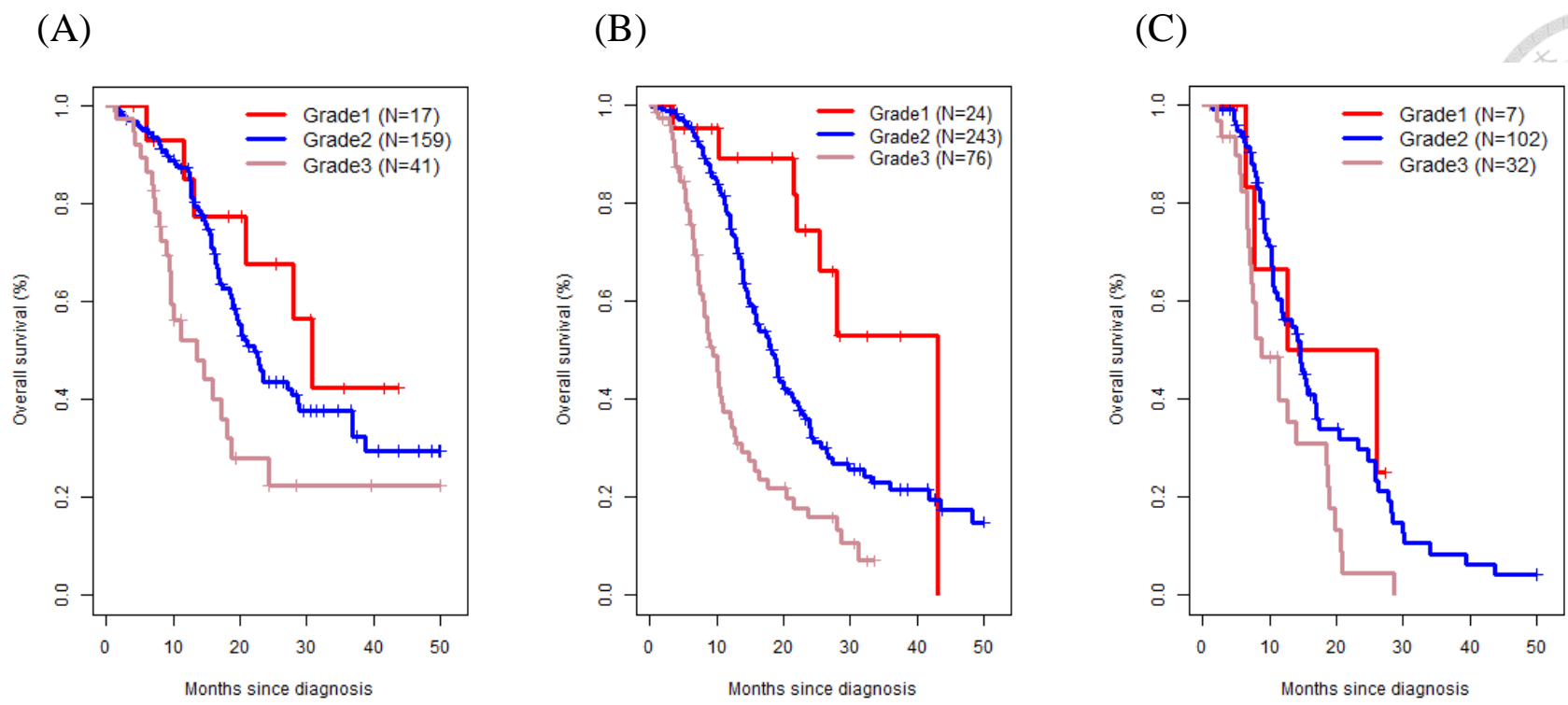
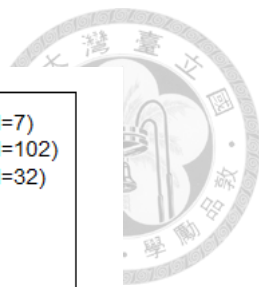
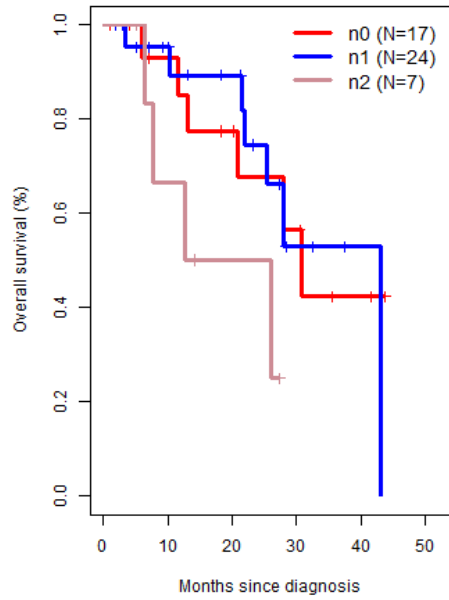
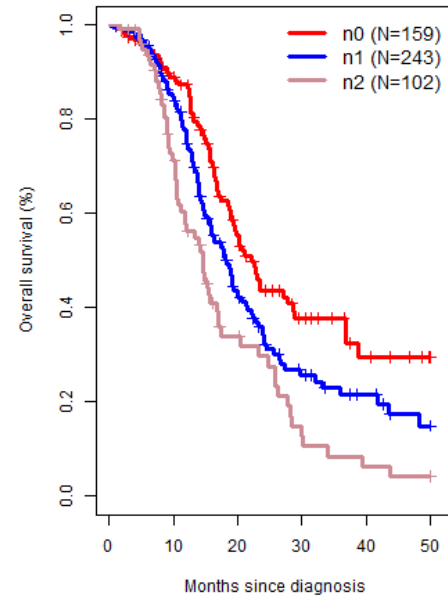


圖 9 台灣癌症登記資料庫(A)固定 N0 下(B)固定 N1 下(C)固定 N2 下的 Kaplan-Meier 分析圖

(A)



(B)



(C)

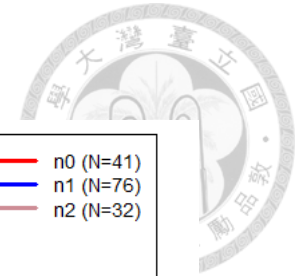
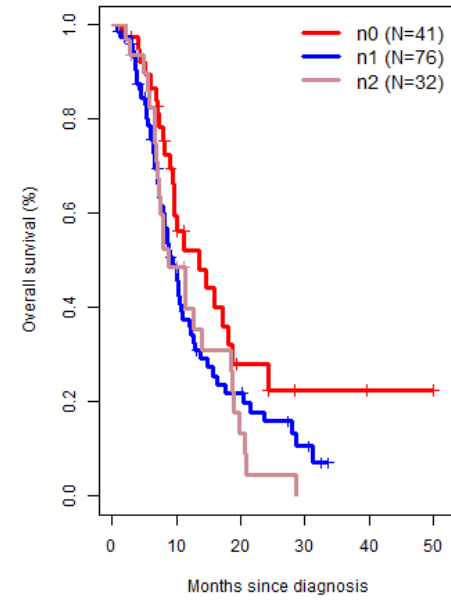


圖 10 台灣癌症登記資料庫(A)固定分化良好下(B)固定分化中度下(C)固定分化不良下的 Kaplan-Meier 分析圖

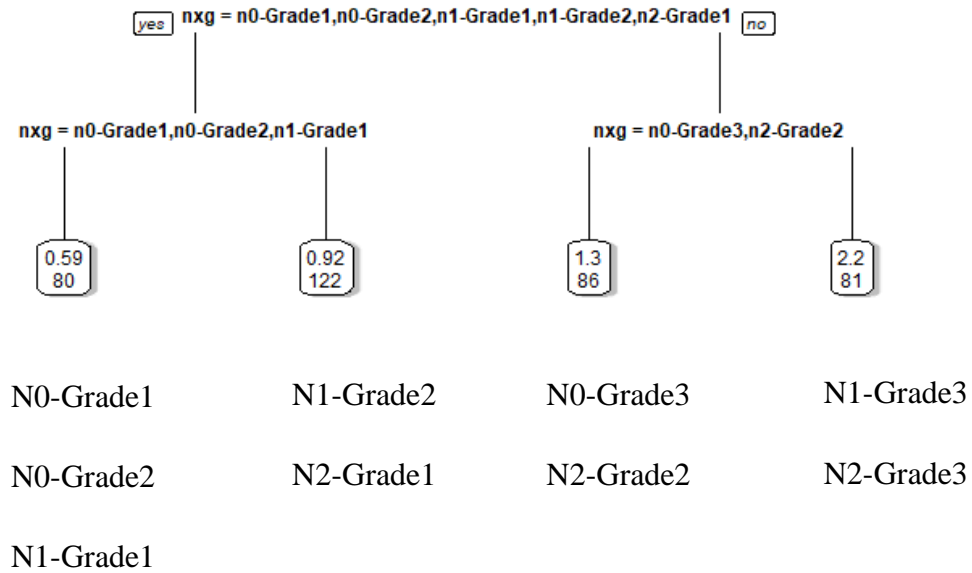


圖 11 台灣癌症登記資料庫之使用兩變數交互作用分群結果

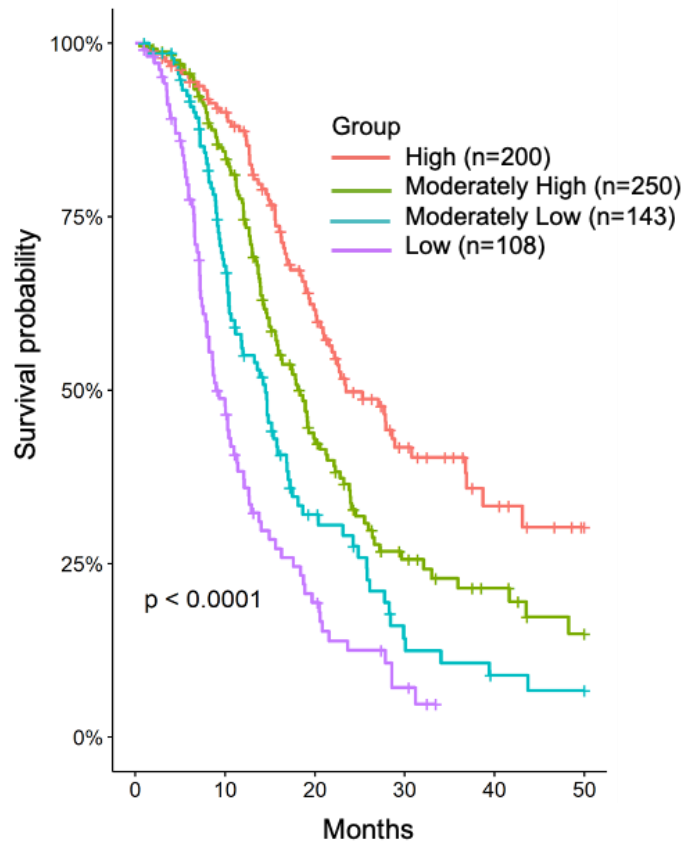


圖 12 台灣癌症登記資料庫四組的 Kaplan-Meier 分析圖

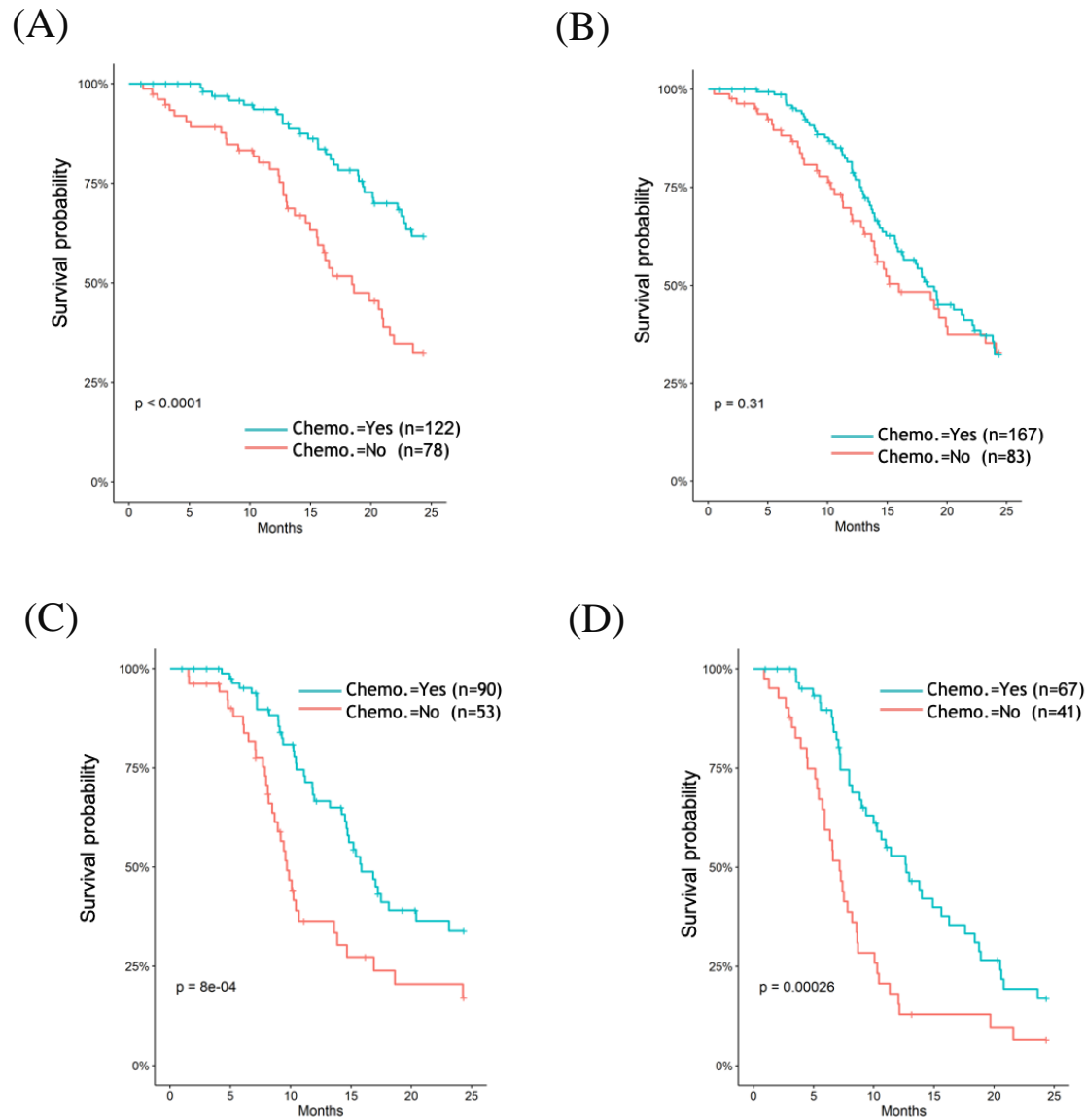


圖 13 台灣癌症登記資料庫預後(A)最高組(B)中高組(C)中低組(D)最差組切兩年化療變數的 Kaplan-Meier 分析圖

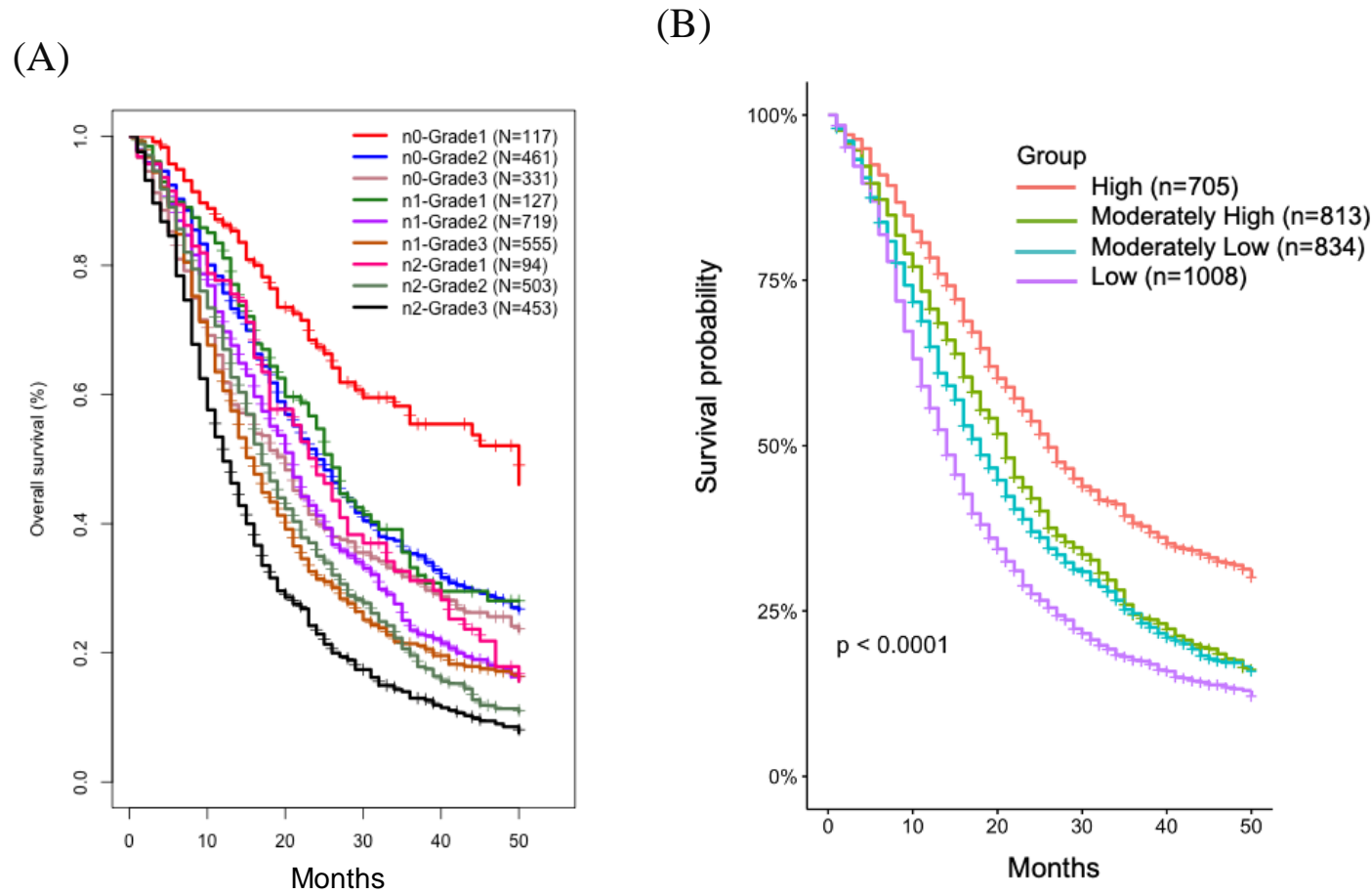


圖 14 美國癌症登記資料庫第三期病人的(A)兩變數交互作用(B)使用台灣癌症分群方法的 Kaplan-Meier 分析圖

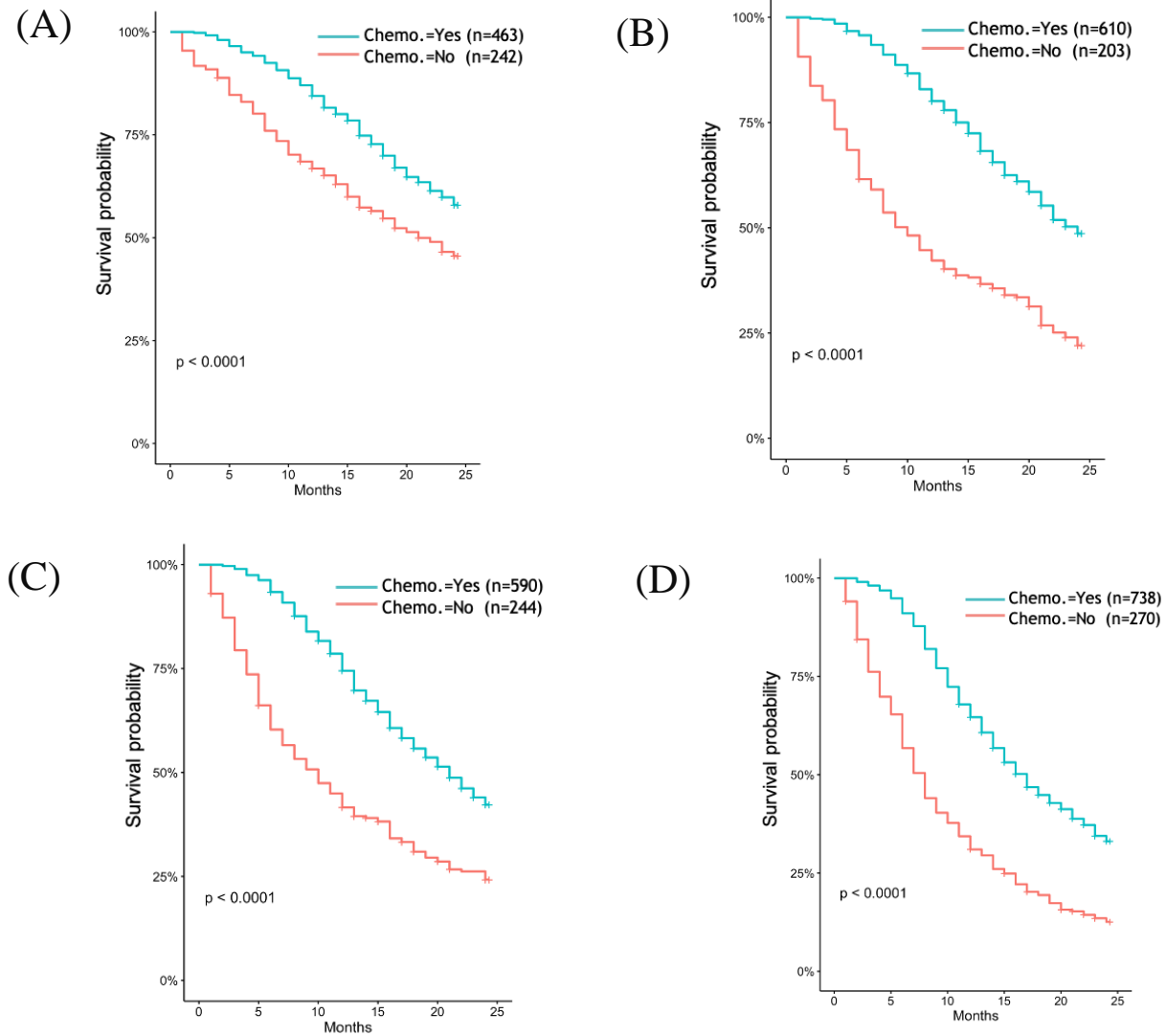


圖 15 美國癌症登記資料庫預後(A)最高組(B)中高組(C)中低組(D)最差組切兩年化療變數的 Kaplan-Meier 分析圖

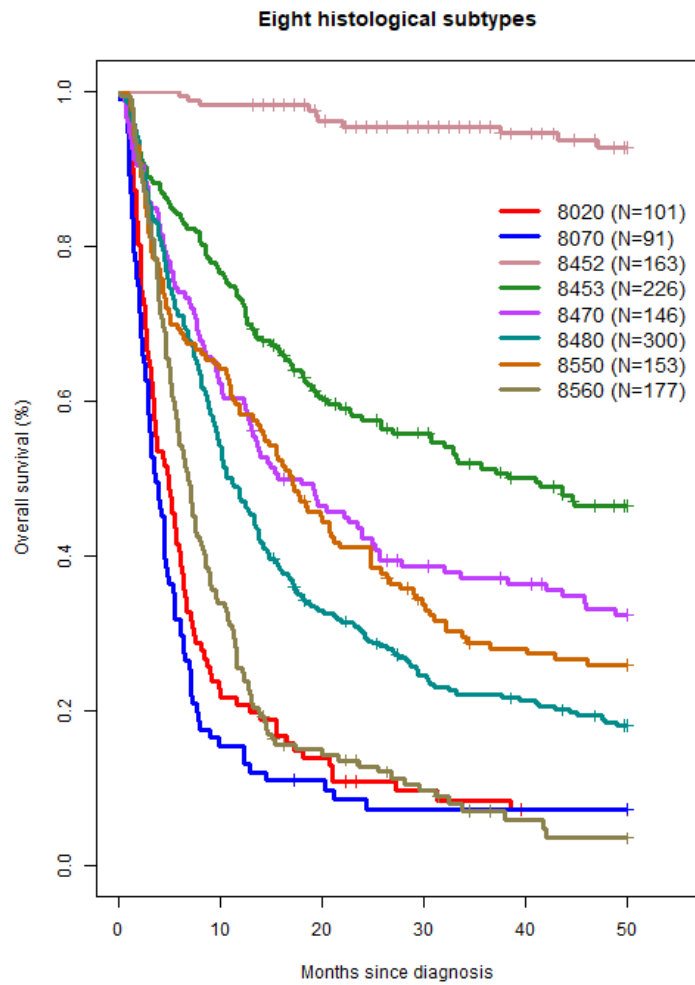


圖 16 台灣癌症登記資料庫組織型態亞型的 Kaplan-Meier 分析圖

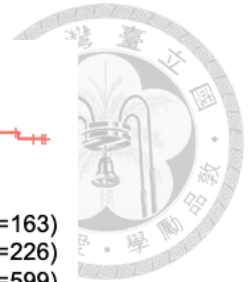
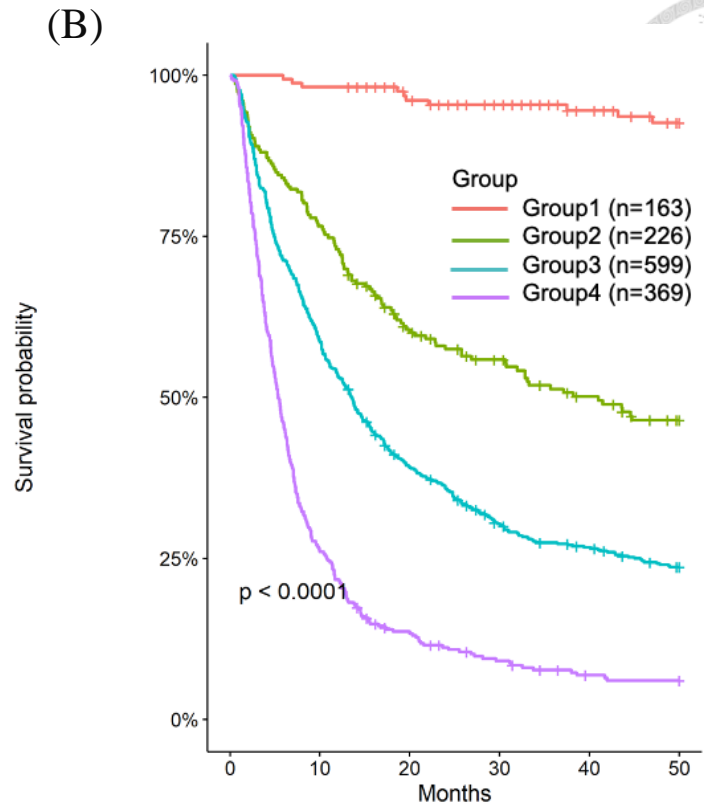
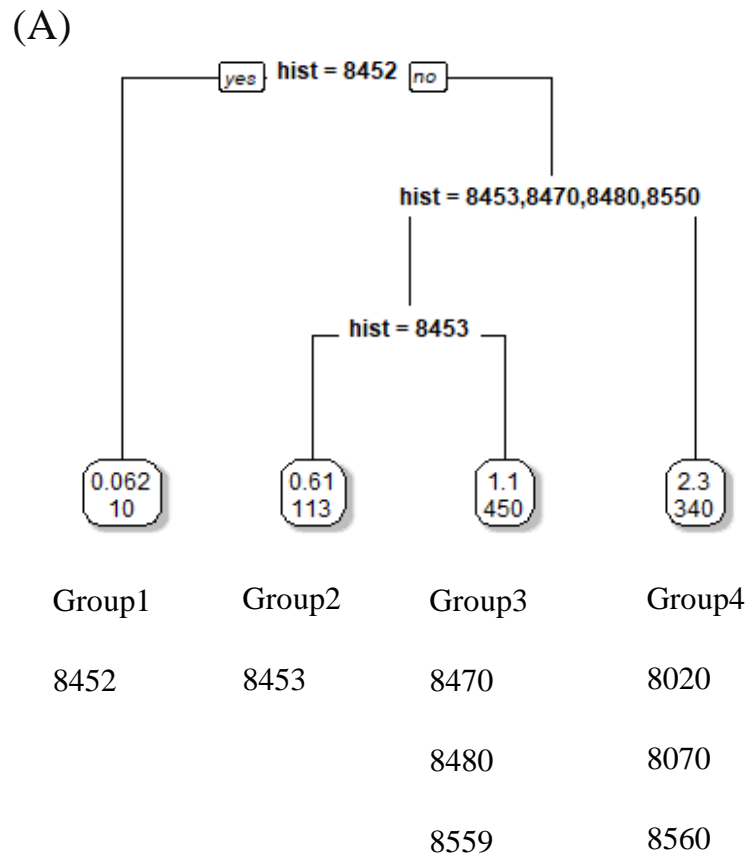
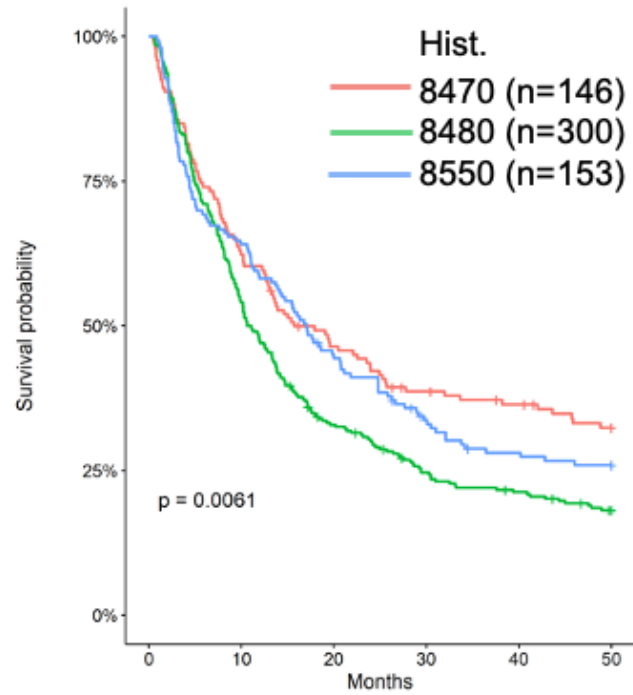


圖 17 台灣癌症登記資料庫之八個組織亞型(A)決策樹分群結果(B)使用台灣癌症分群方法的 Kaplan-Meier 分析圖

(A)



(B)

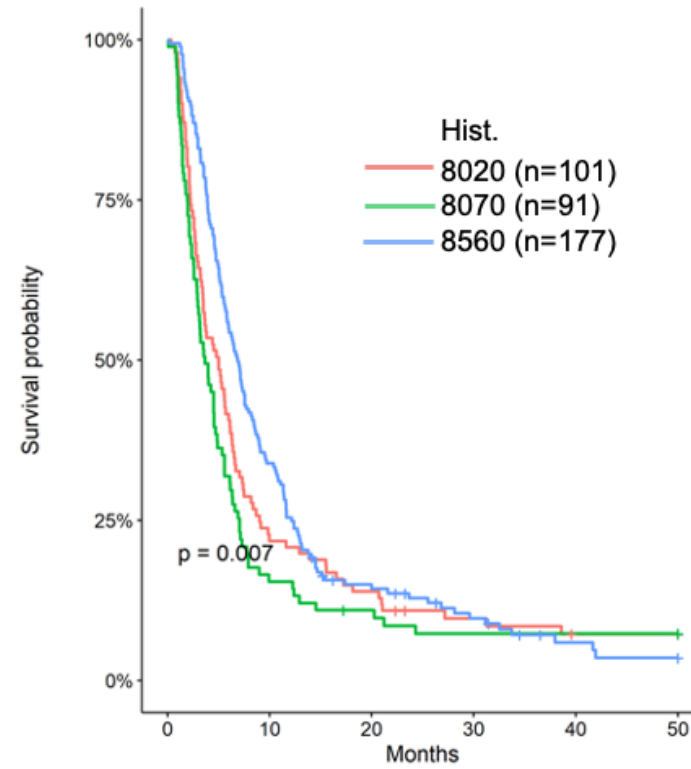
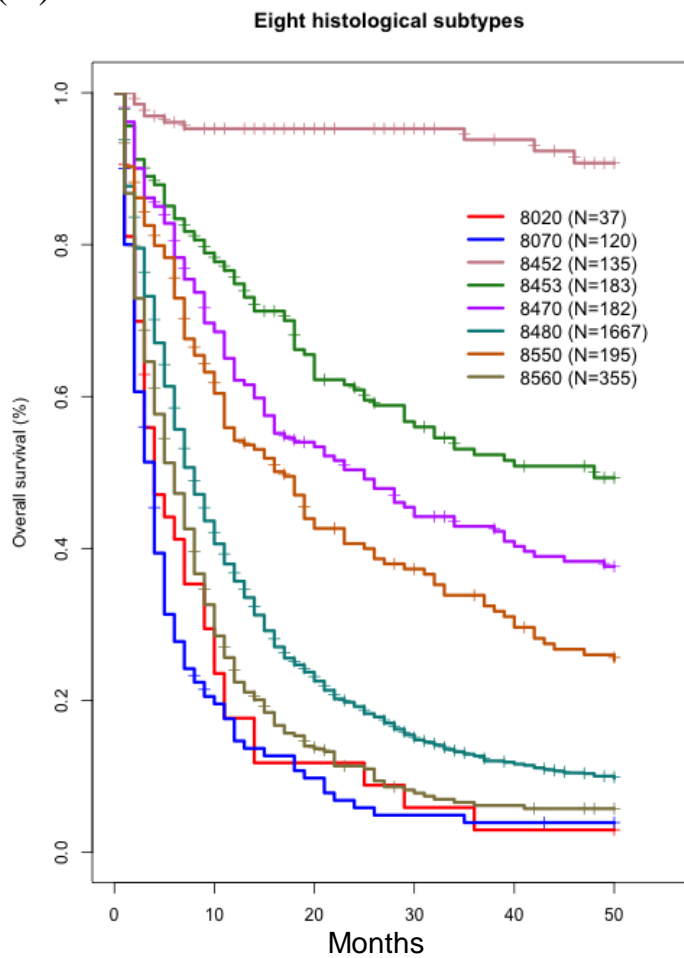


圖 18 台灣癌症登記資料庫分組後(A)Group 3 (B)Group 4 中組內差異的 Kaplan-Meier 分析圖

(A)



(B)

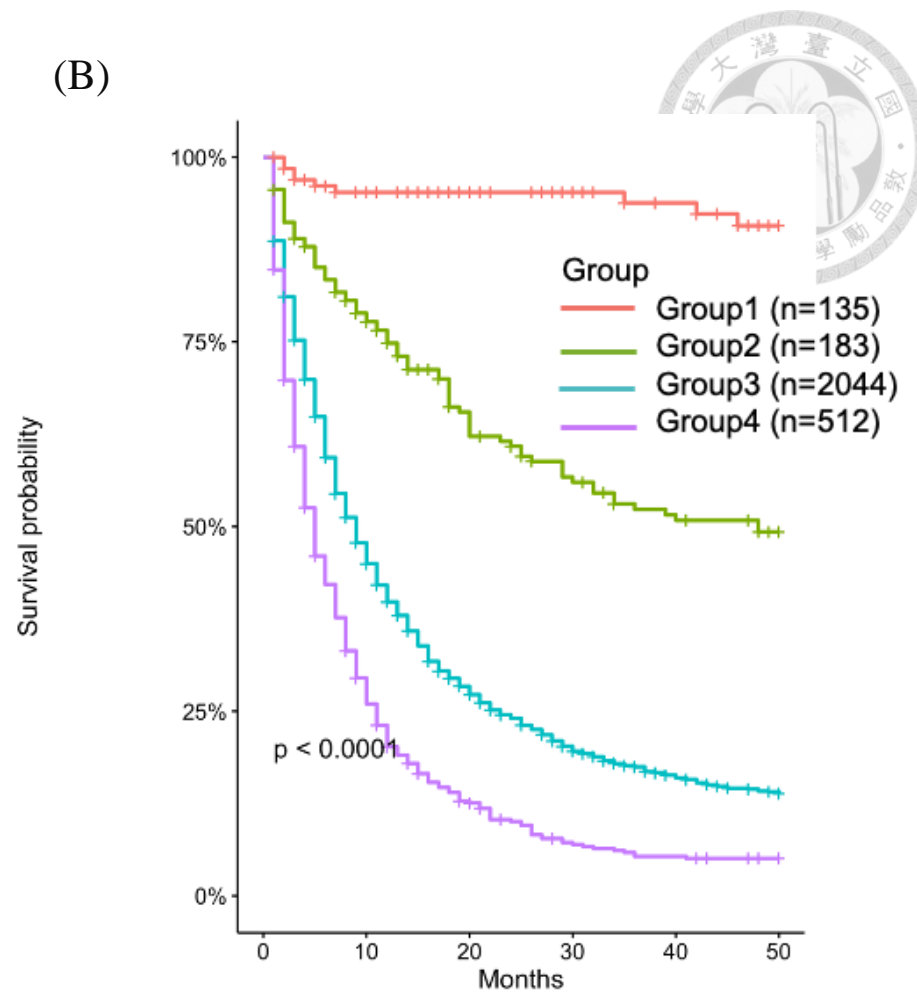


圖 19 美國癌症登記資料庫(A)組織型態(B)分四個組別的 Kaplan-Meier 分析圖

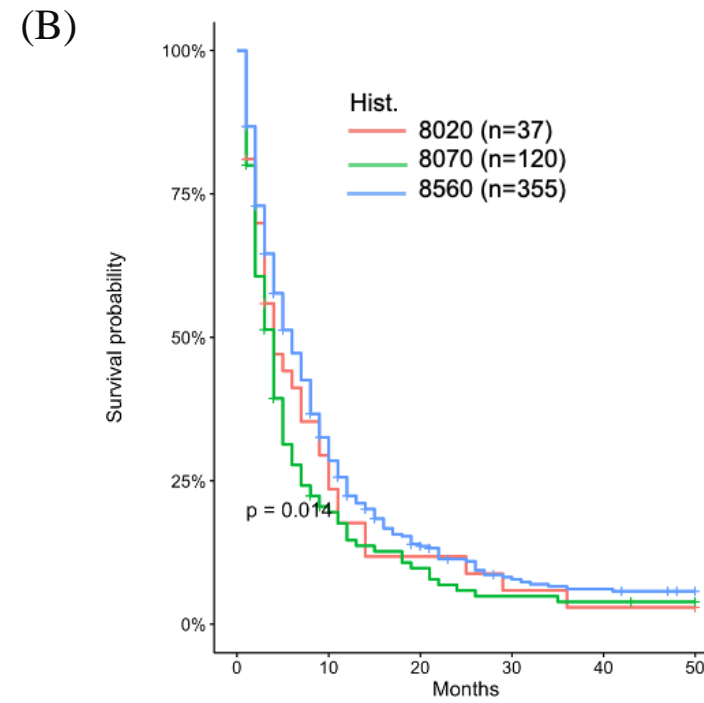
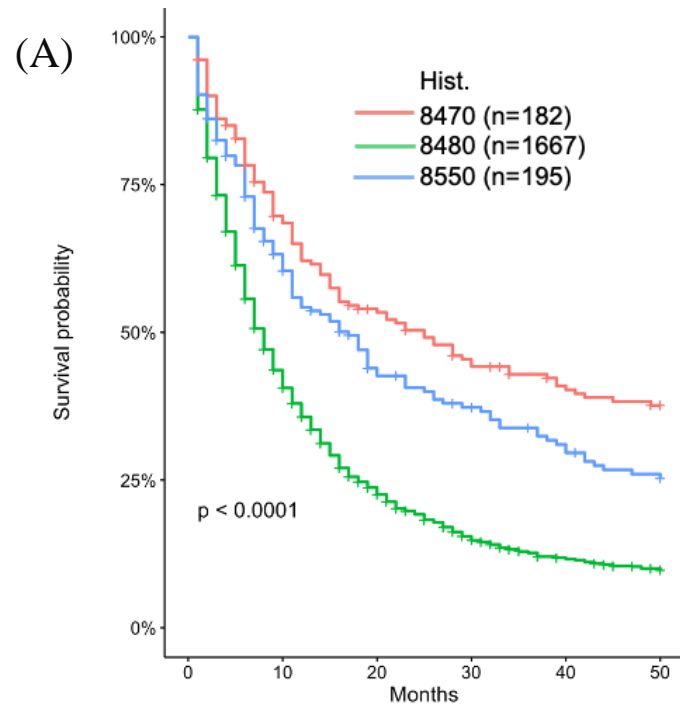


圖 20 美國癌症登記資料庫分組後(A)Group 3(B)Group 4 中組內差異的 Kaplan-Meier 分析圖

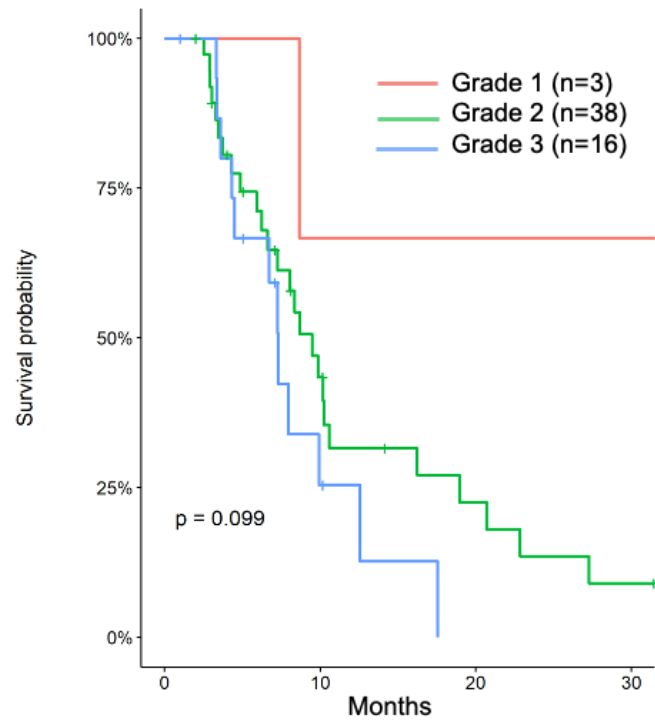


圖 21 台灣癌症登記資料庫之腫瘤分級分化在第四期病人的 Kaplan-Meier 分析圖



表 1 資料描述清單

Dataset	Year	Sample(used)	Stage information	Reference
台灣癌症登記資料庫長表	2013-2017	5685(3269)	With stage	[16,17]
台灣癌症登記資料庫年報檔	1995-2017	33364(31898)	Without stage	
美國癌症登記資料	2000-2016	92804(67370)		[18]
美國癌症登記資料	2010-2015	26784(24035)	AJCC7 same as TCR	



表 2 AJCC7 和 AJCC8 之 TNM 的定義

	AJCC8	AJCC7
T1	size ≤ 20mm	size ≤ 20mm
T2	20mm < size ≤ 40mm	size > 20mm
T3	size > 40mm	Tumor extends beyond the pancreas but without involvement of the celiac axis or the superior mesenteric artery
T4	Tumor involves the celiac axis, common hepatic artery or the superior mesenteric artery	Tumor involves the celiac axis or the superior mesenteric artery (unresectable primary tumor)
N0	No regional lymph node	No regional lymph node
N1	Metastasis in 1-3 lymph nodes	Regional lymph node metastasis
N2	Metastasis ≥ 4 lymph nodes	
M0	No distant metastasis	
M1	Distant metastasis	



表 3 AJCC7 和 AJCC8 版的制定期別定義

	AJCC7			AJCC8		
	T	N	M	T	N	M
stage IA	T1	N0	M0	T1	N0	M0
stage IB	T2	N0	M0	T2	N0	M0
stage II A	T3	N0	M0	T3	N0	M0
stage II B	T1.2.3	N1	M0	T1.2.3	N1	M0
stage III	T4	any	M0	T4 any	any N2	M0
stage IV	any	any	M1	any	any	M1

表 4 台灣癌症登記資料中病人之變數敘述統計

Variables		all (n=3249)		with surgery (n=855)		without surgery (n=2394)	
Sex	male	1782	54.85%	481	56.26%	1301	54.34%
	female	1467	45.15%	374	43.74%	1093	45.66%
Age	≤ 60y	1047	32.23%	323	37.78%	724	30.24%
	> 60y	2202	67.77%	532	62.22%	1670	69.76%
Stage	1	98	3.02%	38	4.44%	60	2.51%
	2	215	6.62%	29	3.39%	186	7.77%
	3	1092	33.61%	726	84.91%	366	15.29%
	4	1844	56.76%	62	7.25%	1782	74.44%
Grade	1	89	2.74%	66	7.72%	23	0.96%
	2	705	21.70%	588	68.77%	117	4.89%
	3	297	9.14%	163	19.06%	134	5.60%
	4	14	0.43%	9	1.05%	5	0.21%
	NA	2144	65.99%	29	3.39%	2115	88.35%
Chemotherapy	yes	2219	68.30%	536	62.69%	1683	70.30%
	no	1030	31.70%	319	37.31%	711	29.70%
Radiation therapy	yes	406	12.50%	104	12.16%	302	12.61%
	no	2843	87.50%	751	87.84%	2092	87.39%



表 5 台灣癌登有手術第三期病人 cox regression 結果

	Variables	HR	P value	95% CI
Continuous variables				
	Age	1.01	0.08	0.99-1.01
	Tumor size	1.00	0.44	0.99-1.00
Categorical variables				
Sex	male	-	-	-
	female	0.99	0.96	0.80-1.22
Grade	1	-	-	-
	2	1.70	0.04	1.03-2.79
	3	3.80	<0.0001	2.25-6.39
Lymph nodes involvement (pn)	0	-	-	-
	1	1.52	<0.0001	1.18-1.94
	2	2.30	<0.0001	1.71-3.08
Chemotherapy	Without	-	-	-
	With	0.58	<0.0001	0.46-0.72
Radiotherapy	Without	-	-	-
	With	0.99	0.93	0.70-1.36

表 6 腫瘤分級分化和淋巴結擴散程度數量表

	Grade 1	Grade 2	Grade 3
N0	17	159	41
N1	24	243	76
N2	7	102	32





表 7 分級和淋巴結擴散程度之交互作用對數秩檢定結果

	n0-Grade1	n0-Grade2	n0-Grade2	n0-Grade2	n0-Grade2	n0-Grade2	n0-Grade2	n0-Grade2
n0-Grade2	0.361	-	-	-	-	-	-	-
n0-Grade3	0.039	0.005	-	-	-	-	-	-
n1-Grade1	0.956	0.192	0.010	-	-	-	-	-
n1-Grade2	0.102	0.027	0.103	0.027	-	-	-	-
n1-Grade3	0.002	<0.0001	0.124	<0.0001	<0.0001	-	-	-
n2-Grade1	0.177	0.355	0.773	0.107	0.845	0.353	-	-
n2-Grade2	0.011	<0.0001	0.956	0.004	0.020	0.020	0.884	-
n2-Grade3	<0.0001	<0.0001	0.117	<0.0001	<0.0001	0.698	0.182	0.014



表 8 分層控制在(A)N0 (B)N1 (C)N2 下 腫瘤分級分化之對數秩檢驗結果

(A)

	Grade1	Grade2
Grade2	0.301	-
Grade3	0.031	0.005

(B)

	Grade1	Grade2
Grade2	0.008	-
Grade3	<0.0001	<0.0001

(C)

	Grade1	Grade2
Grade2	0.834	-
Grade3	0.198	0.016



表 9 分層控制在(A)分化良好(B)分化中度(C)分化不良下 淋巴結侵犯程度之對數秩檢驗結果

(A)

	N0	N1
N1	0.931	-
N2	0.184	0.184

(B)

	N0	N1
N1	0.013	-
N2	<0.0001	0.011

(C)

	N0	N1
N1	0.124	-
N2	0.124	0.601



表 10 美國癌症登記資料庫接受手術第三期病人 cox 回歸模型結果

	Variables	HR	P value	95% CI
Continuous variables				
	Age	1.01	<0.0001	1.00-1.01
	Tumor size	1.01	<0.0001	1.00-1.00
Categorical variables				
Sex	male	-	-	-
	female	1.05	0.28	0.96-1.13
Grade	1	-	-	-
	2	1.54	<0.0001	1.32-1.79
	3	1.93	<0.0001	1.65-2.24
lymph nodes involvement (pn)	0	-	-	-
	1	1.44	<0.0001	1.29-1.59
	2	1.84	<0.0001	1.65-2.05
Chemotherapy	Without	-	-	-
	With	0.54	<0.0001	0.48-0.58
Radiotherapy	Without	-	-	-
	With	0.88	0.01	0.80-0.97

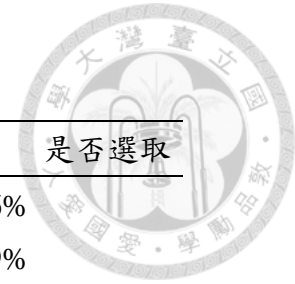


表 11 台灣癌症登記年報檔的組織型態分佈

Hist	名稱	中文名稱	個數	佔比	是否選取
8140	Adenocarcinoma, NOS	胰腺癌	15768	49.56%	
8000	Neoplasm, malignant	胰臟惡性腫瘤	6583	20.69%	
8010	Carcinoma, NOS	未明示型態之胰臟癌	2653	8.34%	
9990	No microscopic confirmation	無顯微鏡確認其組織型態	2502	7.86%	
8500	Invasive carcinoma of no special type	胰臟侵襲性癌	2440	7.67%	
8480	Mucinous adenocarcinoma	黏液性胰腺癌	300	0.94%	v
8453	Intraductal papillary-mucinous carcinoma, invasive	胰管內乳突狀黏液性癌	226	0.71%	v
8560	Adenosquamous carcinoma	胰腺鱗癌	177	0.56%	v
8452	Solid pseudopapillary carcinoma	胰臟實性假乳突狀癌	163	0.51%	v
8470	Mucinous cystadenocarcinoma, NOS	黏液性囊胰腺癌	153	0.48%	v
8550	Acinar cell carcinoma	胰臟腺泡細胞癌	146	0.46%	v
8020	Carcinoma, undifferentiated type, NOS	未分化型之胰臟癌	101	0.32%	v
8070	Squamous cell carcinoma, NOS	鱗狀細胞胰臟癌	91	0.29%	v
8440	Cystadenocarcinoma, NOS	囊胰腺癌	60	0.19%	v
	Other	其他組織型態	454	1.43%	