

國立臺灣大學公衛學院流行病學與預防醫學所

生統組博士論文

Division of Biostatistics, Institute of Epidemiology and Preventive
Medicine, College of Public Health,
National Taiwan University

Doctoral Dissertation

標誌歷程之動態存活預測的統計分析

Statistical Analysis for Dynamic Survival Prediction
Involving Marker Processes

蘇登煌

Deng-Huang Su

指導教授：張淑惠 教授

Advisor : Shu-Hui Chang, Professor

中華民國 105 年 1 月

January, 2016





國立臺灣大學博士學位論文
口試委員會審定書

論文中文題目

標誌歷程之動態存活預測的統計分析

論文英文題目

Statistical Analysis for Dynamic Survival Prediction
Involving Marker Processes

本論文係蘇登煌君（學號 D97842007）在國立臺灣大學
流行病學與預防醫學研究所完成之博士學位論文，於民國
105 年 1 月 27 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：

張淑惠 (簽名)
(指導教授)

邱政民

邱政民

戴政

戴政

鄭系記

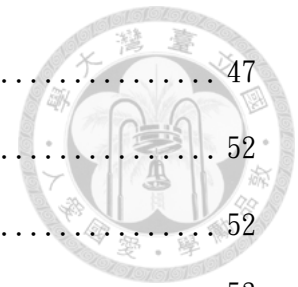
陳秀配

目錄



摘要	iii
第一章 背景和目的	1
第二章 文獻回顧	5
1· 多變數存活資料的分類及特性	5
(I) 平行資料	5
(II) 長期資料	7
2· 多變數存活資料的統計分析	11
(I) 相依性分析	12
(II) 邊際模型的分析	15
(III) 條件與聯合模型	16
(IV) 多階段模式	17
3· 標誌歷程	18
(I) 標誌歷程的應用	19
(II) 右受限的標誌歷程	20
(III) 前進式三階段資料	20
(IV) 前進式三階段資料之統計分析	21
4· 不同 Cox 模式對時間依賴性變數之處理	24
5· 所用之甲狀腺乳突癌資料	24
第三章· 方法	27
1· 符號及資料結構	27
2· 模型	32
3· 估計	34
(I) 以右受限資料估計以標誌歷程為條件的未來存活函數	34
(II) 以存活者的後續有限觀察時間條件存活機率之預測	46

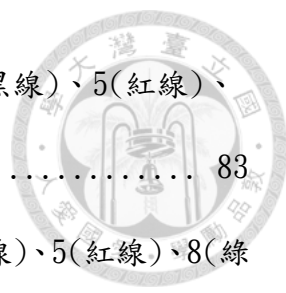
4· 估計量的統計性質	47
第四章 模擬及結果	52
1· 資料生成	52
2· 模擬結果	53
第五章 實例說明	57
第六章 討論和結論	62
參考文獻	90
附錄	105
附錄 A 甲狀腺癌預後因子研究之台大醫院研究倫理委員會通過函	105
附錄 B 無共變數下，條件存活機率 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 的推導	108
附錄 C 條件存活機率 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 的推導	117
附錄 D 輔助定理的證明	124



圖目錄



圖一	進展疾病死亡模式	28
圖二	T_M 及 T_D 在 $t > m$ 的關係圖	35
圖三	模擬資料的標誌時間為 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 及真值、估計值及 95%信賴區間。	74
圖四	甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M \leq m)$ 估計值及 95%信賴區間。	75
圖五	甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M > m)$ 估計值及 95%信賴區間。	76
圖六	甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M \leq m, T_D > m)$ 估計值及 95%信賴區間。	77
圖七	甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M > m, T_D > m)$ 估計值及 95%信賴區間。	78
圖八	甲狀腺乳突癌患者(40 歲內)的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M \leq m)$ 估計值及 95%信賴區間。	79
圖九	甲狀腺乳突癌患者(40 歲內)的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M > m)$ 估計值及 95%信賴區間。	80
圖十	甲狀腺乳突癌患者(40 歲內)的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M \leq m, T_D > m)$ 估計值及 95%信賴區間。	81
圖十一	甲狀腺乳突癌患者(40 歲內)的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t T_M > m, T_D > m)$ 估計值及 95%信賴區間。	82



圖十二 40 歲以上之甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m)$ 估計值及 95%信賴區間。 83

圖十三 40 歲以上之甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M > m)$ 估計值及 95%信賴區間。 84

圖十四 40 歲以上之甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m, T_D > m)$ 估計值及 95%信賴區間。 85

圖十五 40 歲以上之甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M > m, T_D > m)$ 估計值及 95%信賴區間。 86

圖十六 甲狀腺乳突癌患者的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier 估計量)。 87

圖十七 甲狀腺乳突癌患者(40 歲內)的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier 估計量)。 88

圖十八 甲狀腺乳突癌患者(40 歲以上)的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier 估計量)。 89



表目錄


表一 疾病發展路徑	31
表二 $S(t T_M \leq m)$ 的真值及三組右受限資料之模擬結果	65
表三 $S(t T_M > m)$ 的真值及三組右受限資料之模擬結果，第一種方法	66
表四 $S(t T_M > m)$ 的真值及三組右受限資料之模擬結果，第二種方法	67
表五 甲狀腺乳突癌(不分年紀)的 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 的估計值及 95%信賴區間	68
表六 甲狀腺乳突癌(不分年紀)的 $S(t T_M \leq m, T_D > m)$ 及 $S(t T_M > m, T_D > m)$ 的估計值及 95%信賴區間	69
表七 甲狀腺乳突癌(40 歲內)的 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 的估計值及 95%信賴區間	70
表八 甲狀腺乳突癌(40 歲內)的 $S(t T_M \leq m, T_D > m)$ 及 $S(t T_M > m, T_D > m)$ 的估計值及 95%信賴區間	71
表九 甲狀腺乳突癌(40 歲以上)的 $S(t T_M \leq m)$ 及 $S(t T_M > m)$ 的估計值及 95%信賴區間	72
表十 甲狀腺乳突癌(40 歲以上)的 $S(t T_M \leq m, T_D > m)$ 及 $S(t T_M > m, T_D > m)$ 的估計值及 95%信賴區間	73

誌謝

做為一個臨床醫師，要一方面從事門診工作，一方面要從事生物統計的研究，其實是非常困難，故若非很多人工作、學業及生活上的協助，是不可能完成我的論文，故要感謝的人太多了。

首先謝謝我臨床上的恩師—張天鈞教授，我母親是甲狀腺機能亢進症的患者，年輕時便苦於此症，還為此開了兩次刀，而張教授是我母親的主治醫師，後來我因緣際會進了台大內分泌新陳代謝科，成了張教授的學生，在其門下學習診治患者、甲狀腺超音波、細胞穿刺及細胞學等，並開始學習臨床研究。張教授每逢困難的問題，都能用簡單的方法解決，也樂意回答學生的問題，拙於問問題的我，幸得張教授的允許，得以在身旁學習，通過觀察張教授和病人的互動及患者的病歷，完成了臨床的入門。台大住院醫師訓練結束後，幸得張教授的引薦而進入遠東聯合診所，在此也感謝我的所長—黃博昭教授及所有的診所成員，讓我在工作上無後顧之憂。

在預醫所碩士班時，第一次接觸陳秀熙教授上課，就深深被陳教授的課程所吸引，當統計不再是死記的東西，而是可理解可依循的，它不再是高牆內遙不可及的，陳教授讓小池塘中的我領略了大海的廣大，於是我在陳教授的指導下完成了碩士學位，並在同年進入了生統組開始了我的博士班生活，這要感謝陳教授帶我入生統的門。



在生統組的課程中，戴政教授總是不厭其煩地從基礎教導學生，不僅是學術上的教導，也包括人生上的學問，只是戴教授的學問確實深奧，在數統上的學習，他教導了我不一樣的想法，而不是傳統上不停的演算，而在統計的學習，戴教授也是一磚一瓦地教導我從目的、資料結構、假設、模式及使用方法一步一步地架構起統計的結構。他不僅教我 A 的道理，也包括 B 的事。

最感謝的是我的指導教授—張淑惠教授，她給了我充足的空間，在我低潮時，給我驅動力可以再度前進，她總是不厭煩地等待我趕上她的進度，她的邏輯性是我學習的目標，謝謝她的耐心等待及教導，讓我終於能完成博士論文。

最後感謝我的女兒—子容，她的圖讓我的論文增色不少，僅將本論文獻於我的父母及家人—太太采芬、兒子宥銘及女兒子容。



摘要

以臨床工作而言，診療慢性患者的過程所形成的記錄，為一長期追蹤資料型態。在每次患者就診中，臨床醫師會透過一些標誌或事件訊息來了解未來患者的併發症或死亡風險的高低，根據這些風險的高低，醫師需要制定適當的措施來防制或延緩併發症甚至死亡的發生，故如何量化此類風險是臨床上很重要的問題。本論文的目的在於利用患者的動態指標訊息及基本特性來預測未來的存活機率，亦即利用標誌的訊息，不斷地更新對未來存活的預測。時間依賴性 Cox 模式 (time-dependent Cox's model) 是分析存活資料的一種常用的迴歸模式，具有利用長期追蹤資料具時間順序的特色，而明確地架構標誌歷程和基礎共變數跟終止事件間的相關性。但採用以標誌歷程為時間依賴性共變數的 Cox 模式的問題在於此時間依賴性模式為一即時解釋模式，無法用過去標誌歷程直接預測未來存活機率。因此本論文運用貝氏定理作機率的反轉及條件式機率的運算來處理這個問題，如此便可利用共變數和標誌歷程的條件式分布，在發生終止事件的時間點上給予適當的權重上 (Cox 模式中的參數及變數值所構成)，而可在不同標誌歷程及共變數的條件下，估計未來存活機率。此方法之優點在於可免於估計時間依賴性 Cox 模式中的基礎風險函數及標誌時間的邊際分布。本研究將利用模擬以測試所提出方法之表現，並以甲狀腺乳突癌患者為例加以說明所提之方法的運用。

關鍵詞: 追蹤資料、標誌歷程、雙變數存活、多變數存活、存活預測

Abstract

In clinical practice, the records of patients with chronic diseases is a form of the longitudinal data. At each patient's visit, the physician will collect the signs or event information to understand the level of the patient's future risk of complications or death. According to the level of these risks, physicians need to take some appropriate actions to prevent or delay the occurrence of complications or death. So, how to quantify such risks is a clinically important issue. The purpose of this paper is to use the dynamic messages of marker and the patients' basic characteristics to predict the patients' survival. Time-dependent Cox's model is a population regression model which constructs explicit dependence of the hazard of termination time on baseline covariates and marker process by taking the advantage of longitudinal data with chronological features. However, in the time-dependent Cox's model, the effect of the marker on the immediate survival has no meaning of prediction. That is, it is not straightforward to predict the future survival given the past information of the marker process in the time-dependent Cox's model. Therefore, we adopt Bayes' theorem and conditional probability to overcome such problems. We estimate the conditional probability of future survival given the different information of marker process by using the conditional distribution of baseline covariates and marker process given surviving at a time point and the Cox modeling information. The advantage of the proposed method is that marginal distribution of marker process and baseline hazard function in the Cox's model are not required. Simulation studies are conducted to assess the performance of the proposed method. An example of papillary thyroid carcinoma is provided for illustration.


Key Words: longitudinal data, marker process, bivariate survival,
Multivariate survival, prediction



第一章 背景和目的


在醫師的臨床工作所接觸的患者中，有些患者的疾病病程是短期而可治癒的，像是上呼吸道感染，若是患者沒有出現併發症，患者可能只有一至數次的門診。相反地，有些患者本身的疾病則需要長期的治療方可治癒，像是肺結核的治療，往往需要六個月以上的藥物治療才有較高的治癒的機會；另有些疾病則是不可治癒且其器官機能會隨時間逝去而逐漸變差，這類疾病治療的目的在於改善症狀或延緩死亡的發生，像是阿滋海默症(Alzheimer's disease)所引起的老人失智症，目前並沒有藥物可以有效治療其基本病程(Schneider, 2013)，故其治療的方針是改善生活品質或延緩疾病的進展；而有些疾病則是不可治癒但可以控制，像是高血壓、糖尿病及高血脂症等慢性病，透過藥物治療及生活方式的改變，其血壓、血糖及血脂的異常可被控制，這類疾病治療的目的在於防止其相關的併發症或死亡的發生。而在這些慢性病患者的追蹤期間，可能會出現某些臨床訊息而代表疾病的惡化，如糖尿病患者微量白蛋白尿的出現，代表的是患者腎臟功能的初期損害或心臟血管疾病的風險的增加(Zoppini et al., 2012)。這些慢性病患者的存活時間、併發症或某些臨床訊息的發生，往往是醫生或患者注意的重點，亦是長期追蹤資料需要的訊息。

慢性患者的特徵在於需要長期追蹤，而在追蹤中，根據診治醫師的要求及病情的需要，安排患者在在特定的時間會到醫師的診間或醫院，接受問診、理學檢查、抽血、影像檢查和治療。理論上，觀察的時間應是連續的(continuous)，但實際上我們卻只在離散的(discrete)時間點上擷取訊息，這些訊息包括觀察到不同的事件(event)及事件時間或生



物標誌(biological marker)。這樣的資料結構稱為長期資料(longitudinal data)或追蹤資料(follow-up data) (Zoppini et al., 2012)，慢性疾病患者的就診過程便是標準的長期資料。如果在追蹤的過程中，我們只對某特定事件有興趣，此稱為單變數存活資料(univariate survival data)，此為現行的臨床研究中最常見的資料型態;但若被觀察者可出現不同的事件或重覆出現相同的事件，不同的事件間可能存在時序間先後的關係或彼此為互斥的關係，這種資料型態稱為多變數存活資料(multivariate survival data)，其中包涵了競爭風險資料(competing risks data) (Zhang, Zhang and Fine, 2011)、多階段資料(multi-state data) (Meira-Machado et al., 2009)和復發事件資料(recurrent event data) (Zhao et al., 2012)。從長期資料的特性中，我們知道單變數存活資料若也有長期追蹤的過程，其本質上也是一種多變數存活資料，只是終止事件(terminal event)前發生的其他事件，不是我們有興趣的，像是乳癌的患者，若我們只對死亡有興趣，對於追蹤中間發生的局部復發或遠處轉移，雖然有觀察到這些事件，但不呈現這些事件的過程。也有非常單純的單變數存活資料，像是登記檔的資料，可能僅有起初登記的資料及最後死亡檔的資訊 (Bodger et al., 2011)。


而在門診中，常因口耳相傳或地緣的緣故，使得同一家庭的成員或同一公司的職員都成為同一個醫師及同一家醫院的患者。對於家庭而言，其成員具有基因及生活環境上的類似，對於公司的成員而言，他們具有工作環境上的類似，這種被觀察者間具有某種類似的性質，使得在分析上每個個體皆獨立的假設是不存在，這種在群聚中具有相關的資料可稱為平行資料(parallel data)或群集資料(cluster data) (Zhang, Fan and Sun, 2009)。由上述臨床的經驗可知，就醫行為所形成的資料，常為具有長期追蹤及群集的雙重特性之混合多變數存活資料型態。對



於臨床醫師而言，其面對的群體大多是已患病的患者，其臨床工作的目的可依疾病的性質來分，若是可治癒的病，其目的在於治癒患者的疾病，若是慢性病，其目的大多是要預防疾病的惡化所產生的併發症所導致的殘障，甚至死亡。故臨床醫師們希望在每次就診中，透過一些指標訊息或事件訊息來了解未來患者的併發症或死亡的風險高低。透過對併發症或死亡風險的了解，我們可採取一些因應措施去降低患者未來產生不良事件的風險。這些指標訊息或事件訊息包涵了患者的口述病史及症狀、理學檢查、血液檢查、影像檢查及患者在就診間隔所發生的事件記載，這樣的訊息我們可稱為標誌(markers)，而標誌所形成的歷程可稱為標誌歷程(marker processes) (Whitmore, Crowder and Lawless, 1998)。例如甲狀腺癌患者的追蹤中，醫師會檢驗甲狀腺球蛋白(thyroglobulin)、甲狀腺超音波及碘 131 攝影來發現局部復發或遠處轉移(Kucukalic-Selimovic et al., 2012)，這些檢查的結果、局部復發或遠處轉移都可做為甲狀腺癌的生物或事件標誌，因局部復發或遠處轉移這樣的事件(標誌事件)可能對於甲狀腺癌患者的最後終止事件(指死亡)具有重大的影響 (Verburg et al., 2013)，故若患者發生標誌事件時，我們必須要對患者進行更積極的治療以防止或延遲死亡的到來。

因疾病進展所導致的終止事件(死亡或某種併發症)才是臨床上所關心的重點，而臨床上常面臨的問題是假如發生了或尚未發生某種事件(標誌)，此時對患者的預後影響為何?而在不同臨床分期的患者，同一標誌事件的發生與否，是否其預後有所差別?而標誌事件的發生時間的早晚是否會影響其後續存活?而標誌事件影響存活的程度是多少?這些問題都跟患者的生活及未來息息相關，故是患者跟臨床工作者很想知道的問題。

如何能運用上述這些標誌歷程的資訊，以達到預測未來存活機



率，乃本論文之目的。本研究將在標誌歷程和其他共變數的分布並無特定的假設下，使用時間依賴性 Cox 模式(Cox, 1972)來設定風險函數(hazard function)與標誌歷程和其他共變數(covariates)間的關係，運用貝氏定理(Bayes' theorem)做機率的反轉，並利用標誌歷程和其他共變數的條件聯合分布以估計未來存活函數，而達到本研究以動態之標誌歷程預測未來存活機率的目的，此為一半母數估計法(semi-parametric estimator)。以下論文分為下列章節進行:第二章為文獻回顧，將針對多變數存活資料的分類及特性、多變數存活資料的處理及標誌歷程的運用來討論;第三章為方法，首先定義所用的資料型態及符號，後分別針對完整資料結構及右受限資料結構，利用動態的標誌歷程進行對未來存活函數的估計，此章最後為估計式的變異數的推導過程及其漸進性質的證明;第四章為模擬之過程及結果;第五章為以甲狀乳突癌為例說明方法之運用;第六章為討論和結論，主要是說明方法之特點、未來論文所要持續進行之方向，及現階段所得之結果之總結。




第二章 文獻回顧

1. 多變數存活資料的分類及特性

多變數存活資料是一種具有相關性(correlation)的存活資料，其相關性有兩種來源，一種是來自被觀察者與被觀察者間，亦即事件是發生在不同的被觀察者身上，但被觀察者與被觀察者間具有某種相同的特性，像是血緣關係或相同的生活或工作環境。相關性的另一個來源來自同一被觀察者的不同事件間的相關。根據相關性來源的不同，我們可將多變數存活資料分為平行資料和長期資料，以下就其不同的特性來分別討論：


(I) 平行資料

當相關性來自被觀察者與被觀察者間的某項類似的特性時，此時所形成的資料結構稱為平行資料或聚集資料。這種資料中，每個被觀察者都有不同的受限時間(censoring time)，此與長期資料中，同一個被觀察者的多元事件皆受同一個受限時間限制不同。平行資料常見的例子為雙胞胎研究(Twin Study) (Fischer, Harvald and Hauge, 1969)，這樣的研究設計主要是要研究雙胞胎間資料的相依性(dependence)，因同卵雙胞胎與異卵雙胞胎間，其生活方式和其他環境因子皆類似，故兩者間的差異來自基因上的差異，同卵雙胞胎的基因是相同的，而異卵雙胞胎的基因間的相異度就如同一般的兄弟姊妹一樣，兩者的基因相似只有百分之五十，故研究雙胞胎間的參數是否具有相依性，就等於研究同卵及異卵雙胞胎間基因的差異是否與我們有興趣的參數有關，亦



即若不存在相關性，則結果與遺傳無關。這類的研究的例子包括有丹麥雙胞胎研究(Hauge et al., 1968)和澳洲雙胞胎研究(Slutske et al., 2013)。類似的研究概念包括手足研究(sibling study)(Pollard et al., 2010)和親子研究(parent-child study)(Tan et al., 2007)，只是這樣的研究中仍可能存在環境因子的差異，不像雙胞胎研究單純只有基因上的差異，但其優點也在於可研究基因和環境間的交互作用。而更複雜的遺傳研究設計包括了三元體資料(Vefring et al., 2010)、四元體資料、核心家庭資料或其他家族聚集資料，這樣的家族資料主要是用來看資料間是否具有相關性，這種相關性的存在即是要討探基因和我們有興趣的結果的相關性是否存在。除了遺傳研究外，有些平行資料的研究，研究者必需將某一群被觀察者視為同一個群集(cluster)，因他們具有某種類似的特性，如同一所學校，這種資料型態必須要處理群集中的相關性以求得參數的無偏估計值。

另一個平行資料研究的重點在於估計變異數時，其值會受到資料間相關性的影響，如果不考慮相關性時，會得到有偏的(biased)變異數估計量。故在處理資料時，可先測試其相依性，若相依性不存在，則一般的獨立分析方法即可用。但若存在相關性時，一則是我們對於相關性的估計並不在意，則相關性相對於我們有興趣的參數而言，是一個累贅因子(nuisance factor)，我們可運用穩健估計量(robust estimator)去做參數的估計(Al-Khalidi et al., 2011)，而不須在意相關性的實際分布是什麼。另一方面，如果相關性的估計也是我們有興趣的，則必須同時考慮相關性及其他參數，此時可運用模式去模式化相關性及假設在模式化相關性後個體或事件間為條件獨立，再運用資料去估計相關性及所需的參數。這個方法的優點在於可處理複雜的資料，並同時得到相關性及所需的參數的估計值，但缺點是同時下了相關性的模式及條




件獨立的假設，這些假設是很強的假設，若是這些假設不符合實際的資料，則可能導致所得到的估計值是錯誤的，故這樣的方法是需要做模式診斷和驗證(model diagnosis and validation)。另一個複雜的方法，則是假設邊際(marginal)函數、相關性和其聯合函數間的關係來估計相關性及其參數。

(II) 長期資料


另一種多變數存活資料則是同一被觀察者發生多個事件，這些事件同時受同一個受限時間的限制。在單變數存活分析時，我們常假設死亡時間與受限時間之間為獨立，但在此類多變數存活資料中，若考慮間隔時間(sojourn time)時，此時，間隔時間跟受限時間並不獨立，這種相關稱為誘導式相依(reduced dependence)。這種同一被觀察者會發生多重事件的資料，因為要觀察到這些事件的發生，必需要長期的追蹤及觀察，故此類多變數存活資料被稱為長期資料或追蹤資料。

若將長期資料依發生事件是否相同來分類，當發生的事件為相同時，此時資料可稱為復發事件資料(recurrent event data)(Lim and Zhang, 2011)，復發事件為有序事件(ordered event)，因要發生下一次復發事件前，一定要發生前一次的復發事件。而在資料處理上，第一個問題仍是復發事件間是否具有相依性，在工業或日常生活上，復發事件間可能具有獨立性，例如家裏所用的日光燈的置換就可能具有獨立性。當復發事件間的相關性不存在時，則每個事件都可看成是獨立的，如一個個體有三次復發事件，則我們可將此個體的時間切成四段時間，此相當於四個獨立個體，其中有三個個體發生事件，一個未發生事件。但在醫學上，復發事件間往往是具有相關性的，亦即上一次復發往往




會增加下一次復發的風險，另復發的次數也可能會增加下一次復發的風險。若當相關性存在時，假如我們有興趣的是平均復發次數，則事件間的相關性並不是我們所關心的，此時使用穩健估計量去做參數的估計即可，不需考慮事件間的相關性。另一個方法則是將相關性設為一個隱變數(latent variable)，利用模式化相關性和事件的關係去估計我們所需要的參數，並在估計過程中，將相關性利用平均或積分的技巧去除掉。但若我們對相關性有興趣，除模式化相關性外，並將相關性放在模式中去估計參數及相關性。而在復發事件資料處理上，第二個問題是復發事件和終止事件間的關係，復發事件會受終止事件的影響而不再發生，如懷孕為復發事件，停經可視為懷孕的終止事件，懷孕會受到停經的影響，一旦停經後，懷孕便不再發生，在此例中，停經的發生並不受到懷孕的影響。另終止事件可能會受到復發事件的發生而增加風險，如腫瘤的復發會受到個體死亡的緣故而不再發生，且腫瘤的復發往往會造成個體的死亡，故當以腫瘤的復發為研究主體，此時不能將死亡視為獨立性受限(independent censoring)或無訊息性受限(uninformative censoring)，而要將之視為相依性受限(dependent censoring)或訊息性受限(informative censoring)來處理(Liu, Wolfe and Huang, 2004; Ye, Kalbfleisch and Schaubel, 2007)。在復發事件資料中，除了要處理終止事件的相依性受限，有時也要處理間隔時間與受限時間之間的誘導式相依及間隔時間與間隔時間之間的相關性。考量此類資料的複雜性，必須要同時處理復發事件的相關性和其他參數，就如同要處理平行資料一樣，必須要模式化相關性及使用條件獨立的假設。

若是長期資料中，發生事件為不同時，此時可依事件間是否有序來分類。若為有序多重事件，最簡單的是直線漸進的模式，如我們將糖尿病的病程分為正常、糖尿病前期、糖尿病、併發症及死亡五期。



對有序事件而言，每一個事件都是必須照順序，亦即發生下個事件前，必須發生前一事件，不可能跳躍過某個事件而直接進入下一個事件。如果有跳躍發生，表示事件的發生有另外的途徑，若以前面所舉的糖尿病為例，可能有某些人直接死於糖尿病，而並非經由併發症這個階段，此時糖尿病患者從糖尿病到死亡的事件的途徑便有兩條，一條不經由併發症，一條經由併發症。另外，若我們有大量資料，我們可複雜化上述路徑，我們可將糖尿病的併發症再細分為腎病變、心臟血管疾病、腦中風及下肢血管疾病等併發症，而不同的併發症可同時或不同時間發生，但若以第一個發生的併發症為考慮，則這些併發症間彼此為互斥，且相互競爭，我們可稱這些併發症為競爭風險(competing risks)，這樣去定義事件發生的次序及路徑的模式，可稱為多階段模式(multi-state models)(Meira-Machado et al., 2009)，在多階段模式中，事件的發生是一種轉換(transition)，而事件發生後，就是進入一種新的狀態(state)，而狀態的改變也就是轉換。這個模式的優點在於它定義了事件發生的路徑，如此會簡化並限制事件聯合機率(joint probabilities)化為條件機率(conditional probabilities)的過程，因當事件為有序時，下一個事件的機率必須根據上一個事件而決定，事件間的關係就有如親代和子代間的關係。但缺點在於路徑是人為定義，且多階段模式本身即為一非常強烈的假設，此假設若要跟實際的資料符合，必須在定義路徑及事件上去符合生物的特性及實際發生的路徑，一旦假設錯誤，則結果會發生極大的錯誤。若事件間為無序時，亦即在上述的多階段模式中，有多條路徑會發生，這種無序事件亦可用多階段模式來模式化，但要得到一致的(consistent)和無偏的結果，這種資料必須要很龐大。此段中所論述的長期資料，因具有多階段的特性，可稱為多階段事件資料。



另一種多變數存活資料型態為競爭風險資料(competing risk data)(Pintilie, 2007)，它指同一個體有可能歷經不同的事件，而這些事件間彼此互斥，且同一個體最終只能發生一個事件，不能經歷兩個事件，故 Hougaard (2000)雖將競爭風險資料跟其他的多變數存活資料一起討論，但他並不將競爭風險資料視為一種多變數存活資料。通常在處理競爭風險資料時，一般會將不同的競爭風險視為彼此獨立，因此在處理上，當我們在探討某一個死因時，其他的死因都被視為獨立性受限來處理，例如一個甲狀腺癌患者死於車禍時，車禍就可視為甲狀腺癌的競爭死因，且是獨立性受限的。但相同的情形發生在糖尿病患者時，就不一定了，如果他是乘客時，車禍就可視為糖尿病的競爭死因，且是獨立性受限的;但如果他是司機時，可能是發生車禍前發生了低血糖而意識不清，或發生急性心肌梗塞或急性腦中風而發生車禍，此時車禍不再與糖尿病無關，故不能當成獨立性受限。另外，我們以 B 型或 C 型肝炎為例，此兩種肝炎病毒若是在患者感染後演變成急性肝炎而沒有演變成慢性肝炎時，除非患者是得到猛暴性肝炎而導致死亡，不然他們是可痊癒的，另一種可能是他們沒有痊癒，而變成慢性肝炎。若患者走向慢性化，則臨床上，他們可能會有慢性肝炎、肝硬化及肝癌的可能性。有些患者曾經慢性肝炎、肝硬化、肝癌而死亡;有些患者只經慢性肝炎、肝硬化而死亡，則他們在因肝硬化而死亡前並未發生肝癌;有些患者則直接由慢性肝炎而直接發生肝癌而死亡。對於一個同時患有肝癌及肝硬化的患者而言，他可能因肝癌或肝硬化死亡，故肝癌或肝硬化彼此為互斥且競爭，此兩者彼此可視為 B 型或 C 型肝炎患者之競爭死因。但從上述複雜之路徑來看，肝硬化與肝癌之間存在著某種相關性，若將二者間視為獨立，則可能會導致錯誤之結論。從以上例子可知，在競爭風險資料中，多數事件皆可視為獨立，


但若在一些特殊事件中，事件間是有相關的。另外，在多階段模式中，就如上述的糖尿病併發症的例子來看，糖尿病併發症間是競爭風險，故要了解多重變數存活資料，了解競爭風險也是必要的。

最後一種多變數存活資料是半競爭風險資料(semi-competing risk data)(Fine, Jiang and Chappell, 2001)，它也可視為多階段模式中的一種特例，稱為前進式三階段資料(progressive three state data)。其特徵在於到終止事件前有二條路徑，一條不經非終止事件而直接到終止事件，一條經非終止事件而到終止事件，故為一種多階段模式，且這個資料結構中，事件的進行是前進而不後退，又有三個狀態，故稱為前進式三階段資料。若跟競爭風險資料做比較，終止事件是非終止事件的受限事件，但非終止事件卻不是終止事件的受限事件，這跟競爭風險資料中的事件互為彼此的受限事件不同，且一旦個體發生非終止事件後，往往發生終止事件的風險便會增高。

複雜的多變數存活資料可同時包括平行資料及長期資料，就如同在臨床工作中所遇到的患者一樣，且也可能同時包涵有復發事件、有序事件、競爭風險及半競爭風險事件。

2. 多變數存活資料的統計分析

從以上的多變數存活資料的介紹中，其分析方法可分為四類，其一是分析兩個事件時間或兩個個體(平行資料的兩個個體或長期資料的時間中，同一個體的兩個事件)之間的相依性;另一個是事件的邊際風險函數(marginal hazard functions)或邊際強度歷程(marginal intensity process)的分析，此分析重點並不在事件的相依性上，故將相依性當做累贅因子來處理。另一種分析方法為條件模式(conditional model)及聯



合機率模式(joint probability model)，條件模式是使用隨機變數去模式化相依性，這個不可觀察到的隨機變數稱為脆弱性(frailty)，它可用於描述單變數存活資料中未觀察到的變數或多變數存活資料中事件和事件之間或人與人間的相關性，通常我們會假設在脆弱性及其他共變數存在下，事件和事件間或人與人間是獨立的，如此便可將多變數存活資料解離成數段獨立的單變數存活資料，還有另一種是聯合機率模式，copula 模式是其中常用的一種(Othus and Li, 2010)，它運用 copula 變數去模式化不同的邊際存活函數間的相關性，而將一個多變數聯合存活函數分解成事件的邊際存活函數與 copula 變數的結合。最後一種方法則是使用多階段模式去處理多變數存活資料，因多變數存活資料大都可化為多階段模式，此模式中最常見的處理方法是使用馬可夫鏈(Markov Chain)的模式(Sweeting, Farewell and De Angelis, 2010)，此模式假設未來一刻發生事件的機率及路徑只跟前一刻的狀態有關，跟前一刻所在的狀態之前的歷史無關，故其假設也是一種條件獨立。以下將分四小節分別討論此四種方法。

(I) 相依性分析

在多變數存活資料分析中，通常只用一個參數去描述其相關性，故為總體相依性(global dependence)。若干雙變數存活時間的總體相依性估計量已經被提出，包括 Spearman's correlation coefficient，Pearson correlation coefficients、Kendall's tau(Oakes, 1982a)和中位數一致性(Median concordance)。Spearman's correlation coefficient 是用於二個可排序的變數間的相依性測量，而 Pearson correlation coefficients 多是用於二個常態分布的變數間的相依性測量，且兩者要用於完整資料

(Hougaard, 2000)，故不太適合於存活資料的分析。總體相依性是一種類獨立(quasi-independence)的測量，因當相依性為零時，表二個變數間不相關，不代表兩者為獨立，但實務上，若相依性為零時，我們將二個事件時間之間視為獨立。Kendall's tau 是目前存活分析中常用的對總體相依性的無母數估計，不僅用於二個事件時間的相依性測量，也已推廣至復發事件(Fu, Su and Chang, 2015)。

總體相依性是假設其相依性與時間無關，但是若兩個事件時間的相依性具有時間的相關性時，則總體相依性會掩蓋相依性和時間之間的關聯，Hougaard (2000)將相依性與時間之間的框架分為早/晚、立即、短期和長期的相依。當相依性具有上述的時間變化，則總體相依性用來測量相依性是不適合的。此時可用局部相依性(local dependence)或分段相依性(piece-wise dependence)來表示相依性的時間變化。

交叉比函數(cross-ratio function)是人們最常使用的局部相依性測量 (Clayton, 1978; Oakes, 1982b)，交叉比函數 θ 最初定義如下

$$\frac{\lambda_s(s_0 | t = t_0)}{\lambda_s(s_0 | t > t_0)} = \frac{\lambda_f(t_0 | s = s_0)}{\lambda_f(t_0 | s > s_0)} = \theta, \quad (2-1)$$

其意指若有父子二人， $\lambda_s(s_0 | t = t_0)$ 為子在父發病時間為 t_0 時，在時間 s_0 的風險函數， $\lambda_s(s_0 | t > t_0)$ 則為子在父發病時間大於 t_0 時，在時間 s_0 的風險函數。同理， $\lambda_f(t_0 | s = s_0)$ 為父在子發病時間為 s_0 時，在時間 t_0 的風險函數， $\lambda_f(t_0 | s > s_0)$ 則為父在子發病時間大於 s_0 時，在時間 t_0 的風險函數。這個定義具有雙向的意義，在相關性測量上，是跟方向無關，故採何種測量並無差別，但在因果關係上，卻是具有方向性的，故以父來看子是較有意義的運用。而此定義吸引的地方在於它具有良好的危險比的解釋，Day, Bryant 和 Lefkopoulou (1997)將其解釋為預測性

危險比(predictive hazard ratio)，且他們也發展出當標記時間少於終止事件時間時，若此交叉比函數 θ 為定值時的估計方法。Clayton copula 模型常用來估計這些交叉比函數，Clayton copula 模型具有以下形式

$$C(u, v) = \{u^{1-\alpha} + v^{1-\alpha} - 1\}^{1/(1-\alpha)}, \quad \alpha \geq 1 \quad (2-2)$$

u 、 v 為雙變數存活函數， α 為其 copula 變數。當雙變數存活函數屬於 Clayton copula 一族時，在(2-2)式中的 α 可表達成在(2-1)式中的 θ (Clayton, 1978)，Oakes (Oakes, 1989) 提供了另一種方法測量局部相依性的方法，他定義

$$\theta(t) = \frac{P\{(T^{(a)}, T^{(b)}) \text{ concordant} \mid \tilde{T}^{(ab)} = t\}}{P\{(T^{(a)}, T^{(b)}) \text{ disconcordant} \mid \tilde{T}^{(ab)} = t\}},$$

其中 $T^{(a)} = (T_1^{(a)}, T_2^{(a)})$ 和 $T^{(b)} = (T_1^{(b)}, T_2^{(b)})$ 為二隨機配對組，

$\tilde{T}^{(ab)} = \min(T^{(a)}, T^{(b)})$ ，則 $\theta(t)$ 為當 $\tilde{T}^{(ab)} = t$ 時，一致配對的機率對上不一致配對的機率之比率，而這個比率 $\{\theta(t) - 1\} / \{\theta(t) + 1\}$ 則是 Kendall's tau 的條件式版本。而阿基米德 copula 家族(Archimedean copula family)中的其他成員也被用來測量雙變數存活函數模型的相關性(Shih and Louis, 1995)。在交叉比和 Kendall's tau 的條件式版本的基礎上，Fan, Hsu 和 Prentice(2000)提出了兩種方法來估計在有限時間內的雙變數存活函數的相依性;Nan 等人(2006)則採用分段時間內交叉比為常數的假設去估計標記事件的年齡和停經年齡之間的關聯。而運用局部相依性的加權也可用來估計總體相依性，這方法稱為加權平均倒數交叉比(weighted average reciprocal cross-ratio) (Fan et al., 2000)，而這種總體相依性的估計的重點在於要使用何種加權函數去加權局部相依性。

由上所述，交叉比函數、Kendall's tau 和 copula 為現今測量多變數函數之相依性之主要方向。



(II) 邊際模型的分析

邊際模型主要的目的是根據依變數的邊際分布，去看其共變數對其邊際存活函數的影響，其事件或人之間的相依性並不是研究的重點 (Spiekerman and Lin, 1998)，在應用上，若是平行資料時，常使用獨立工作模式 (independence working model, IWM) 的假設來做參數的估計，其結果雖為不偏，但其對變異數的估計卻因未考量其相依性而為有偏的，故在變異數的估計時，雖不考量相依性的特定分布，卻仍將相依性納入變異數的估計中，此時對變異數常用的穩健估計量為三明治式變異數估計量 (sandwich variance estimator)。

若是復發事件資料，邊際迴歸模型已經被一些作者所採用，舉例來說，Li 和 Lagakos (1997) 在對事件定義後，將死亡視為復發事件的受限變數，及每次復發的存活時間為復發事件時間或死亡的最小時間，而採用 WLW 邊際模型 (Wei, Lin and Weissfeld, 1989) 的方法用在多變數事件存活資料。Cook 和 Lawless (1997) 利用在某一特定時間內的存活者的復發事件的平均頻率/速率的函數用來分析復發事件。另外，Ghosh 和 Lin (2000) 提出了復發事件的累計數目的邊際平均數的無母數估計法，他們 (Ghosh and Lin, 2002) 還開發了另外對於復發事件的累計數目的邊際平均數的兩種半母數法的估計方法，因估計法涉及右受限時間是累積機率，但右受限時間會有受終止時間受限的問題，一種是使用跟受限機率成反比的加權法 (inverse probability of censoring weighting, IPCW) 來消除其干擾，另一種類似的技術，是以存活機率成反比的加權法 (inverse probability of survival weighting, IPSW)。Wang 和 Chang (1999) 提出對間隔時間的無母數估計法，另外，Chang (2004) 專注在間隔時間和其共變數的關係，雖然在其模式中，有利用到脆弱性變數，但她利用期望值和 log-rank 的估計法去除脆弱性變數的影響，而得到

參數的估計值。在這些文章中，邊際模型並不特定或模式化事件之間的相依性。



(III) 條件與聯合模型

如果我們想要模式化多變數事件的聯合機率，就應將事件之間的相依性考慮在模式中。在事件風險函數的假設中，脆弱性變數模型利用脆弱性變數來指定相依性，如利用未終止事件和終止事件的脆弱性變數去指定事件間相依性，或利用復發事件間的脆弱性變數去指定復發事件間的相依性，或人和人間的脆弱性變數去指定人和人間的相依性；條件在這些脆弱性變數上，未終止事件和終止事件是獨立的，復發事件間是獨立的，人和人間是獨立的，故群集中未終止事件和終止事件風險的聯合函數，可以利用觀察到的共變數和未觀察的脆弱性變數來呈現其事件風險函數。若是共同隨機效應模型，則用一個共同脆弱性變數(shared frailty)去指定上述的相依性，而不是用多個的脆弱性變數去指定個別的相依性，在應用上，如 Lancaster 和 Intrator (1998) 利用共同隨機效應模型透過一個共同的不可測的個人所特有的脆弱性變數，去模型化愛滋病毒感染的病人的住院和存活的聯合分布。


在脆弱性變數模型(Huang and Wang, 2004; Huang and Liu, 2007; Lancaster and Intrator, 1998; Liu et al., 2004; Ye et al., 2007)中，若存在脆弱性變數下，事件歷程是一個非均質的 Poisson 過程(nonhomogeneous Poisson process)，這是此類模式在估計參數時的核心想法。而在處理脆弱性變數時，因為它為一隱性變數，是未知的，必須要像處理遺漏值(missing value)一樣，故處理方式也有多種方法。像 Liu 等人(2004)使用一個共同脆弱性變數去指定復發事件間的相依性及復發事件和死亡間的相依性，他們使用一個平均值為 1，變異數為 θ 的 gamma frailty 來描述此共同脆弱性變數，而在估計時，是使用 E-M (expectation-maximum)

的方法。而 Huang 和 Liu (2007)也是用相同的概念，只是重點在於間隔時間(gap times)。

copula 模式常用於經濟上的分析，而其模式在存活上，是結合存活時間邊際分布和 copula 變數去分析多變數存活聯合函數的一種模式 (Day et al., 1997; Oakes, 1989)。但 copula 模式是一個非常強烈的假設，在實際的資料應用上是否符合，仍是一個需要注意的重點。且 copula 模式是要用到存活時間的邊際分布，不同的存活時間之間彼此並不會被對方受限，彼此間也沒有方向性，故要應用在實際的資料應用上，要先了解 copula 模式的限制，若存活時間彼此間有受限或存活時間的邊際分布存在的估計問題的資料，若使用 copula 模式，可能不太適合。值得注意的是，在雙變數存活 (T_1, T_2) 資料中， $T_1 > T_2$ 和 $T_1 < T_2$ 都可被直接觀察到。

(IV) 多階段模式


當被觀察者在觀察期間經歷不同的階段時，多階段模式會是一個自然發展出來的模式。在多階段模式中，必須要定義狀態空間(state space)，狀態可以定義為事件的發生或某種生化值的變化，而狀態的變化是一個隨機過程(stochastic process)，而時間可以是離散的或連續的。當多階段模式用於臨床資料時，必須要定義狀態及路徑，透過事件發生的路徑的定義，可簡化並限制事件聯合機率化為條件機率的過程。當事件為有序時，我們假設下一個事件的機率只根據上一個事件而決定，這樣的假設稱為馬可夫過程(Markov process)(Wu, Yen and Chen, 2004)。如將處在狀態的間隔時間的分布考慮在上述的模式中，則可稱為半馬可夫模式(semi-Markov model)(Kang and Lagakos, 2007)。若所觀



察的事件是受看不見的狀態所支配，而狀態的變換仍依從馬可夫過程，這種模式則為隱藏式馬可夫模式(hidden Markov model)(Bartolomeo, Trerotoli and Serio, 2011; Titman and Sharples, 2010)。若將半馬可夫模式和隱藏式馬可夫模式結合而具有兩者特徵的模式稱為隱藏式半馬可夫模式(hidden semi-Markov model)(Aydin, Altunbasak and Borodovsky, 2006)。當下一個事件的機率不只根據上一個事件而決定，而是前二個事件(含)以上而決定時，則稱為高階馬可夫模式(higher order Markov model)(Seifert et al., 2012)。因馬可夫模式常涉及多個參數的估計，在樣本數有限的狀況下，常是使用母數估計法。但多階段模式的路徑為人為定義，且多階段模式本身即為一非常強烈的假設，故模式檢定(model checking)對此模式是重要的。

3 · 標誌歷程


在長期資料的研究中，我們會從個案的追蹤得到額外的訊息，此訊息涉及到個人的存活特性。此類訊息在醫學上往往可作為預後指標，我們可以稱它們為標誌。標誌可能是一個生物標誌，被觀察者目前的狀態，或多種因子(factor)的總和。標誌歷程(marker process)是一個隨機過程，它記錄著標誌在觀察時間內的變化(Whitmore et al., 1998)。標誌歷程的路徑可能會提供個案走向終止事件或其他事件的進展訊息。例如，免疫標誌可提供免疫性疾病的進展訊息，如血液中 CD4 細胞的數目，可以用在已感染愛滋病毒的人做為定期測量的工具，以提供愛滋病進展的訊息(Choi et al., 1993)，當 CD4 細胞數目少於每微升 200 個，感染者便容易出現多種伺機性感染或一些不常見的腫瘤，此時稱為後天性免疫不全症候群(acquired immunodeficiency syndrome，



AIDS)，故治療的起點可以 CD4 細胞的數目當做參考的重點。CD4 細胞數目的變化可做為一個標誌歷程，可用來預測感染愛滋病毒的人的病程。標誌也可能是追蹤觀察患有特定疾病的病人時，其在疾病的過程中多種非致命的疾病狀態的一種狀態，通常這種非致命的疾病狀態仍會影響最後的存活率。標誌過程是由被研究和觀察的個案所產生，只有在個案仍存活和未受限的情況下，它才可被觀察(Fine et al., 2001)。

(I) 標誌歷程的應用

有幾種可能的標誌歷程的統計應用。首先，它可以作為回歸模型的時間依賴性共變數。Liao 等人(2005)運用一組隨時間變化的生化指數做為共變量而形成的數值，作為小型肝細胞癌的預後指標。其次，標誌的過程可以視為輔助變量以估計邊際存活函數，Lagakos (1977)在可觀察到 $X \leq Y$ 的區域，在 (X, Y) 具有某特定分布假設下，利用標誌歷程 (X) 去改善存活函數 (Y) 的估計。第三，標誌歷程可以作為臨床試驗中的治療效果的替代終點。使用替代終點的主要動機之一是減少臨床研究的時間，如 Flandre 和 O'Quigley (1995) 在一個兩階段程序流程中，運用標誌歷程做為替代終點以減少追蹤的時間，且其結果也沒有損失太多的訊息。而在乳癌的研究中，Neri 等人(2007)發現，復發的時間似乎是遠端轉移和存活的顯著預後因素，此提供了復發的時間做為存活的替代終點的可能性。第四，標誌歷程可以用來做為在臨床試驗中的治療效果的主要結果。標誌歷程是無病存活分析(disease free survival analysis)的主要終點之一，在這種分析中，它們是終止事件的競爭風險。在一個匯集 18 個第三階段(phase III)的輔助治療結腸癌的試驗的資料分析中，相較於傳統的結果-因任何原因死亡，無病存活分析是一個合適的結果(Sargent et al., 2005)。第五，標誌歷程可以作為時間分層的依據，如在史丹佛心臟移植的研究中，心臟移植的時間被用來作為時



間分層的依據(Kalbfleisch and Prentice, 2002)。最後，標誌歷程可以作為多階段模型中的中介狀態，在以多次上消化道內視鏡檢查的結果的基礎上，從而建立起從正常到幽門螺旋桿菌感染後，到最後形成胃癌的多階段模式為例，這些上消化道內視鏡檢查的結果即是標誌歷程(Liu et al., 2006)。而上面所述，即為標誌歷程可能的實際應用。

(II) 右受限的標誌歷程

標誌歷程易受到右受限的影響，包括標誌路徑的觀察被終止事件所終止、研究終止或觀察個案隨機的失去追蹤。如果受限是由於研究終止或觀察個案隨機的失去追蹤而發生，那麼它可以看作是一個標誌事件和終止事件的獨立性受限。相比之下，當受限時間取決於未觀測到的終止事件時間或終止事件的發生，此時便出現標誌事件的相依性受限。在臨床研究中，我們通常對標誌的進程和終止事件的關係感興趣，所以標誌歷程同時受到相依性受限和獨立性受限，故在處理標誌的進程和終止事件時，必須同時處理相依性受限和獨立性受限的問題。

(III) 前進式三階段資料


標誌歷程可以轉化為事件過程，如果我們將標誌的變化定義為在時間域的事件，如此，終止事件合併標誌歷程可被視為雙變數或多變數存活資料。如我們僅考慮第一次非終止事件發生的時間，此在隨機模式(stochastic model)中被稱為首返時間(first return time)，此為標誌歷程應用上基本的模式，此模式 Fine 等人(2001)將此看作是一個競爭風險問題的一種變形，因終止事件的發生會使非終止事件的觀察停止，故為非終止事件的受限事件，但反過來，非終止事件的發生，不會讓終止事件不會發生，故不為終止事件的受限事件，此與競爭風險不同，故他們把它叫做半競爭風險資料。而另一種看法，此一進程可被看作

是一個病情惡化的自然過程中的多階段模式的一種特定變形。因疾病的侵略性質，我們可稱為前進性三階段資料。




(IV) 前進式三階段資料之統計分析

Cox (1972) 所提出之時間依賴性 Cox 模型可用來假設標誌歷程和終止事件之間的關係，這是一個合理和自然的方式來解釋標誌歷程和終止事件的模式，這可能也是目前臨床醫師較為熟悉用於解釋標誌歷程和終止事件的關係的迴歸模式(Boberg et al., 2002; Hartmann et al., 2012; Williams et al., 2010)。時間依賴性 Cox 模型從原型來看，它是利用即時的訊息去預測下一刻的風險，故為解釋，而不是預測。而 Day 等人(1997) 所提出的預測性危險比，使得標誌歷程對於終止事件的發生，不止只有具有解釋的意味，也具有預測的地位。Fine 等人(2001) 在這種資料結構中，他們所著重是在終止事件的相依性受限上，如何去做標誌歷程的標誌時間(或稱中介時間)的估計及其相依性的估計。相較於 Fine 等人使用 Clayton copula 的假設下對相依性的估計，Wang (2003)則擴大到其他 copula 的假設下，對其相依性的估計。Fine 和 Peng (2007)則將終止事件當做標誌歷程的相依性受限，其重點在於影響標誌歷程的標誌時間的共變數的係數估計，其模式為一時間相依性之 copula 模式，其估計法為一非線性估計方程式(nonlinear estimating equations)。Hsieh, Wang 和 Ding(2008)則採用 Chang (2000)的 Two-sample 的做法的概念，他們利用 copula 模式化標誌時間和終止時間的關係，在第一階段時，先估計所謂的干擾參數(nuisance parameters)，如標誌時間和終止時間的邊際分布及 copula 參數，在第二階段時，再設定標誌時間和其共參數的不同模式關係，其重點亦在於影響標誌時間的參數的係數估計，之後他們用半參數轉換模式(semiparametric transformation model)去設定標誌時間和終止時間和其



共變數間的關係，再提出了同時可估計影響標誌時間和終止時間的參數估計方法(Ding et al., 2009)，此方法是根據 log-rank 的方法。而 Chen (2012)稍後也提出了估計影響標誌時間和終止時間的參數估計方法，他對標誌時間和終止時間和其共變數間的模式也是為半參數轉換模式，而標誌時間和終止時間的聯合分布亦是用 copula 模式，但他用的是最大似法(maximum likelihood)估計。在標誌歷程的應用上，我們提到標誌歷程可以作為臨床試驗中的治療效果的替代終點，而如何評估此種替代性(surrogacy)是一個需要解決的問題，Ghosh (2009) 提出了在單一臨床實驗中，如何在這種資料結構中評估用標誌歷程做為終止事件的替代終點的方法，他對標誌時間和終止時間的模式為加速失效時間模型(accelerated failure time model)，之後他們利用同樣的模式又提出了在多個相近的臨床實驗中，如何評估用標誌歷程做為終止事件的替代終點的方法(Ghosh, Taylor and Sargent, 2012)。Lakhal, Rivest 和 Abdous(2008)提出了標誌歷程和終止事件間相依性的估計方法，並提出了標誌歷程和終止事件的聯合存活函數的估計方法，其方法的延伸固然讓標誌歷程對於終止事件的發生具有預測的地位，但其使用了 copula 的假設，在實際的資料運用上是否符合是個問題。另外，當標誌事件未發生時，標誌歷程對於終止事件的發生的預測並無估計，也缺少其他影響標誌歷程及終止事件的共變數之係數估計。

在上述的標誌歷程的統計運用中，可看到前輩們研究的方向，一個方向是對於標誌歷程和終止事件間相依性的估計，主要是使用 copula 家族的模式(Fine et al., 2001; Lakhal et al., 2008; Wang, 2003);有些人則著重在影響標誌歷程和終止事件的共變數的迴歸參數估計(Chen, 2012; Ding et al., 2009; Hsieh et al., 2008; Peng and Fine, 2007);有些人則專注在標誌歷程做為終止事件的替代終點(Ghosh, 2009; Ghosh et al., 2012);



而標誌時間分布的估計也是另一個方向(Fine et al., 2001)。若將脆弱性模型運用在此，其重點也是在影響標誌歷程和終止事件的共變數的迴歸參數估計(Chang, 2004; Huang and Wang, 2004; Huang and Liu, 2007; Liu et al., 2004)。唯終止事件往往是臨床上最重要的事件，故應是臨床研究上最重要的目標，而標誌歷程往往跟終止事件有關，使得臨床醫師在患者的追蹤中，其重點在於標誌歷程，希望通過對標誌歷程和其他共變數的了解，能夠推估未來存活機率，Lakhal 等人(2008)提出的方法(Archimedean copula)雖能估計標誌歷程和終止事件的聯合存活函數，但他們模式中並未探討其他共變數的影響。

臨床上常面臨的問題是假如發生了或尚未發生某種事件(標誌)，此將對患者的預後產生什麼影響?而在不同臨床分級的患者，標誌事件的發生與否，是否其未來預後有差別，也是很重要的課題，如乳癌一期跟三期的患者，如發生復發，兩者在存活上有沒有差異，而復發的時間的早晚是不是跟其存活有關，這些例子都是患者跟臨床工作者很想知道的問題，且影響的程度也是他們息息相關所關注的問題。而目前的研究方法，從上面的方法，可知並無法完全解決大家的需求，故本研究試圖為上述的問題，提出解決方法，希望能針對同一種疾病但不同嚴重程度的患者，在追蹤中出現或未出現某種徵兆，而能預測其後續的存活機率，如此可量化標誌事件對終止事件的影響。本研究提出了一個半參數模型，將利用某段時間內的標誌歷程的訊息，並將結合共變數的訊息，以用來估計未來某時間時終止事件的存活機率。本研究使用無偏的資料及利用時間依賴性 Cox 模型來解釋標誌歷程和終止事件的相依性結構，並推廣 Xu 和 O'Quigley (2000)的方法，利用標誌歷程和共變數的條件式分布來預測未來存活機率。




4. 不同 Cox 模式對時間依賴性變數之處理

過去使用 Cox 模式處理時間依賴性變數時，這些變數是當做不隨時間變動的變數，即在追蹤中有進行治療(如移植)或有遠處轉移者，一開始便被當作有反應者或轉移者，因為在存活分析時會產生起點的誤差，會誇大治療的結果或低估轉移的後果，這樣的問題在 1970 年代開始被重視(Gail, 1972)，而癌症治療也有相同的問題，重大事件分析(landmark analysis)便被提出來處理這個問題(Anderson, Cain and Gelber, 1983)，重大事件分析的方法是用重大事件的時間做分界，指在 Cox 模式中，分在某個特定時間內有發生重大事件的族群及沒有發生重大事件的族群，再比較兩者存活的差別，此法必須要知道在特定時間時是否已發生重大事件，故必須將觀察時間小於特定時間的個案去除，故無法充分利用資料中所有之訊息。而相較於此，時間依賴性 Cox 模式則會充分利用訊息，故有平滑(smoothing)的效果，但時間依賴性 Cox 模式為一即時解釋模式，無法用過去標誌歷程直接預測未來存活機率，故如果要用標誌歷程預測終止事件，需要其他的轉變。

5. 所用之甲狀腺乳突癌資料

甲狀腺癌為內分泌系統中發生率最高的癌症，而在台灣的甲狀腺



癌中，發生率最高的是甲狀腺乳突癌。為討探甲狀腺癌相關之預後因子，以台大醫院就診之甲狀腺癌患者為世代進行存活研究(相關台大論理委員會通過文件，見附錄 A)。此研究採醫院為基礎的回溯性研究，其研究世代的構成為在台大醫院看診之甲狀腺癌患者。首先利用疾病診斷碼 193 收集疑似甲狀腺癌的住院或門診之台大醫院個案號碼，共 3319 個疑似案例。並利用病歷回顧的方法去確定這些案例是否為甲狀腺癌患者，其中在 1985 年至 2007 年，共有 2634 個甲狀腺癌的診斷案例。而甲狀腺乳突癌的案例為 2093 個(佔 79.5%)。再去除 475 個病理報告不完整的乳突癌個案，最後留下 1636 個案例。我們收集個案的性別、出生年月日、臨床症狀、就醫時間、甲狀腺超音波的特徵、甲狀腺細針穿刺細胞學結果、開刀時間及手術發現、術後併發症、病理結果、術後放射線治療、就診時間及追蹤結果。研究的成果刊登於 *Clinical Endocrinology*(Su, Chang and Chang, 2015)。

我們以 2009 年年底為觀察終點，平均追蹤時間為 97.3 ± 69.5 個月，(中位數為 84 個月，範圍為 1-299 個月)。在這些個案中，至少 535 個案例(32.7%)在開刀時已有淋巴結轉移，另外在觀察期間，有 90 個案例(5.5%)有遠處轉移，其中 43 個在一開始的癌症分期評估中發現，而另外的 47 個則是在追蹤中出現遠處轉移。他們平均發現轉移的平均時間為 5.63 ± 5.14 年(中位數為 3.52 年)。此外，有 134 個案例(8.2%)



在追蹤期間至少有一次局部復發，而發現局部復發的平均時間為 3.89 ± 3.67 年(中位數為 3.02 年)。

在不分死亡原因的存活分析中，我們採用時間依賴性 Cox 模式來做相關預後因子的多因子分析，其結果顯示男性(風險比為 1.96，95% 信賴區間為 1.41~2.74)、較大的年紀(40-60 歲對小於 40 歲之風險比為 5.55，95% 信賴區間為 3.21~9.58 及大於 60 歲對小於 40 歲之風險比為 24.55，信賴區間為 14.10~42.74)、有遠處轉移(風險比為 6.83，95% CI 信賴區間為 4.51~10.36)、局部復發(風險比為 1.98，95% 信賴區間為 1.30~3.02)、較大的腫瘤(最大徑 > 4 公分)(風險比為 2.20，95% 信賴區間為 1.43~3.39)、甲狀腺旁組織侵犯(風險比為 1.55，95% 信賴區間為 1.11~2.16)和淋巴結轉移(風險比為 2.42，95% 信賴區間為 1.33~4.40)為其危險因子。唯一具保護作用的術後使用碘-131 做甲狀腺組織的清除，使用較高的碘-131 劑量有較佳的保護作用(使用大於 30 mCi 的劑量對不使用的風險比為 0.41，95% 信賴區間為 0.28~0.62 及小於 30 mCi 的劑量對不使用的風險比為 0.65，95% 信賴區間為 0.43~1.00)。我將以此甲狀腺乳突癌的患者為例，做為本論文之應用範例。



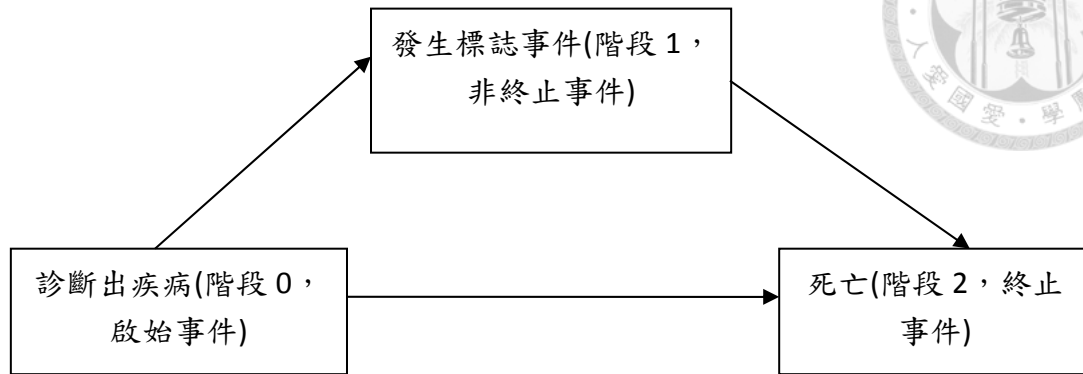
第三章·方法

方法將分四個小節進行，第一小節將先說明並定義符號，且說明並定義所使用之完整資料結構及實際上的右受限資料型態。第二小節則說明所使用之時間依賴性 Cox 模式。第三節依不同的資料型態及不同的時間依賴性 Cox 模式依序推導估計方法，並解決實際上可能遇到之右受限問題及最後存活機率不為零的問題。第四節則為此估計量之一致性(consistence)及漸進常態性(normality)的特性。

1·符號及資料結構

本研究的資料結構為含標誌事件及終止事件的多變數存活資料，為一長期追蹤資料，我們採用的原型為前進式三階段資料，或稱為半競爭風險資料，為上述資料結構之最簡單模式，以做為未來擴展至更複雜模式之基礎。進展疾病死亡模式(progressive illness death model)為前進式三階段資料典型的例子(Kalbfleisch and Prentice, 2002)，我們以甲狀腺乳突癌來說明進展疾病死亡模式，個體在診斷出疾病後進入研究，則診斷出疾病為階段 0(啟始狀態)，如甲狀腺乳突癌患者進行甲狀腺切除手術時，即可稱為階段 0，而在進入階段 0 後，個體的疾病發展過程可能為路徑一，即先發生標誌事件(階段 1，非終止狀態)，如甲狀腺乳突癌患者發生肺部轉移或局部復發，則肺部轉移或局部復發為階段 1，而進入階段 1 後再發生甲狀腺癌相關死亡(階段 2，終止狀態)；而另一疾病發展過程為路徑二，由疾病診斷(階段 0)後，不經標誌事件而直接死亡(階段 2，終止狀態)，如甲狀腺乳突癌直接侵襲氣管而造成

個體死亡。所以個體可能的疾病發展過程如圖 1。



圖一 進展疾病死亡模式

在研究進行中，有時會因為個體在研究期間退出研究、失去聯絡，或在研究終止前還沒有發生終止事件，而無法觀察到終止事件或標誌事件的發生，進而無法得到完整的存活時間或標誌時間，這種導致存活訊息不全的事件稱為右受限(right censoring)事件，而此種無法完全觀察到完整存活時間的資料為右受限資料。在我們的研究中，主要探討的是標誌過程對於終止事件發生的預測，所用之資料結構亦為右受限資料。以下先定義資料所需的符號及資料結構。

T_D ：從疾病診斷到發生終止事件的時間，為存活時間(survival time)，在 t 時間時，其機率密度函數(probability distribution function, pdf) 為 $f(t)$ ，累積分配函數(cumulative distribution function, cdf) 為 $F(t)$ ，其存活函數為 $S(t) = P(T_D > t) = 1 - F(t)$ ，另 $f(\cdot|\cdot)$ 為 T_D 的條件機率密度函數。

T_M ：從疾病診斷到發生標誌事件的時間，為標誌時間(marker time)，在 t 時間時，其機率密度函數為 $g(t)$ ，累積分配函數為 $G(t)$ ，但 T_M 會受到 T_D 的發生而受限，故上述的密度函數為一假想存在的函數，但可做為條件式機率密度函數之過渡函數， $g(\cdot|\cdot)$ 為 T_M 的條件機率密度函數，因 T_M 會受到 T_D 的發生而受限之故，故在 $T_D \geq u$ 條



件下，在 t 時間時， T_M 其條件機率密度函數為 $g(t|T_D \geq u)$ ，條件累積分配函數為 $G(t|T_D \geq u)$ 。

$\delta_G = I(T_M \leq T_D)$ ：表示有無發生標誌事件的指標函數，

$$\delta_G = \begin{cases} 1, & \text{發生標誌事件} \\ 0, & \text{沒發生標誌事件} \end{cases}。$$

Z ：疾病診斷時的共變數，可化為基礎風險分數(baseline risk score)，並為風險分組之依據， Z 的邊際機率密度函數為 $w(z)$ ，和標誌時間之聯合密度函數為 $q(t, z)$ 。而 $q(t, z)$ 的分布會受到 T_D 的發生而產生影響，故 $q(t, z)$ 也是一過渡函數，如同 T_M 的分布一樣，本文所需要的是 $q(t, z)$ 的條件式機率，故在 $T_D \geq u$ 的條件下， (T_M, Z) 的條件機率密度函數為 $q(t, z|T_D \geq u)$ ，而對於 $q(t, z|T_D \geq u)$ 的分布，並沒有去指定某種特定分布，而是用其資料的實際分布去代替其值。若每個個體皆可完整觀察到由疾病診斷到發生終止事件時，則此完整資料的結構可表示為 $\{T_D, T_M^*, \delta_G, Z\}$ 。此時由完整資料中的指標函數 δ_G ，便可以知道疾病的發展過程，若 $\delta_G = 1$ 為路徑一，若 $\delta_G = 0$ 則為路徑二。

考慮實際情形時，個體存活時間、標誌時間因右受限情況而無法完整觀察到，為解決此問題，故必須有獨立受限的假設。若右受限為獨立受限時，其意義在此為右受限時間跟個體存活時間、標誌時間和疾病診斷時的共變數無關，即 $C \perp (T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 。在描述此右受限資料時，除上述定義外，尚需定義以下符號：

C ：從疾病診斷到發生右受限事件的時間，為右受限時間(right censoring time)，累積分配函數為 $S_c(t)$ 。

$T^* = \min(T_D, C)$ ：從疾病診斷到觀察到發生終止事件或右受限事件



的時間，為觀察存活時間。

$X = \min(\bar{T}, T_M)$ ：從疾病診斷到觀察到發生標誌事件、終止事件或右受限事件的時間。

$\delta_D = I(T_D \leq C)$ ：表示有無觀察到發生終止事件的指標函數，

$$\delta_D = \begin{cases} 1, & \text{觀察到終止事件發生} \\ 0, & \text{沒觀察到終止事件發生} \end{cases}。$$

$\delta_M = I(T_M \leq \bar{T})$ ：表示觀察到發生標誌事件的指標函數，

$$\delta_M = \begin{cases} 1, & \text{觀察到標誌事件發生} \\ 0, & \text{沒觀察到標誌事件發生} \end{cases}。$$

則觀察到的右受限資料結構可表示為 $\{\bar{T}, X, \delta_D, \delta_M, Z\}$ 。此時由觀察資料中的兩個指標函數 (δ_D, δ_M) ，便可以知道疾病的發展過程， (δ_D, δ_M) 有下列 4 種情形： $(\delta_D, \delta_M) = \{(1,1), (1,0), (0,1), (0,0)\}$ ，下面分別說明此 4 種情形：

(1) $(\delta_D, \delta_M) = (1,1)$

表示疾病的發展過程為圖 1 中的路徑一，即由疾病診斷發展到標誌事件，再到終止事件，可觀察到標誌事件及終止事件的時間，即 $\bar{T} = T_D$ 及 $X = T_M$ 。

(2) $(\delta_D, \delta_M) = (1,0)$

表示疾病發展過程為圖 1 中的路徑二，在疾病診斷後，未經標誌事件而直接發展到終止事件，故只觀察到終止事件的時間，即 $\bar{T} = T_D$ 且 $X = T_D$ 。

(3) $(\delta_D, \delta_M) = (0,1)$

表示疾病發展過程也是路徑一，由疾病診斷發展到標誌事件，但在觀察到終止事件發生前先發生了右設限事件，可觀察到標誌事件及



右受限事件的時間，也就是 $\bar{T} = C$ 和 $X = T_M$ 。

(4) $(\delta_D, \delta_M) = (0, 0)$

表示在疾病診斷後，尚未發展到標誌事件或終止事件之前，已先觀察到右受限事件的發生，則我們無法得知發展過程為路徑一或路徑二，故無法判斷疾病的發展過程，只觀察到右受限事件的時間，即 $\bar{T} = C$ 和 $X = C$ 。

由上述(1)~(4)可整理成下表一。

表一 疾病發展路徑

δ_D	δ_M	\bar{T}	X	發展路徑
1	1	T_D	T_M	路徑一
1	0	T_D	$\bar{T} = T_D$	路徑二
0	1	C	T_M	路徑一
0	0	C	$\bar{T} = C$	無法判斷

接下來定義標誌事件的歷程(path)及存活時間的風險函數。當資料為完整資料時，在 t 時間的標誌狀態定義為 $m^*(t) = I(T_M < t, T_M \leq T_D)$ ，而從 0 時到 t 時間為止所對應的標誌歷史(marker history)為

$M^*(t) = \{m^*(u), 0 \leq u < t\}$ ，標誌歷史為一左連續過程。

在給定標誌歷史 $M^*(t)$ 和共變數 Z 的條件下，在 t 時間時的存活時間 T_D 的條件風險函數定義為

$$h(t | M^*(t), Z) = \lim_{dt \rightarrow 0} \frac{P(T_D \in [t, t + dt] | T_D \geq t, M^*(t), Z)}{dt} \quad (3-1)$$

當觀察資料為有右受限資料時，在 t 時間觀察到的觀察標誌狀態為 $m(t) = I(X < t, T_M \leq \bar{T})$ ，對於觀察資料 $\{\bar{T}, X, \delta_D, \delta_M, Z\}$ ，在給定觀察到 t 時

間為止的觀察標誌歷史 $M(t) = \{m(u), 0 \leq u < t\}$ 和共變數 Z 的條件下， t 時間時，觀察存活時間 \tilde{T} 的條件風險函數定義為

$$\tilde{h}(t | M(t), Z) = \lim_{dt \rightarrow 0} \frac{P(\tilde{T} \in [t, t + dt), \delta_D = 1 | \tilde{T} \geq t, M(t), Z)}{dt} \quad (3-2)$$

2. 模型

Cox 模式可用來模式化共變數和風險函數間的關係，通過時間依賴性 Cox 模式，我們可架構標誌歷程跟終止事件間的相關性。在本論文中，我們是運用時間依賴性 Cox 模式來模式化標誌歷程及基礎共變數(baseline covariates)和終止事件間的關係，假設 Z 為 $p+1$ 向量，為個案之基礎共變數， $k(T_M)$ 為 T_M 的函數，為一已知型態的標誌事件時間的函數，此為一彈性的做法，可讓模式有更大運用的空間，讓 $Z_{p+2} = (Z, m(t), K(T_M))$ 為 $(p+2) \times 1$ 向量， $B_{p+2} = (B, \beta_1, \beta_2)$ 為 $(p+2) \times 1$ 向量， B 為基礎共變數之係數， β_1 為標誌狀態之係數，而 β_2 為 $k(T_M)$ 之係數，當資料為完整資料時，以 Cox 模式表示標誌事件和終止事件的關係時，考慮以下二種情境：

第一種情境，考慮終止事件的發生，只跟標誌事件相關，此時的 Cox 模式設定為

$$\begin{aligned} h(t | Z, M^*(t)) &\equiv h(t | Z, m^*(t), T_M) \\ &\equiv h_0(t) \exp\{\beta_1 m^*(t) + \beta_2 m^*(t) k(T_M)\} \end{aligned} \quad (3-3)$$

此模式為一時間依賴性 Cox 模式(time-dependent Cox's model)，它界定了終止事件的立即發生風險跟標誌事件是否發生及發生的時間的關係，此兩者的關係是即時的，在此模式中，標誌事件是否發生及發生的時間解釋了下一刻終止事件的立即發生風險，故不具預測的意義，



只能解釋二者之相關程度。(3-3)式因忽略了共變數的影響，故會簡化算式，可做為推導之原型，但在實際的運用上，此假設不易達到，故可能會導致有偏的估計。

第二種情境，如我們認為終止事件的發生，跟標誌事件和共變數皆相關，此時我們所設定的 Cox 模式為下式：

$$\begin{aligned} h(t|Z, M^*(t)) &\equiv h(t|Z, m^*(t), T_M) \\ &\equiv h_0(t) \exp\{B^T Z + \beta_1 m^*(t) + \beta_2 m^*(t)k(T_M)\}, \end{aligned} \quad (3-4)$$

此式在推導上較(3-3)式為複雜，但在實際運用上，應較為正確的，因此式考慮共變數的影響，將可能影響終止事件的因素皆考慮在內，也因為此式運用上之複雜性，乃能創造本文之獨創性及實用性，此模式在推導上，將會依不同的共變數及標誌歷程的分布做適當的權重，此和上述(3-3)式每個個案的權重皆相同不同。

如實際觀察資料為右受限資料時，則在 t 時間時，如上述(3-3)和(3-4)式之 Cox 模式分別設定為

$$\begin{aligned} \tilde{h}(t|Z, M(t)) &\equiv \tilde{h}(t|Z, m(t), T_M) \\ &\equiv \tilde{h}_0(t) \exp\{\beta_1 m(t) + \beta_2 m(t)k(T_M)\} \end{aligned} \quad (3-5)$$

和

$$\begin{aligned} \tilde{h}(t|Z, M(t)) &\equiv \tilde{h}(t|Z, m(t), T_M) \\ &\equiv \tilde{h}_0(t) \exp\{B^T Z + \beta_1 m(t) + \beta_2 m(t)k(T_M)\}. \end{aligned} \quad (3-6)$$

根據上述模式來利用標誌事件和基礎共變數的訊息推估未來存活機率，亦即估計 $S(t|T_M \leq m)$ 和 $S(t|T_M > m)$ 之條件存活機率，當共變數 (Z) 為類別變數時，令其全域為 H_T ，則其全域 (H_T) 為所有不同的條件的基礎共變數所形成的子集合， H 為不同的條件的基礎共變數所形成的子集合。當 Z 為連續變數時，如風險分數，其值域為連續的，如則其全域



為所有可能數之所形成的範圍，令其全域為 H_T ，為 $[-b, a]$ ，其意乃其範圍為有界的(bounded)。故以下的重點為如何利用透過標誌事件及基礎共變數的訊息來得到此估計式。

3 · 估計

以下將針對右受限資料下，在有基礎共變數的情況下，推導 $S(t|T_M \leq m)$ 和 $S(t|T_M > m)$ 等條件存活機率的估計式：

(I) 以右受限資料估計以標誌歷程為條件的未來存活函數

以下考慮，當 $t > m$ 時，利用標誌歷程來預測未來存活函數：

$S(t|T_M \leq m)$ 和 $S(t|T_M > m)$ 只有在個案仍存活和未受限的情況下，標誌事件才可能被觀察，一旦個案的終止事件發生，則標誌歷程亦即終止，此意即 Fine 等人(2001)所述之在半競爭風險資料中， (T_M, T_D) 雙變數聯合存活函數只有在終止事件時間大於標誌事件時間之象限可被估計之意。在完整資料中， $T_M \leq T_D$ 的比例是可估算的，故 $T_M > T_D$ 的比例也是可被估算的。

在推導 $S(t|T_M \leq m)$ 和 $S(t|T_M > m)$ 之前，先要推導分母 $P(T_M \leq m)$ 及 $P(T_M > m)$ 。圖二為了解 T_M 及 T_D 在 $t > m$ 的關係之重要圖形，從圖中可看出：

區域 1 的面積之機率： $P(T_M \leq m, T_D \geq t)$;

區域 1+2 的面積之機率： $P(T_M \leq m, T_D > m)$

$$= P(T_M \leq m, T_D > m | T_M \leq T_D) P(T_M \leq T_D);$$

區域 3 的面積之機率，即 m 時間內觀察標誌事件之機率：

$$P(T_M \leq m, T_D \geq T_M, T_D \leq m)$$

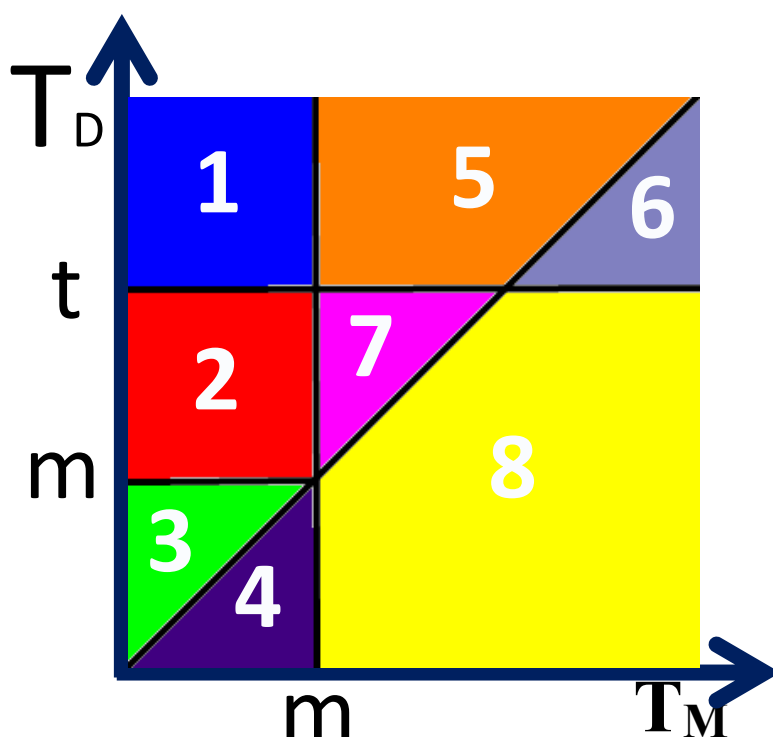


$$= P(T_M \leq m, T_D \leq m | T_M \leq T_D) P(T_M \leq T_D);$$

區域 4 的面積之機率，即 m 時間內無法觀察標誌事件之機率：

$$P(T_M \leq m, T_D < T_M) = P(T_M \leq m | T_M > T_D) P(T_M > T_D);$$

區域 1+2+3+4 的面積之機率： $P(T_M \leq m)$;



圖二 T_M 及 T_D 在 $t > m$ 的關係圖

區域 5 的面積之機率： $P(T_D \geq T_M > m, T_D > t)$;

區域 6 的面積之機率： $P(T_M > m, T_D > t, T_D < T_M)$;

區域 5+7 的面積之機率： $P(T_M > m | T_M \leq T_D) P(T_M \leq T_D)$;

區域 5+6 面積之機率： $P(T_M > m, T_D > t)$;

區域 6+8 面積之機率： $P(T_M > m | T_M > T_D) P(T_M > T_D)$

$$= P(T_M > T_D) - P(T_M \leq m | T_M > T_D) P(T_M > T_D);$$



區域 4+6+8 面積之機率： $P(T_D < T_M)$;

區域 5+6+7+8 面積之機率： $P(T_M > m)$ 。

在此種資料結構中， $P(T_M \leq m)$ 及 $P(T_M > m)$ 會有兩種路徑，參考圖二及附錄 B，兩者形式如下：

$$\begin{aligned} & P(T_M \leq m) \\ &= P(T_M \leq m, T_M \leq T_D) + P(T_M \leq m, T_M > T_D) \\ &= P(T_M \leq m, T_D > m, T_M \leq T_D) + P(T_M \leq m, T_D \leq m, T_M \leq T_D) + P(T_M \leq m, T_M > T_D); \end{aligned} \quad (3-7)$$

$$\begin{aligned} & P(T_M > m) \\ &= P(T_M > m, T_M \leq T_D) + P(T_M > m, T_M > T_D) \\ &= P(T_M > m, T_M \leq T_D) + 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M > T_D)。 \end{aligned} \quad (3-8)$$

由(3-7)式可知，在計算 $P(T_M \leq m)$ 時，在此資料結構中，僅只觀察到 $P(T_M \leq m, T_M \leq T_D)$ 部份，若用無母數法計算，並無法充分使用所有的訊息而需要大量的資料才能估算，但因在早期標誌時間內的個案數往往很少，故通過模式化，可充分運用訊息並可計算

$P(T_M \leq m, T_M > T_D)$ ，如此便可有效率法來計算 $P(T_M \leq m)$ ，故必須使用模式的假設去計算 $P(T_M \leq m, T_M > T_D)$ ，另 $P(T_M \leq m, T_M \leq T_D)$ 及

$P(T_M \leq m, T_M > T_D)$ 將以 T_D 的值做切割，因 $P(T_M \leq m, T_M \leq T_D, T_D = u)$

在不同的 T_D 時，才能提供我們所需要的訊息。另要在另一個切點

($T_D = m$) 做切割，因我們使用時間依賴性 Cox 模式會受到切點

($T_D = m$) 的影響而使得模式有所不同。在 $t \leq m$ 時，是不須考慮標誌事

件對終止事件的影響，但當 $t > m$ 時，便要將標誌事件納入 Cox 模式。

$$\begin{aligned} P(T_M \leq m, T_M \leq T_D) &= \int_0^{\infty} P(T_M \leq m, T_M \leq T_D, T_D = u) du \\ &= \int_0^m P(T_M \leq u, T_D = u) du + \int_m^{\infty} P(T_M \leq m, T_D = u) du, \end{aligned}$$



$$\begin{aligned}
 P(T_M \leq m, T_M > T_D) &= \int_0^m P(T_M \leq m, T_M > T_D, T_D = u) du \\
 &= \int_0^m P(u < T_M \leq m, T_D = u) du, \tag{3-9}
 \end{aligned}$$

當 $t > m$ 時，利用上述式子，可得

$$\begin{aligned}
 &S(t|T_M \leq m) \\
 &= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z|T_D = u) dz dF(u)}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, Z = z|T_D = u) dz dF(u) + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z|T_D = u) dz dF(u)} \tag{3-10}
 \end{aligned}$$

為估計(3-10)式，如何在適當的切點做切割仍是個重點，而要計算上式，要先計算給予 $T_D = u$ ， (T_M, Z) 的條件機率密度函數 $q(r, z|T_D = u)$ ，我們需要估計 $P(T_M \leq m|T_D = u)$ 、 $P(u < T_M \leq m|T_D = u)$ 和 $P(T_M \leq u|T_D = u)$ 。而為了得到這三個值，我們則要計算要先計算給予 $T_D = u$ ， (T_M, Z) 的條件機率密度函數 $q(r, z|T_D = u)$ ：

$$\begin{aligned}
 q(r, z|T_D = u) &= \frac{f(u|T_M = r, Z = z)q(r, z)}{f(u)} \\
 &= \frac{f(u|T_M = r, Z = z)q(r, z)}{\int_0^\infty \int_{Z \in H_T} f(u|T_M = v, Z = s)q(v, s) ds dv} \tag{3-11}
 \end{aligned}$$

上述式子中之 $q(r, z|T_D = u)$ 為一假想過渡機率函數，此邊際函數在實際上是否可計算的是有爭議的(Lakhali et al., 2008)，但在以下的算式中，它將會被運用貝式定理而取代成一實際存在的條件機率函數，另外還會再導入(3-4)式的 Cox 模式中風險函數和其共變數的假設進入算式中。

$$\begin{aligned}
 &h(t|Z, M^*(t)) \equiv h(t|Z, m^*(t), T_M) \\
 &\equiv h_0(t) \exp\{B^T Z + \beta_1 m^*(t) + \beta_2 m^*(t)k(T_M)\} \circ
 \end{aligned}$$

另對於 $f(u)$ 而言，會有兩個可能路徑，一個是經由標誌事件，此



路徑要受 $T_M \leq T_D$ 的限制，另一個未經由標誌事件，二者受惠於Cox模式的導入，將可獲得運用資料的好處，故 $f(u)$ 會有兩個部份，則(3-11)式可寫為

$$q(r, z | T_D = u) = \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)]q(r, z | T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)]q(v, s | T_D \geq u)dsdv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s)q(v, s | T_D \geq u)dsdv} \quad (3-12)$$

(3-12)式為本文中重要的式子(詳細見附錄C)，貝式定理的運用是畫龍點睛的重要一步。運用(3-12)式，在適當的 T_M 範圍對 $q(r, z | T_D = u)$ 積分便可得到(3-10)式中分子和分母所需要的 $P(T_M \leq m | T_D = u)$ 、 $P(u < T_M \leq m | T_D = u)$ 和 $P(T_M \leq u | T_D = u)$ ，如此在 $T_D = u$ 的條件下，將 $T_M \leq m$ 和 $Z \in H_T$ (H_T 為共變數之全域)之聯合機率 $P(T_M \leq m | T_D = u)$ 化為不同範圍的 $P(T_M \leq m, Z = z | T_D = u)$ 、 $P(u < T_M \leq m, Z = z | T_D = u)$ 、 $P(T_M \leq u, Z = z | T_D = u)$ 等三個機率的總和。即可用來求取 $S(t | T_M \leq m)$ ，其過程如下：

$$S(t | T_M \leq m) = \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u)dzdF(u)}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u)dzdF(u) + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u)dzdF(u)}$$

其中上式分子為

$$P(T_M \leq m, T_D \geq t) = \int_t^\infty \int_{Z \in H_T} \frac{\int_0^m \exp[B^T z + \beta_1 + \beta_2 k(r)]q(r, z | T_D \geq u)dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)]q(v, s | T_D \geq u)dsdv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s)q(v, s | T_D \geq u)dsdv} dzdF(u),$$

而其分母分為兩部份，表示如下：

$$P(T_D \leq m, T_M \leq m)$$

$$\begin{aligned}
&= \int_0^m \int_{Z \in H_T} \frac{\int_0^u \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr + \int_u^m \exp(B^T z) q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u), \\
&\quad P(T_D > m, T_M \leq m) \\
&= \int_m^\infty \int_{Z \in H_T} \frac{\int_0^m \exp[B^T s + \beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u).
\end{aligned}$$

實際資料為右受限資料，故以下要解決右受限問題。我們假設右受限為獨立性受限，此乃實際資料在運用上之最重要假設，此假設讓實際上觀察到的資料能做為完整資料之參數估計，而不會有偏差。在運用右受限資料之標誌歷程來預測未來的存活函數前，我們先要有以下的輔助定理，即(3-2)式和(3-1)式相等：

輔助定理 1。若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則 $\tilde{h}(t | M(t), Z) = h(t | M^*(t), Z)$ 。(證明附於附錄 D)

在右受限資料中，重點在於利用觀察的資料去求取(3-12)式的 $q(r, z | T_D \geq u)$ ，以下為所需要的另二個輔助定理。

輔助定理 2 若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則

$$P(X \leq t, \delta_M = 1 | \tilde{T} \geq t) = P(T_M \leq t | T_D \geq t) = G(t | T_D \geq t);$$

$$P(X > t, \delta_M = 1 | \tilde{T} \geq t) = P(T_M > t, T_D \geq T_M | T_D \geq t)。$$

(證明附於附錄 D)。

輔助定理 3 若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則

$$q(X = r, \delta_M = 1, Z = z | \tilde{T} \geq u) = q(r, z | T_D \geq u)，$$

(證明附於附錄 D)。

根據輔助定理 2 及輔助定理 3，我們可用 $\hat{P}(X \leq t, \delta_M = 1 | \tilde{T} \geq t)$ 來估計

$G(t | T_D \geq t)$ ，用 $\hat{g}(X = r, \delta_M = 1 | \tilde{T} \geq u)$ 來估計 $g(r | T_D \geq u)$ ，用



$\hat{q}(X = r, \delta_M = 1, Z = z | \tilde{T} \geq u)$ 來估計 $q(r, z | T_D \geq u)$ 。如此，我們可用實際資料來估計其機率，但是在右受限的影響下，無法直接以經驗估計方式估計 $P(u \leq T_M \leq m | T_D \geq u)$ ，因為

$$\begin{aligned} & P(u \leq X_i \leq m, \tilde{T}_i \geq u, \delta_{M_i} = 1) \\ &= E \left\{ E \left[I(u \leq X_i \leq m, \tilde{T}_i \geq u, \delta_{M_i} = 1) | T_{M_i}, T_{D_i} \right] \right\} \\ &= E \left\{ I(u \leq T_{M_i} \leq m, T_{D_i} \geq u) E \left[I(C_i \geq \max(T_{M_i}, u)) | T_{M_i}, T_{D_i} \right] \right\} \\ &= E \left\{ I(u \leq T_{M_i} \leq m, T_{D_i} \geq u) \cdot S_c(T_{M_i}) \right\}; \end{aligned}$$

與 C 的存活函數有關，由於 $P(\tilde{T}_i \geq u) = P(T_{D_i} \geq u) S_c(u)$ ，因此

$$\begin{aligned} & E \left\{ I(u \leq X_i \leq m, \tilde{T}_i \geq u, \delta_{M_i} = 1) \cdot S_c(u) / S_c(T_{M_i}) \right\} \\ &= E \left\{ I(u \leq T_{M_i} \leq m, T_{D_i} \geq u) \cdot S_c(u) \right\} \\ &= P(u \leq T_{M_i} \leq m, T_{D_i} \geq u) S_c(u), \end{aligned}$$

則

$$\begin{aligned} & P(u \leq T_{M_i} \leq m | T_{D_i} \geq u) \\ &= E \left\{ I(u \leq X_i \leq m, \tilde{T}_i \geq u, \delta_{M_i} = 1) \cdot S_c(u) / S_c(T_{M_i}) \right\} / P(\tilde{T}_i \geq u). \end{aligned}$$

根據上式，可得 $P(u \leq T_M \leq m | T_D \geq u)$ 的估計式為：

$$\begin{aligned} & \hat{P}(u \leq T_M \leq m, T_D \geq T_M | T_D \geq u) \\ &= \sum_{i=1}^n \left\{ (I(u \leq X_i \leq m, \tilde{T}_i \geq u, \delta_{M_i} = 1) / n) \cdot \hat{S}_c(u) / \hat{S}_c(\tilde{T}_{M_i}) / (\sum_{i=1}^n I(\tilde{T}_i \geq u) / n) \right\}. \end{aligned}$$

而其他的式子的條件機率，於分母和分子都是 $C > u$ 的受限條件，可同時抵消。

在右受限資料下，根據前面的推導，仍可以得到

$$\hat{S}(t | T_M \leq m) = \frac{p\hat{C}_1}{p\hat{D}_1 + p\hat{C}_2},$$



其中 $p\hat{C}_1$ 為估計的圖二中的區域 1 面積之機率，即

$$\begin{aligned} p\hat{C}_1 &= \hat{P}(T_M \leq m, T_D \geq t) \\ &= \sum_{i=1}^n I(\tilde{T}_i \geq t) \frac{\sum_{j=1}^n p\hat{C}_{1ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}(\tilde{T}_i), \end{aligned}$$

上式中，

$$p\hat{C}_{1ij} = \exp(B^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(X_j \leq m, \bar{F}_j \geq \tilde{T}_i, \delta_{M_j} = 1),$$

$$p\hat{C}_{2ij} = \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(X_j \leq \tilde{T}_i, \tilde{T}_j \geq \tilde{T}_i, \delta_{M_j} = 1),$$

$$p\hat{D}_{1ij} = \exp(\hat{B}^T Z_j) I(X_i \geq \tilde{T}_i, \bar{F}_j \geq \tilde{T}_i),$$

且 $\hat{F}(\cdot)$ 為 T_D 的累積分布函數之 Kaplan-Meier 估計函數。

$p\hat{C}_2$ 是估計圖二中的區域 1+2 面積之機率，即

$$\begin{aligned} p\hat{C}_2 &= \hat{P}(T_M \leq m, T_D \geq m) \\ &= \sum_{i=1}^n I(\tilde{T}_i \geq m) \frac{\sum_{j=1}^n p\hat{C}_{1ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}(\tilde{T}_i). \end{aligned}$$

$p\hat{D}_1$ 代表的圖二中的區域 3+4 面積之機率，則

$$\begin{aligned} p\hat{D}_1 &= \hat{P}(T_M \leq m, T_D \leq m) \\ &= \sum_{i=1}^n I(T_{D_i} < m) \frac{\sum_{j=1}^n p\hat{C}_{3ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}(\tilde{T}_i), \end{aligned}$$

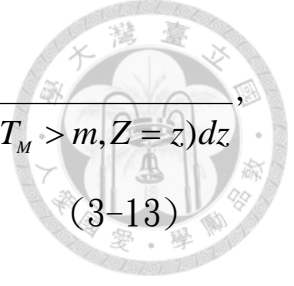
其中

$$p\hat{C}_{3ij} = \exp(B^T Z_j + \hat{\beta}_2 + \hat{\beta}_2 k(T_{M_j})) I(X_j \leq \tilde{T}_i, \tilde{T}_j \geq \tilde{T}_i, \delta_{M_j} = 1)$$

$$+ \exp(B^T Z_j) I(\tilde{T}_i < X_j \leq m, \bar{F}_j \geq \tilde{T}_i, \delta_{M_j} = 1) \bar{S}_c(\tilde{T}_i) / \bar{S}_c(X_j)$$

如要得 $S(t|T_M > m)$ ，則如同 $S(t|T_M \leq m)$ 的過程。

$$S(t|T_M > m)$$



$$= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M > m, Z = z | T_D = u) dz dF(u)}{\int_m^\infty \int_{Z \in H_T} P(u \geq T_M > m, Z = z | T_D = u) dz dF(u) + \int_{Z \in H_T} P(T_M > T_D, T_M > m, Z = z) dz} \quad (3-13)$$

上式(3-13)的分子為

$$\int_t^\infty \int_{Z \in H_T} \left\{ \frac{\int_m^u \exp[B^T z + \beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr + \int_u^\infty \exp(B^T z) q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} \right\} dz dF(u)^\circ$$

上式(3-13)的分母分為兩部分，其中

$$\begin{aligned} & \int_m^\infty \int_{Z \in H_T} P(u \geq T_M > m, Z = z | T_D = u) dz dF(u) \\ &= \int_m^\infty \int_{Z \in H_T} \frac{\int_m^u \exp[B^T z + \beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u); \\ & \int_{Z \in H_T} P(T_M > T_D, T_M > m, Z = z) dz \\ &= \int_{Z \in H_T} (w(z) - P(T_M \leq T_D, Z = z) - P(T_M \leq m, T_M > T_D, Z = z)) dz \circ \end{aligned}$$

接續上述的推導，運用類似的方法，可以得到右受限資料下 $S(t | T_M > m)$

的估計如下：

$$\hat{S}(t | T_M > m) = \frac{p\hat{C}_1^\#}{p\hat{C}_2^\# + p\hat{D}_1^\#} ,$$

其中 $p\hat{C}_1^\#$ 代表的圖二中的區域 5+6 面積之機率，即

$$\begin{aligned} p\hat{C}_1^\# &= \hat{P}(T_M > m, T_D \geq t) \\ &= \sum_{i=1}^n I(\tilde{T}_i \geq t) \left[\frac{\sum_{j=1}^n p\hat{C}_{1ij}^{**} + \sum_{j=1}^n p\hat{D}_{1ij}^{**}}{\sum_{j=1}^n p\hat{C}_{2ij}^{**} + \sum_{j=1}^n p\hat{D}_{2ij}^{**}} \right] d\hat{F}(\tilde{T}_i) , \end{aligned}$$

其中

$$p\hat{C}_{ij}^{**} = \exp(B^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(m \leq X_j \leq \tilde{T}_i, \tilde{T}_j \geq \tilde{T}_i, \delta_{M_j} = 1) ,$$



$$p\hat{D}_{1ij}^{**} = \exp(B^T Z_j) I(\bar{F}_j \geq \tilde{T}_i, X_j \geq \tilde{T}_i),$$

$$p\hat{C}_{2ij}^{**} = \exp(\hat{B}^T s + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(X_j \leq \tilde{T}_i, \bar{F}_j \geq \tilde{T}_i, \delta_{M_j} = 1),$$

$$p\hat{D}_{2ij}^{**} = \exp(\hat{B}^T s) I(X_j \geq \tilde{T}_i, \bar{F}_j \geq \tilde{T}_i),$$

$p\hat{C}_2^\#$ 則為估計圖二中的區域 5+7 面積之機率，故

$$\begin{aligned} p\hat{C}_2^\# &= \hat{P}(T_M > m, T_D \geq T_M) \\ &= \sum_{i=1}^n I(\tilde{T}_i \geq m) \left[\frac{\{\sum_{j=1}^n p\hat{C}_{1ij}^{**}\}}{\sum_{j=1}^n p\hat{C}_{2ij}^{**} + \sum_{j=1}^n p\hat{D}_{2ij}^{**}} \right] d\hat{F}(\tilde{T}_i), \end{aligned}$$

$p\hat{D}_1^\#$ 則為圖二中的區域 6+8 面積之機率，則

$$\begin{aligned} p\hat{D}_1^\# &= \hat{P}(T_M > m, T_D < T_M) \\ &= 1 - \frac{\sum_{i=1}^n I(X_i \leq \tilde{T}_i, \delta_{M_i} = 1) / \hat{S}_c(X_i)}{n} - \sum_{i=1}^n I(\tilde{T}_i \leq m) \frac{\sum_{j=1}^n p\hat{C}_{4ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}(\tilde{T}_i), \end{aligned} \quad (3-14)$$

其中

$$p\hat{C}_{4ij} = \exp(B^T Z_j) I(\tilde{T}_i < X_j \leq m, \bar{F}_j \geq \tilde{T}_i)。$$

估計 $P(T_M > m)$ 的第二種方法為利用 $1 - P(T_M \leq m)$ 來估計。

$1 - P(T_M \leq T_D)$ 的估計在右受限的資料中是另一個問題，在實際運用上，可利用如(3-14)式中的跟受限機率成反比的加權法(IPCW)來估計。上述的結果可運用在最大存活時間為有限時。

在右受限觀察資料時，在會受 $\sum_{i=1}^n dF(\tilde{T}_i)$ 是否為 1 的影響，上述的結果可用於 $\sum_{i=1}^n dF(\tilde{T}_i) = 1$ 時或最大存活時間為有限時，以下將針對 $\sum_{i=1}^n dF(\tilde{T}_i)$ 小於 1 來討論。

讓 $0 < t_1 < t_2 < \dots < t_k$ 為其 k 個發生終止事件的時間，當最後的被觀

察者不是發生終止事件時，便會發生 $\sum_{i=1}^k d\hat{F}(t_i) < 1$ 的狀況，另 m 要小於最大可觀察死亡時間，即 t_k 時，根據全機率法則(the law of total probability)，在運算時，要將機率分為 $\sum_{i=1}^k d\hat{F}(t_i)$ 及 $1 - \sum_{i=1}^k d\hat{F}(t_i)$ 兩部份，其運算如下

$$\begin{aligned} & \hat{S}(t|T_M \leq m) \\ &= \frac{\sum_{t_i \geq t} \hat{P}(T_M \leq m|T_D = t_i)d\hat{F}(t_i) + \hat{P}(T_M \leq m|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))}{\sum_{t_i \leq m} \hat{P}(T_M \leq m|T_D = t_i)d\hat{F}(t_i) + \sum_{t_i > m} \hat{P}(T_M \leq m|T_D = t_i)d\hat{F}(t_i) + \hat{P}(T_M \leq m|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))} \end{aligned}$$

其中上式分子跟上述(3-10)式的估計式比較，只差要多計算分子和分母中的 $\hat{P}(T_M \leq m|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))$ ，而

$$\hat{P}(T_M \leq m|T_D > t_k) = \frac{\sum_{i=1}^n I(\tilde{T}_i > t_k, X_i \leq m)}{\sum_{i=1}^n I(\tilde{T}_i > t_k)}。$$

同樣地，

$$\begin{aligned} & \hat{S}(t|T_M > m) \\ &= \frac{\sum_{t_i \geq t} \hat{P}(T_M > m|T_D = t_i)d\hat{F}(t_i) + \hat{P}(T_M > m|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))}{\sum_{t_i > m} \hat{P}(T_M > m, T_M < T_D|T_D = t_i)d\hat{F}(t_i) + \hat{P}(T_M > m, T_M < T_D|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i)) + 1 - \hat{P}(T_M \leq T_D) - P(T_M < m, T_M > T_D)} \end{aligned}$$

此式跟(3-13)式的估計式比較，差別在於分子中的

$\hat{P}(T_M > m|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))$ 跟分母中的

$\hat{P}(T_M > m, T_M \leq T_D|T_D > t_k)(1 - \sum_{i=1}^k d\hat{F}(t_i))$ ，其中

$$\hat{P}(T_M > m|T_D > t_k) = \frac{\sum_{i=1}^n I(\tilde{T}_i > t_k, X_i > m)}{\sum_{i=1}^n I(\tilde{T}_i > t_k)}，$$

$$\hat{P}(T_M > m, T_M \leq T_D|T_D > t_k) = \frac{\sum_{i=1}^n I(\tilde{T}_i > t_k, X_i > m, \delta_{M_i} = 1)}{\sum_{i=1}^n I(\tilde{T}_i > t_k)}。$$

結合上述的結果，便可解決右受限資料及 $\sum_{i=1}^k d\hat{F}(t_i) < 1$ 的問題，而能運用在實際的資料上。

類似的利用標誌歷程去預測未來存活函數的方法，可見於 Lakhali 等人(2008)、Porta 等人(2012)和 Maugeun 等人(2013)的文章中。而他們的差別在於 Lakhali 等人(2008)是利用 Archimedean copula 的模式，且他們所在意的是在此模式假設下，標誌時間的估計及未發生標誌事件時，未來發生終止事件的機率，對於某時間內已發生標誌事件時，對未來發生終止事件的影響，則沒有著墨。而 Porta 等人(2012)則是利用多階段模式及 Cox 模式來做預測，其文章前段有提到基礎共變數在每個階段對應變數的影響，但在做標誌事件對終止事件的預測時，並未對基礎共變數做處理。而 Maugeun 等人(2013)則是利用聯合脆弱性模型及 Cox 模式做預測，他們利用腫瘤復發的次數來預測死亡的機率，其優點在於利用脆弱性變數去除復發和死亡間的相關性，而彼此間為條件式獨立。但其模式仍可見其只能針對某特定的基礎共變數來做預測。可見在上述模式中，要將基礎共變數及標誌歷程的變化做未來終止事件的預測是有難度的，原因在於必須分別模式化基礎共變數與各階段或各事件間的關係。而我們採用的時間依賴性 Cox 模式除了將標誌歷程當時間依賴性變數而架構其和終止事件的相關性，也可加入基礎共變數於模式中，而同時處理基礎共變數對其的影響。

當 Z 為類別變數時，求取某群具有某種特質的患者之以標誌歷程為條件的未來存活函數，即 $S(t|T_M > m, Z = z)$ 和 $S(t|T_M \leq m, Z = z)$ ，為上述之特例。

在本節中，對於有基礎共變數影響下，利用右受限資料下之標誌歷程去預測未來的存活函數的過程中，此乃利用 Cox 模式之優點，因可同時處理共變數和標誌歷程與終止事件間的相關性。此為優於其他

方法(Lakhal et al., 2008; Mauguén et al., 2013; Porta et al., 2012)之處，因他們必須分別模式化共變數和各階段或各事件間的相關性。

本論文的目的，即在於利用共變數和標誌歷程的分布，給予適當的權重，而能預測未來終止事件的發生，上述的式子中充分運用了資料中所有的訊息，也提供了當個案事件數過少(如重大事件分析會去除觀察時間短於特定時間)而產生不穩定的估計的解決方法，而達到平滑的目的。

(II) 以存活者的後續有限觀察時間條件存活機率之預測

對存活者的後續有限觀察時間條件存活機率的預測，此時被觀察者至少要存活到 m 時間，在臨床上之意為患者在就診時仍續存活，而在就診時，根據原有之資訊及現在標誌的狀態去預測未來患者的存活機率，而在統計上即求取 $S(t|T_M \leq m, T_D \geq m)$ 和 $S(t|T_M > m, T_D \geq m)$ 。先計算 $S(t|T_M \leq m, T_D \geq m)$ 如下：

$$\begin{aligned} S(t|T_M \leq m, T_D \geq m) &= \frac{P(T_D > t, T_M \leq m)}{P(T_M \leq m, T_D \geq m)} \\ &= \frac{\int_t^{\infty} P(T_M \leq m | T_D = u) dF(u)}{\int_m^{\infty} P(T_M \leq m | T_D = u) dF(u)} \end{aligned}$$

從圖二來說明，即分子為圖二中之區域 1 面積機率，而分母為圖二中之區域 1+2 面積機率。跟(3-10)式的估計式比較，上式分子跟其分子是相同的，而上式分母跟分子型態相同，只是積分的範圍不同，分子是積分 $[t, \infty)$ 的範圍，而分母則是積分 $[m, \infty)$ 的範圍。同樣地，要計算 $S(t|T_M > m, T_D \geq m)$ 如下：

$$S(t|T_M > m, T_D \geq m) = \frac{P(T_M > m, T_D > t)}{P(T_M > m, T_D \geq m)}$$



$$= \frac{\int_t^\infty P(T_M > m | T_D = u) dF(u)}{\int_m^\infty P(T_M > m | T_D = u) dF(u)}。$$

從圖二來說明，即分子為圖二中之區域 5+6 面積機率，而分母為圖二中之區域 7 面積機率及部分的區域 8 面積機率。跟(3-20)式比較，上式分子跟(3-14)式分子是相同的，而上式分母跟分子型態相同，只是積分的範圍不同，分子是積分 $[t, \infty)$ 的範圍，而分母則是積分 $[m, \infty)$ 的範圍。兩者皆可從前述的推導中得到計算式。如遇到 $\sum_{i=1}^k d\hat{F}(t_i) < 1$ 的狀況，處理方法也同前一節。

4. 估計量的統計性質

本節將探討在右受限資料且有共變數 Z 下， $\hat{S}(t|T_M \leq m)$ $= \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S}))$ 的漸進統計性質。假設每個被觀察體間彼此獨立且具相同分布(independently and identically distributed, iid)，也假設 C 的分布函數等於 1 的支持域最小數值較 T_D 及 T_M 的的分布函數對應之支持域最小數值為大。

首先，假設 B, β 's, $S(t) = (1 - F(t))$ 和 $S_c(t)$ 真值為已知的情況下，估計量 $\hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S))$ 根據上節可表示如下，

$$\hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S)) = \frac{p\hat{C}_1(B, \beta, S)}{p\hat{D}_1(B, \beta, S_c, S) + p\hat{C}_2(B, \beta, S)}$$

在此， $p\hat{C}_1(B, \beta, S)$ ， $p\hat{D}_1(B, \beta, S_c, S)$ 和 $p\hat{C}_2(B, \beta, S)$ 分別為機率

$$pC_1 = P(T_M \leq m, T_D \geq t) \text{、} pD_1 = P(T_M \leq m, T_D \leq m) \text{ 及}$$



$pC_2 = P(T_M \leq m, T_D \geq m)$ 的估計量，所以，當 $p\hat{C}_1(B, \beta, S)$ ， $p\hat{D}_1(B, \beta, S, S_c)$ 和 $p\hat{C}_2(B, \beta, S)$ 具一致性(consistency)時，即可得證 $\hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S))$ 為 $S(pC_1(B, \beta, S), pD_1(B, \beta, S, S_c), pC_2(B, \beta, S))$ 的一致估計量。因為

$$p\hat{C}_1(B, \beta, S) = \sum_{i=1}^n \frac{n^{-1} \sum_{j=1}^n p\hat{C}_{1ij}(B, \beta) \cdot I(\tilde{T}_i \geq t)}{n^{-1} \sum_{j=1}^n p\hat{C}_{2ij}(B, \beta) + n^{-1} \sum_{j=1}^n p\hat{D}_{1ij}(B, \beta)} dF(\tilde{T}_i)$$

為以下三種 iid 指標函數平均之組合，後再以 $dF(t) = S(t^-) - S(t)$ 加權

$$\text{組合， } p\hat{C}_{1ij}(B, \beta) = \exp(B^T Z_j + \beta_1 + \beta_2 k(T_{M_j})) I(X_j \leq m, \bar{F}_j \geq \tilde{T}_i, \delta_{M_j} = 1),$$

$$p\hat{C}_{2ij}(B, \beta) = \exp(B^T Z_j + \beta_1 + \beta_2 k(T_{M_j})) I(X_j \leq \tilde{T}_i, \tilde{T}_j \geq \tilde{T}_i, \delta_{M_j} = 1),$$

$$p\hat{D}_{1ij}(B, \beta) = \exp(B^T Z_j) I(X_i \geq \tilde{T}_i, \bar{F}_j \geq \tilde{T}_i)。$$

故，由大數法則，當 $n \rightarrow \infty$ 時，可以得到 $p\hat{C}_1(B, \beta, S)$ 為 $P(T_M \leq m, T_D \geq t)$ 的一致估計量。同理可得證， $p\hat{C}_2(B, \beta, S)$ 的一致性質。另外，

$$p\hat{D}_1(B, \beta, S_c, S) = \sum_{i=1}^n \frac{n^{-1} \sum_{j=1}^n p\hat{C}_{3ij}(B, \beta, S_c) \cdot I(\tilde{T}_i < m)}{n^{-1} \sum_{j=1}^n p\hat{C}_{2ij}(B, \beta) + n^{-1} \sum_{j=1}^n p\hat{D}_{1ij}(B, \beta)} dF(\tilde{T}_i)$$

其中

$$p\hat{C}_{3ij}(B, \beta, S_c) = \exp(B^T Z_j + \beta_1 + \beta_2 k(T_{M_j})) I(X_j \leq \tilde{T}_i, \tilde{T}_j \geq \tilde{T}_i, \delta_{M_j} = 1)$$

$$+ \exp(B^T Z_j) I(\tilde{T}_i < X_j \leq m, \bar{F}_j \geq \tilde{T}_i, \delta_{M_j} = 1) \cdot S_c(\tilde{T}_i) / S_c(X_j)$$

也是一 iid 指標函數，同理，仍可以得到 $p\hat{D}_1(B, \beta, S_c, S)$ 的一致性質。

因 $p\hat{C}_1(B, \beta, S)$ ， $p\hat{D}_1(B, \beta, S, S_c)$ 和 $p\hat{C}_2(B, \beta, S)$ 具一致性(consistency)

時，故 $\hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S}))$ 為

$S(pC_1(B, \beta, S), pD_1(B, \beta, S, S_c), pC_2(B, \beta, S))$ 的一致估計量。

接著，探討

$\sqrt{n} \left\{ \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S})) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\}$
 的漸進常態分布的性質。首先，假設 B, β 's, $S(t)(=1-F(t))$ 和 $S_c(t)$ 真值

為已知的情況下，當 $n \rightarrow \infty$ 時，

$\sqrt{n} \left\{ \hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S)) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\}$
 近似於下列之線性組合：

$$J_s(B, \beta, S, S_c) \cdot \mathbf{W}_n,$$

在此

$$J_s(B, \beta, S, S_c) = \left(\frac{\partial S(pC_1, pC_2, pD_1)}{\partial pC_1}, \frac{\partial S(pC_1, pC_2, pD_1)}{\partial pD_1}, \frac{\partial S(pC_1, pC_2, pD_1)}{\partial pC_2} \right)$$

$$\text{且 } \mathbf{W}_n = \begin{pmatrix} \sqrt{n} (p\hat{C}_1(B, \beta, S) - pC_1(B, \beta, S)) \\ \sqrt{n} (p\hat{D}_1(B, \beta, S, S_c) - pD_1(B, \beta, S, S_c)) \\ \sqrt{n} (p\hat{C}_2(B, \beta, S) - pC_2(B, \beta, S)) \end{pmatrix}.$$

因此，只要證得 \mathbf{W}_n 的分布近似三變量常態分布，其平均向量為零向量且變異數-共變數矩陣 $\Sigma_w(B, \beta, S, S_c)$ 為一正定矩陣(positive-definite matrix)，由於 \mathbf{W}_n 為四種 iid 指標函數平均之組合，再以

$dF(t) = S(t^-) - S(t)$ 加權組合，因此根據中央極限定理，可以得到 \mathbf{W}_n 的漸進常態分布性質，則可以得到

$\sqrt{n} \left\{ \hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S)) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\}$
 的漸進常態分布的大樣本性質，其漸進變異數-共變數矩陣為
 $(J_s \Sigma_w (J_s)')(B, \beta, S, S_c)$ 。

當 B, β 's, $S(t)(=1-F(t))$ 和 $S_c(t)$ 真值為未知的情況下，則

$$\begin{aligned} & \sqrt{n} \left\{ \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S})) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\} \\ &= \sqrt{n} \left\{ \hat{S}(p\hat{C}_1(B, \beta, S), p\hat{D}_1(B, \beta, S, S_c), p\hat{C}_2(B, \beta, S)) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\} \\ &+ \sqrt{n} \left\{ \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S})) - S(p\hat{C}_1(B, \beta, S), p\hat{C}_2(B, \beta, S, S_c), p\hat{D}_1(B, \beta, S)) \right\} \end{aligned}$$

$$+\sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),pD_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),pC_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},S),p\hat{C}_2(\hat{B},\hat{\beta},S,S_c),pD_1(\hat{B},\hat{\beta},S))\right\}.$$

其中從上述中可得第一項為

$$\sqrt{n}\left\{\hat{S}(p\hat{C}_1(B,\beta,S),p\hat{D}_1(B,\beta,S,S_c),p\hat{C}_2(B,\beta,S)) - S(pC_1(B,\beta,S),pC_2(B,\beta,S,S_c),pD_1(B,\beta,S))\right\} \\ \approx J_s(B,\beta,S,S_c) \cdot \mathbf{W}_n,$$

第二項利用泰勒展式為

$$\sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},S),p\hat{D}_1(\hat{B},\hat{\beta},S,S_c),pC_2(\hat{B},\hat{\beta},S)) - S(p\hat{C}_1(B,\beta,S),p\hat{C}_2(B,\beta,S,S_c),pD_1(B,\beta,S))\right\} \\ \approx J_s(B,\beta,S,S_c) \cdot \mathbf{L}' \cdot \mathbf{V}_n,$$

其中 \mathbf{L} 為一 2×3 正定矩陣：

$$\mathbf{L} = \begin{pmatrix} \frac{\partial pC_1}{\partial B} & \frac{\partial pD_1}{\partial B} & \frac{\partial pC_2}{\partial B} \\ \frac{\partial pC_1}{\partial \beta} & \frac{\partial pD_1}{\partial \beta} & \frac{\partial pC_2}{\partial \beta} \end{pmatrix}$$

而且， \mathbf{V}_n 向量之轉置向量為 $(\mathbf{V}_n)' = (\sqrt{n}(\hat{B} - B), \sqrt{n}(\hat{\beta} - \beta))$ ，

第三項再分為兩項

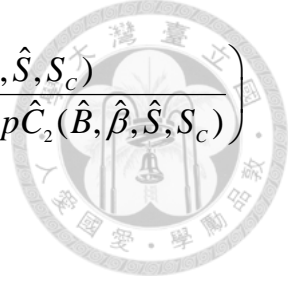
$$\sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),pD_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),p\hat{C}_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},S),pC_2(\hat{B},\hat{\beta},S,S_c),pD_1(\hat{B},\hat{\beta},S))\right\} \\ = \sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),p\hat{D}_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),p\hat{C}_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},S),pC_2(\hat{B},\hat{\beta},S,S_c),pD_1(\hat{B},\hat{\beta},S))\right\} \\ + \sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),p\hat{D}_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),p\hat{C}_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),pC_2(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),pD_1(\hat{B},\hat{\beta},\hat{S}))\right\},$$

其中 \hat{B} 、 $\hat{\beta}$ 為 B 、 β 的一致估計量，則可以得到以下近似式

$$\sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),p\hat{D}_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),p\hat{C}_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},S),pC_2(\hat{B},\hat{\beta},S,S_c),pD_1(\hat{B},\hat{\beta},S))\right\} \\ = \sqrt{n}\left(\frac{p\hat{C}_1(\hat{B},\hat{\beta},\hat{S},S_c)}{p\hat{D}_1(\hat{B},\hat{\beta},\hat{S},S_c) + p\hat{C}_2(\hat{B},\hat{\beta},\hat{S},S_c)} - \frac{p\hat{C}_1(\hat{B},\hat{\beta},S,S_c)}{p\hat{D}_1(\hat{B},\hat{\beta},S,S_c) + p\hat{C}_2(\hat{B},\hat{\beta},S,S_c)}\right) \\ \approx \sqrt{n}J_s(B,\beta,S,S_c) \cdot \mathbf{U}_n$$

以及

$$\sqrt{n}\left\{\hat{S}(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),p\hat{D}_1(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),p\hat{C}_2(\hat{B},\hat{\beta},\hat{S})) - S(p\hat{C}_1(\hat{B},\hat{\beta},\hat{S}),pC_2(\hat{B},\hat{\beta},\hat{S},\hat{S}_c),pD_1(\hat{B},\hat{\beta},\hat{S}))\right\}$$



$$= \sqrt{n} \left(\frac{p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c)}{p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c) + p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c)} - \frac{p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}, S_c)}{p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, S_c) + p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S}, S_c)} \right) \\ \approx \sqrt{n} J_s(B, \beta, S, S_c) \cdot S_n \circ$$

其中，

$$U_n = \begin{pmatrix} \sqrt{n} (pC_1(B, \beta, \hat{S}) - pC_1(B, \beta, S)) \\ \sqrt{n} (pD_1(B, \beta, \hat{S}, \hat{S}_c) - pD_1(B, \beta, S, \hat{S}_c)) \\ \sqrt{n} (pC_2(B, \beta, \hat{S}) - pC_2(B, \beta, S)) \end{pmatrix} \\ S_n = \begin{pmatrix} 0 \\ \sqrt{n} (pD_1(B, \beta, S, \hat{S}_c) - pD_1(B, \beta, S, S_c)) \\ 0 \end{pmatrix} \circ$$

由於 U_n 和 S_n 可應用 martingale 的技巧推導得到其近似常態分布性質 (如 Fine and Gray, 1999)， W_n 與 V_n 之近似分布為多變量常態分布，而且

$$\sqrt{n} \left\{ \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S})) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\}$$

近似於此線性組合 $J_s \cdot W_n + J_s \cdot U_n + J_s \cdot S_n + J_s \cdot L' \cdot V_n$ ，所以，根據 martingale 中央極限定理 (Kalbfleisch and Prentice, 2002)，可以得證

$$\sqrt{n} \left\{ \hat{S}(p\hat{C}_1(\hat{B}, \hat{\beta}, \hat{S}), p\hat{D}_1(\hat{B}, \hat{\beta}, \hat{S}, \hat{S}_c), p\hat{C}_2(\hat{B}, \hat{\beta}, \hat{S})) - S(pC_1(B, \beta, S), pC_2(B, \beta, S, S_c), pD_1(B, \beta, S)) \right\} \\ = \sqrt{n} \left\{ \hat{S}(t|T_M \leq m) - S(t|T_M \leq m) \right\}$$

具有近似常態分布之統計性質，但是該漸進變異數相當複雜，因此本研究採用 bootstrap 抽樣方法得到其變異數的估計值。



第四章 模擬及結果

為了解估計量在右受限資料及完整資料間的差異，必須模擬生成右受限資料及完整資料，並計算兩者的偏誤(bias)，另一方面，如要了解估計量的一致性、漸進常態性，可模擬生成不同的樣本數及模擬次數的右受限資料，透過其估計量的分佈的比較，便可了解。以下分成二節說明，第一節資料為模擬資料生成之步驟，第二節則是模擬之結果。

1. 資料生成

在產生資料時，主要是要生成共變數、標誌時間、存活時間、及右受限時間，在根據這些生成的資料做成觀察資料，產生資料的步驟如下：

1. 假設共變數(Z ，性別)為二項分佈，男性($Z = 1$)佔 $1/4$ ，方法為先以均勻分布(0, 1)產生一變數，若此變數小於 $1/4$ 為男性，若不小於 $1/4$ 則為女性($Z = 0$)。
2. 假設標誌時間 T_M 服從指數分佈，設定條件為 $T_M \sim \exp(\lambda_M)$ ， $\lambda_M = 0.0125 \exp(Z)$ ，方法為先以均勻分布(0, 1)產生一變數(uT_M)，則 $T_M = -\ln(uT_M) / \lambda_M$ 。
3. 假設存活時間 T_D 服從混合指數分佈(mixed exponential distribution)(Lagakos, 1976)，根據(3-6)式 Cox 模式的假設，在 u 時



間的標誌過程的情況下，假設存活時間的風險函數為 $\tilde{h}(t|Z, M(t)) \equiv \lambda_D \exp\{B^T Z + \beta_1 m(t)\}$ ，令 $\lambda_D = \exp(0.05)$ 、 $B = 0.7$ 及 $\beta_1 = 1.5$ 。即未發生標誌事件前之存活時間的風險函數為 $\exp(0.05)\exp\{0.7^* Z\}$ ；而發生標誌事件後之存活時間的風險函數為 $\exp(0.05)\exp\{0.7^* Z + 1.5\}$ 。在上述假設下，發生標誌事件之機率決定於標誌事件及發生路徑二之終止事件兩者之速率，亦兩者可視為競爭之關係(Lagakos, 1976)。根據標誌事件時間及發生路徑二之終止事件時間的大小，便可決定路徑及是否發生標誌事件，若標誌事件大於發生路徑二之終止事件時間，則為路徑二，即標誌事件不發生；相反地，則標誌事件發生(Austin, 2012)。方法為先以均勻分布 $(0, 1)$ 產生一變數 (uT_D) ，再比較 $\ln(uT_D) + \exp(0.05)\exp(0.7^* Z)T_M$ 是否小於零，若小於零，為標誌事件發生，存活時間為 $T_M - \{\ln(uT_D) + \exp(0.05)\exp(0.7^* Z)T_M\} / \{\exp(0.05)\exp(0.7^* Z + 1.5)\}$ ，反之存活時間則為 $-\ln(uT_D) / \{\exp(0.05)\exp(0.7^* Z + 1.5)\}$ 。

4. 假設受限時間 C 服從指數分佈，設定條件為 $C \sim \exp(0.02)$ ，方法為先以均勻分布 $(0, 1)$ 產生一變數 (uC) ，則 $C = -\ln(uC) / 0.02$ 。
5. 比較 T_D 、 T_D 及 C 的大小，便可產生觀察存活時間、觀察標誌時間及各項指標。

2. 模擬結果

由第一節產生資料的步驟，分別產生了 1000 組樣本數為 500，750



及 1000 等三組右受限資料，而真值則為 5000 次樣本數為 1000 所組成的無受限資料的分佈所產生，真值之性別變項之係數平均值為 0.702，真值為 0.7，而標誌事件之係數平均值為 1.505，真值為 1.5，產生標誌事件之比例為 0.213，真值為 0.213，兩者相近。

1000 組樣本數為 500 的右受限資料，其性別變項之係數平均值為 0.700(變異數為 0.0139)，標誌事件之係數平均值為 1.512(變異數為 0.0139)，而產生標誌事件之比例依跟受限機率成反比的加權法(IPCW)處理右受限後為 0.213(變異數為 0.000486)，跟真值(0.213)相似，而受限比率為 0.226。1000 組樣本數為 750 的右受限資料，其性別變項之係數平均值為 0.702(變異數為 0.0099)，標誌事件之係數平均值為 1.512(變異數為 0.0126)，變異數因樣本數增加而下降，而產生標誌事件之比例依 IPCW 處理受限後為 0.212(變異數為 0.000321)，跟真值(0.213)相似，而受限比率為 0.227。而 1000 組樣本數為 1000 的右受限資料，其性別變項之係數平均值為 0.702(變異數為 0.0074)，標誌事件之係數平均值為 1.504(變異數為 0.0085)，變異數因樣本數增加而為三組資料中最低，而產生標誌事件之比例依 IPCW 處理受限後為 0.212(變異數為 0.000225)，跟真值(0.213)相似，而受限比率為 0.225。

表二為 $S(t|T_M \leq m)$ 的真值及三組右受限資料之模擬結果，右受限資料跟真值之偏誤(bias)除了在標誌時間為 2 及 3 時預測標誌事件後數




年較差(最大為 $-1.61*10^{-2}$)，但會隨樣本數增加而下降。 $S(t|T_M \leq m)$ 的標準差的估計乃是一種 Monte Carlo 方法，因為有 1000 組樣本所組成的分佈，自然可計算其變異數，在模擬中所求得的標準差皆為此法。而 $S(t|T_M \leq m)$ 的標準差數較大(最大為 0.127)，會隨樣本數增加而下降，此兩者的現象乃受 $T_M \leq m$ 的個數過少之故，這在實例中也有相同的狀況。

表三為 $S(t|T_M > m)$ 的真值及三組右受限資料之模擬結果，此為估計 $S(t|T_M > m)$ 的第一種方法，乃利用

$P(T_M > m) = 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M \leq T_D)$ 來估計。右受限資料跟真值之偏誤(bias)在樣本數為 500 時，可見到偏誤已很小(最大為 $1.18*10^{-3}$)，可見利用 IPCW 及運用條件式機率的方法能有效去除獨立右受限的干擾。另一個結果為標準差很小(最大為 0.026)，此跟表二不同，此因 $T_M > m$ 的個數夠多之故，故 $S(t|T_M \leq m)$ 跟 $S(t|T_M > m)$ 的標準差跟其樣本數有關，為一抵換(trade-off)。

表四為估計 $S(t|T_M > m)$ 的第二種方法的模擬結果，此方法乃利用 $P(T_M > m) = 1 - P(T_M \leq m)$ 來估計。右受限資料跟真值之偏誤(bias)在樣本數為 500 時，可見到偏誤已很小(最大為 $1.56*10^{-3}$)，此跟表三相同。另一個結果為標準差很小(最大為 0.025)，從偏誤跟標準差來看，這二個方法在模擬上是沒有差別，但在實例上，第二種方法可能較佳，因



$P(T_M \leq m)$ 的估計是以有觀察到標誌事件的個案估計，而在短時間，受右受限的干擾可能較少，則估計 $P(T_M \leq m)$ 在實際情況下可能偏誤較小，故以下實例在估計 $S(t|T_M > m)$ 時是用第二種方法，而在估計 $S(t|T_M > m, T_D > m)$ 因估計量的特性則是採用第一種方法。

圖三為以樣本數為 1000 的右受限資料為例，其中黑線為標誌時間為 2 年內時，依序紅線為 5 年內及綠線為 8 年內的後十年之存活曲線。因估計值跟真值間的偏誤很少，故兩者從圖三中是無法區別。從圖三可看到 $S(t|T_M > m)$ 跟 $S(t|T_M \leq m)$ 存在極大的差異，故透過本研究中方法可運用標誌歷程及基礎共變數來預測未來終止事件的機率。下一章將運用此方法於甲狀腺乳突癌的患者的資料，以了解遠處轉移或局部復發對其未來存活的影響。



第五章 實例說明


首先以 1636 個患者為例，其中性別(男性比女性)、腫瘤大小(小於 4 公分比上大於 4 公分及未知大小)、年紀(40 歲以上比 40 歲內)、有無使用碘-131 做甲狀腺組織的清除及有無甲狀腺旁組織侵犯做為基礎變數，而遠處轉移或局部復發則為標誌事件。其係數值分別為 0.505(變異數為 0.0379)、0.775(變異數為 0.0335)、2.221(變異數為 0.0738)、-0.509(變異數為 0.0363)及 0.578(變異數為 0.0362)，而標誌事件之係數為 1.948(變異數為 0.0453)。這 1636 個甲狀腺乳突癌患者(不分年紀)的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及 95%信賴區間列於表五。 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的變異數仍依 Bootstrap 方法求得，其方法為產生 1000 組相同樣本數(抽到收回)之分佈，通過此 1000 組樣本便可求得變異數，實例的變異數皆依此法。圖四則為標誌時間小於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線，而圖五則為標誌時間大於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線。圖四中，紅線與綠線交叉，這跟模擬中的圖形不同，顯示綠線預後不佳，這可能跟標誌時間在 6-8 年間的患者的預後不佳有關，圖四的 95%信賴區間隨時間增加而增加且甚寬，這甚寬的信賴區間有賴樣本數的增加而改善。而圖五則跟模擬中的圖形的排列類似，可能這群患



者並無其他干擾因子存在，其 95%信賴區間也隨時間增加而增加。

臨床上， $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 較 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 較實用，其意為患者就診時，通過對標誌歷程的了解而能預測患者的未來存活的機率。表六為甲狀腺乳突癌患者(不分年紀)的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 的估計值及 95%信賴區間，而圖六是標誌時間為小於 2(黑線)、5(紅線)及 8(綠線)年的 $S(t|T_M \leq m, T_D > m)$ 的圖形，圖七則是相對應的 $S(t|T_M > m, T_D > m)$ 的圖形。從圖六及圖七可看出甲狀腺乳突癌是一種預後很好的癌症，甚至 $S(t|T_M \leq m, T_D > m)$ 在十年後仍有 0.6 以上，亦即標誌事件在 8 年內發生者，如能活過 8 年者，其在第 18 年存活之機率達六成之上。而標誌事件未在某特定年內發生者，其預後更佳，其特定年後第十年的存活機率皆大於 0.85，這也看出標誌事件(遠處轉移或局部復發)對甲狀腺乳突癌患者的存活影響很大。另圖六中，紅線與綠線近於交叉，顯示綠線預後不佳，這可能跟標誌時間在 6-8 年間的患者的預後不佳有關，此跟圖四一致。

在上述的例子，可看到年紀對甲狀腺乳突癌患者的預後影響很大，故利用分組的概念將年紀分成 40 歲內跟 40 歲以上來分析。先看 40 歲內的分析，共有 778 個個案，其中性別(男性比女性)、腫瘤大小(小於 4 公分比上大於 4 公分及未知大小)、有無使用碘-131 做甲狀腺組織



的清除及有無甲狀腺旁組織侵犯做為基礎共變數，而遠處轉移或局部復發則為標誌事件，此時年紀就不是基礎共變數。其係數值分別為 1.129(變異數為 0.274)、0.108(變異數為 0.587)、-0.0824(變異數為 0.285)及 0.201(變異數為 0.309)，而標誌事件之係數為 0.896(變異數為 0.479)。這 778 個年紀在 40 歲內的甲狀腺乳突癌患者的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及 95%信賴區間列於表七，圖八則是其標誌時間不大於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線，而圖九則為標誌時間大於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線。從圖八及圖九可看到此群患者預後相當好，且標誌事件對終止事件的影響較小。


表八為甲狀腺乳突癌患者(年紀在 40 歲內)的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 的估計值及 95%信賴區間，而圖十是標誌時間為不大於 2(黑線)、5(紅線)及 8(綠線)年的 $S(t|T_M \leq m, T_D > m)$ 的圖形，圖十一則是相對應的 $S(t|T_M > m, T_D > m)$ 的圖形。雖然這群患者預後相當好，甚至標誌事件為 8 年的 $S(t|T_M \leq m, T_D > m)$ 在十年後仍有 0.88 以上，亦即標誌事件在 8 年內發生者，如能活過 8 年者，其在第 18 年存活之機率達八成八之上。而標誌事件未在某特定年發生者，其預後更佳，其特定年後第十年的存活機率皆大於 0.94。當此群患者活過六年以上，其存活率下降，這可能跟競爭死因有關。



接下看 40 歲以上的 858 個個案，為做比較，性別(男性比女性)、腫瘤大小(小於 4 公分比上大於 4 公分及未知大小)、有無使用碘-131 做甲狀腺組織的清除及有無甲狀腺旁組織侵犯為基礎變數，而遠處轉移或局部復發則為標誌事件。其係數值分別為 0.429(變異數為 0.0464)、0.851(變異數為 0.0386)、-0.577(變異數為 0.0407)及 0.594(變異數為 0.0446)，而標誌事件之係數為 2.076(變異數為 0.0521)。這 858 個年紀在 40 歲以上的甲狀腺乳突癌患者的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及 95%信賴區間列於表九，圖十二則是其標誌時間小於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線，而圖十三則為標誌時間大於 2(黑線)、5(紅線)及 8(綠線)年後十年之預測存活曲線。其中可看到在圖十二中，三條線有交叉，若以紅線(5 年)為比較基點，此意謂標誌事件(遠處轉移或局部復發)小於 2 年的預後是好的，而在 6-8 年發生標誌事件的預後甚差。

表十為甲狀腺乳突癌患者(年紀在 40 歲以上)的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 的估計值及 95%信賴區間，而圖十四是標誌時間為小於 2(黑線)、5(紅線)及 8(綠線)年的 $S(t|T_M \leq m, T_D > m)$ 的圖形，圖十五則是相對應的 $S(t|T_M > m, T_D > m)$ 的圖形。雖然這群患者比起年紀在 40 歲內的患者的預後就比較差，以標誌事件為 8 年內的

$S(t|T_M \leq m, T_D > m)$ 在十年後只有 0.224，亦即標誌事件在 8 年內發生



者，如能活過 8 年者，其在第 18 年存活之機率只有二成二左右，若跟 5 年內的 $S(t|T_M \leq m, T_D > m)$ 比較，應是 7-8 年中發生標誌事件者差後非常差。而標誌事件未在某特定年發生者，其預後較一般癌症為佳，其特定年後第十年的存活機率皆大於 0.75。但比較年紀在 40 歲內的患者，則預後變差。

圖十六至十八為以 Kaplan-Meier 法估計發生標誌事件後及未發生標誌事件之存活機率，從圖十六可看出轉移或復發對甲狀腺癌患者之存活影響很大，尤其是年紀在 40 歲以上的族群(圖十八)，年紀在 40 歲內族群(圖十七)則在後期才有差別。如用此法估計存活者的後續有限觀察時間條件存活機率，則必須去除相當多的個案，此致此估計值將相對不穩定，而本文之方法可充分運用訊息而達到平滑的估計值。



第六章 討論和結論

本研究運用 IPCW 的方法處理右受限資料，從模擬的結果來看，其結果甚佳，故大多數的模擬中的估計值跟真值中的偏誤很小。而在實例中， $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 提供了以標誌事件預測終止事件的效力，而 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 則可提供臨床醫師運用標誌事件來預測未來終止事件的訊息，而不同標誌時間的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 間的比較，則可以了解某標誌時間對其預後的影響，故本研究提供了標誌事件作為預測終止事件的方法。


使用無母數方法估計 $P(T_M \leq m)$ 時，會遇到估計 $P(T_M \leq m, T_D > T_M)$ 的問題，因為右受限的限制及終止事件的發生，使得 $P(T_M \leq m, T_D > T_M)$ 的部分是無法觀察，因此部分的樣本數通常很少，使用無母數法可能無法使用充分的訊息。但本論文通過依賴性 Cox 模式的引入及 $P(u < T_M \leq m, T_D > u)$ 的部分來估計 $P(T_M \leq m, T_D > T_M)$ ，如此可充分利用資料所具有的訊息。但觀察資料會有右受限的情況，必須用 IPCW 來解決右受限的問題(Fu et al., 2015)，但本文是使用條件機率式 $P(u < T_M \leq m | T_D > u)$ 來估計，此時會有分子及分母同時有右受限的情況，但分子會受到存活時間及右受限時間的雙重受限(暫譯為 dual censoring，跟雙側受限(double censoring)不同)，分母只受右受限時間的受限，兩者右受限的條件是不同的，故分子及分母須同時利用 IPCW 法處理，這應是未來在處理這種雙變數資料要考慮的問題，而本文也提供了解決的方法。而在估計 $P(T_M \leq m)$ 時，有兩個方法，第一個方法 $P(T_M > m) = 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M \leq T_D)$ 會需要估計 $P(T_M \leq T_D)$ ，而

估計 $P(T_M \leq T_D)$ 亦是用 IPCW 來處理(見附錄 C)，另一個方法則是 $P(T_M > m) = 1 - P(T_M \leq m)$ ，則不須估計 $P(T_M \leq T_D)$ ，且此兩方法在模擬上並沒有差別。

甲狀腺乳突癌的患者預後良好(Su et al., 2015)，尤其是 40 歲內的患者，甚至當 $m=8$ 的 $S(t|T_M \leq m, T_D > m)$ 在十年後仍有 0.88 以上，亦即標誌事件在 8 年內發生者，如能活過 8 年者，其在第 18 年存活之機率達八成八之上，這是其他癌症少見的。只是甲狀腺乳突癌預後很好，會受到其他競爭死因的影響，尤其是在 40 歲以上的患者這可從圖十三看出， $m=8$ 時，存活率下降很快。亦即如出現存活曲線交叉的現象，表示後面出現標誌事件的患者反而預後不好，故透過不同時間的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 可以了解標誌時間對其存活機率的現象，亦即不同的標誌時間可能有不一樣的影響。

$S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 在臨床上的運用較不容易理解，而 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$ 比較適用於臨床上的解釋。但 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 具有測量工具上的涵意，即具有敏感性及特異性的性質，未來具有發展成了解測量工具好壞的方法。故未來努力的方向為持續發展 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的用途，另實例中是用分組的方法了解條件存活機率在不同類別的差異，但因分組後，個案數會變少，而導致變異數變大的問題，故將擴展本式的方法至 Z 的部分範圍，透過樣本的訊息的充分應用而降低變異數。雖我們可透過不同的條件存活曲線來了解不同標誌時間對存活的影响，但未來仍應量化不同標誌時間對存活的影响，並應利用這些方法用在臨床上之應用，以提供臨床上的判斷，例如甲狀腺乳突癌患者。

本論文的目的是在於利用患者的動態指標訊息及基本特性來預測未



來的存活機率，亦即利用標誌的訊息，去不斷地更新對未來存活的預測。本文的主要方法是採用時間依賴性 Cox 模式去明確地架構標誌歷程、共變數和終止事件間的相關性與時序性，且對於標誌歷程和其他共變數的分布未有特定的假設，讓方法具有非常大的彈性，此有別於一般以標誌歷程預測存活機率的方法，多必須分別模式化共變數和各階段或各事件間的所有相關性(Lakhal et al., 2008; Mauguen et al., 2013; Porta et al., 2012)。並運用 Xu 等人(2000)對貝氏定理的運用而作機率的反轉及條件式機率的運算，但因估計式中用的是時間依賴性 Cox 模式，這也擴展了他們(Xu and O'Quigley, 2000)限於靜態的共變數的方法而可使用動態的時間依賴性變數。而運用貝氏定理作機率的反轉，也使得原來不可估計的標誌時間邊際機率函數轉為可估計的條件式機率函數。且本文不須計算基礎風險函數，也不同于一般以標誌歷程預測存活機率，多需要計算基礎風險函數(Lakhal et al., 2008; Mauguen et al., 2013; Porta et al., 2012)。本論文利用共變數和標誌歷程的條件式分布，在給予適當的權重上(Cox 模式中的參數及變數值所構成)，可在不同標誌歷程及共變數的條件下，估計未來存活機率，具臨床上運用之價值。

表二 $S(t|T_M \leq m)$ 的真值及三組右受限資料之模擬結果

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
真值 M+1	0.494	0.418	0.356	0.306	0.265	0.232	0.203
N=500 B, SD	-1.61*10 ⁻² , 0.118	-9.49*10 ⁻³ , 0.104	-2.94*10 ⁻³ , 0.092	1.15*10 ⁻³ , 0.085	-1.49*10 ⁻⁴ , 0.078	-5.53*10 ⁻⁴ , 0.072	5.79*10 ⁻⁴ , 0.069
N=750 B, SD	-8.19*10 ⁻³ , 0.094	-3.71*10 ⁻⁴ , 0.084	1.77*10 ⁻³ , 0.075	1.17*10 ⁻³ , 0.068	9.88*10 ⁻⁴ , 0.062	-9.65*10 ⁻⁵ , 0.058	4.81*10 ⁻⁴ , 0.053
N=1000 B, SD	-5.63*10 ⁻³ , 0.083	-2.01*10 ⁻³ , 0.070	9.44*10 ⁻⁴ , 0.062	1.61*10 ⁻³ , 0.057	4.98*10 ⁻⁴ , 0.053	-7.26*10 ⁻⁴ , 0.050	-3.88*10 ⁻⁴ , 0.046
真值 M+2	0.364	0.310	0.266	0.230	0.201	0.176	0.155
N=500 B, SD	-1.61*10 ⁻² , 0.127	-8.23*10 ⁻³ , 0.107	-2.32*10 ⁻³ , 0.092	6.14*10 ⁻⁴ , 0.084	1.03*10 ⁻⁴ , 0.075	-9.24*10 ⁻⁴ , 0.068	5.80*10 ⁻⁴ , 0.065
N=750 B, SD	-4.54*10 ⁻³ , 0.102	7.73*10 ⁻⁴ , 0.087	2.18*10 ⁻³ , 0.075	1.36*10 ⁻³ , 0.066	-1.13*10 ⁻⁴ , 0.060	-2.97*10 ⁻⁴ , 0.055	-2.13*10 ⁻⁴ , 0.050
N=1000 B, SD	-4.18*10 ⁻³ , 0.089	-2.13*10 ⁻³ , 0.073	3.33*10 ⁻⁴ , 0.063	8.83*10 ⁻⁴ , 0.056	8.52*10 ⁻⁴ , 0.052	-4.48*10 ⁻⁴ , 0.047	-5.75*10 ⁻⁴ , 0.044
真值 M+3	0.271	0.233	0.201	0.175	0.153	0.135	0.119
N=500 B, SD	-1.41*10 ⁻² , 0.127	-6.91*10 ⁻³ , 0.104	-2.93*10 ⁻³ , 0.090	1.16*10 ⁻³ , 0.079	1.56*10 ⁻⁴ , 0.070	-8.17*10 ⁻⁴ , 0.063	6.24*10 ⁻⁴ , 0.060
N=750 B, SD	-2.37*10 ⁻³ , 0.103	1.15*10 ⁻³ , 0.086	2.90*10 ⁻³ , 0.072	6.54*10 ⁻⁴ , 0.063	-5.81*10 ⁻⁴ , 0.056	-6.75*10 ⁻⁴ , 0.051	-2.61*10 ⁻⁴ , 0.046
N=1000 B, SD	-3.48*10 ⁻³ , 0.089	-2.40*10 ⁻³ , 0.072	-1.39*10 ⁻⁴ , 0.061	1.04*10 ⁻⁴ , 0.053	8.88*10 ⁻⁴ , 0.049	-4.13*10 ⁻⁴ , 0.045	-6.25*10 ⁻⁴ , 0.041
真值 M+4	0.204	0.177	0.153	0.134	0.118	0.105	0.093
N=500 B, SD	-1.28*10 ⁻² , 0.120	-6.67*10 ⁻³ , 0.100	-2.61*10 ⁻³ , 0.084	1.41*10 ⁻³ , 0.074	2.96*10 ⁻⁴ , 0.066	-7.45*10 ⁻⁴ , 0.058	4.41*10 ⁻⁴ , 0.055
N=750 B, SD	-5.67*10 ⁻⁴ , 0.100	1.74*10 ⁻³ , 0.082	2.49*10 ⁻³ , 0.068	2.87*10 ⁻⁴ , 0.058	-9.52*10 ⁻⁴ , 0.052	-7.04*10 ⁻⁴ , 0.047	-5.78*10 ⁻⁴ , 0.043
N=1000 B, SD	-2.80*10 ⁻³ , 0.086	-2.72*10 ⁻³ , 0.069	1.72*10 ⁻⁴ , 0.057	8.85*10 ⁻⁴ , 0.050	7.11*10 ⁻⁴ , 0.046	-4.37*10 ⁻⁴ , 0.042	-1.08*10 ⁻³ , 0.037
真值 M+5	0.155	0.135	0.118	0.104	0.091	0.081	0.072
N=500 B, SD	-1.21*10 ⁻² , 0.113	-5.64*10 ⁻³ , 0.092	-2.09*10 ⁻³ , 0.077	1.55*10 ⁻³ , 0.069	4.07*10 ⁻⁴ , 0.060	-7.76*10 ⁻⁴ , 0.053	1.46*10 ⁻⁴ , 0.050
N=750 B, SD	5.81*10 ⁻⁴ , 0.094	1.45*10 ⁻³ , 0.077	2.26*10 ⁻³ , 0.063	-8.05*10 ⁻⁵ , 0.054	-9.66*10 ⁻⁴ , 0.047	-9.18*10 ⁻⁴ , 0.043	-7.01*10 ⁻⁴ , 0.039
N=1000 B, SD	-2.28*10 ⁻³ , 0.081	-2.29*10 ⁻³ , 0.064	1.16*10 ⁻⁴ , 0.053	5.48*10 ⁻⁴ , 0.047	6.13*10 ⁻⁴ , 0.042	-8.67*10 ⁻⁴ , 0.038	-9.55*10 ⁻⁴ , 0.035
真值 M+6	0.119	0.104	0.091	0.081	0.071	0.063	0.056
N=500 B, SD	-1.04*10 ⁻² , 0.105	-4.56*10 ⁻³ , 0.085	-1.76*10 ⁻³ , 0.071	1.44*10 ⁻³ , 0.063	4.22*10 ⁻⁴ , 0.055	-9.97*10 ⁻⁴ , 0.048	1.39*10 ⁻⁴ , 0.046
N=750 B, SD	4.71*10 ⁻⁴ , 0.087	1.49*10 ⁻³ , 0.071	1.85*10 ⁻³ , 0.059	-2.95*10 ⁻⁴ , 0.049	-1.08*10 ⁻³ , 0.043	-9.91*10 ⁻⁴ , 0.039	-9.43*10 ⁻⁴ , 0.035
N=1000 B, SD	-1.52*10 ⁻³ , 0.075	-2.09*10 ⁻³ , 0.059	-6.46*10 ⁻⁵ , 0.049	2.83*10 ⁻⁴ , 0.043	8.58*10 ⁻⁵ , 0.039	-7.75*10 ⁻⁴ , 0.035	-1.04*10 ⁻³ , 0.032
真值 M+7	0.092	0.081	0.071	0.063	0.056	0.050	0.044
N=500 B, SD	-8.33*10 ⁻³ , 0.096	-3.74*10 ⁻³ , 0.078	-1.58*10 ⁻³ , 0.065	1.36*10 ⁻³ , 0.059	2.52*10 ⁻⁴ , 0.050	-9.47*10 ⁻⁴ , 0.044	3.05*10 ⁻⁴ , 0.041
N=750 B, SD	4.81*10 ⁻⁴ , 0.080	1.40*10 ⁻³ , 0.065	1.48*10 ⁻³ , 0.053	-5.75*10 ⁻⁴ , 0.045	-1.00*10 ⁻³ , 0.039	-1.23*10 ⁻³ , 0.035	-6.46*10 ⁻⁴ , 0.032
N=1000 B, SD	-1.23*10 ⁻³ , 0.069	-2.14*10 ⁻³ , 0.054	-2.77*10 ⁻⁴ , 0.045	-3.12*10 ⁻⁴ , 0.040	-7.28*10 ⁻⁵ , 0.036	-8.71*10 ⁻⁴ , 0.032	-9.11*10 ⁻⁴ , 0.029
真值 M+8	0.071	0.063	0.056	0.049	0.044	0.039	0.035
N=500 B, SD	-6.40*10 ⁻³ , 0.087	-3.08*10 ⁻³ , 0.071	-1.37*10 ⁻³ , 0.060	1.09*10 ⁻³ , 0.054	2.18*10 ⁻⁴ , 0.046	-6.97*10 ⁻⁴ , 0.040	1.92*10 ⁻⁴ , 0.037
N=750 B, SD	6.26*10 ⁻⁴ , 0.073	1.19*10 ⁻³ , 0.058	9.71*10 ⁻⁴ , 0.049	-6.12*10 ⁻⁴ , 0.041	-1.14*10 ⁻³ , 0.035	-9.24*10 ⁻⁴ , 0.032	-6.88*10 ⁻⁴ , 0.028
N=1000 B, SD	-1.25*10 ⁻³ , 0.063	-2.13*10 ⁻³ , 0.049	-7.7*10 ⁻⁴ , 0.041	-6.31*10 ⁻⁴ , 0.036	-3.28*10 ⁻⁴ , 0.032	-7.74*10 ⁻⁴ , 0.029	-9.51*10 ⁻⁴ , 0.026
真值 M+9	0.056	0.049	0.044	0.039	0.034	0.031	0.028
N=500 B, SD	-4.94*10 ⁻³ , 0.080	-2.28*10 ⁻³ , 0.065	-1.33*10 ⁻³ , 0.055	1.02*10 ⁻³ , 0.049	2.45*10 ⁻⁴ , 0.042	-6.31*10 ⁻⁴ , 0.036	1.70*10 ⁻⁴ , 0.034
N=750 B, SD	6.34*10 ⁻⁴ , 0.067	7.18*10 ⁻⁴ , 0.053	7.44*10 ⁻⁴ , 0.045	-8.16*10 ⁻⁴ , 0.037	-8.73*10 ⁻⁴ , 0.032	-9.27*10 ⁻⁴ , 0.028	-6.74*10 ⁻⁴ , 0.025
N=1000 B, SD	-1.18*10 ⁻³ , 0.057	-2.36*10 ⁻³ , 0.045	-1.11*10 ⁻³ , 0.038	-8.65*10 ⁻⁴ , 0.033	-3.23*10 ⁻⁴ , 0.029	-8.73*10 ⁻⁴ , 0.026	-8.81*10 ⁻⁴ , 0.024
真值 M+10	0.044	0.039	0.035	0.031	0.027	0.024	0.022
N=500 B, SD	-3.62*10 ⁻³ , 0.073	-1.65*10 ⁻³ , 0.059	-9.73*10 ⁻⁴ , 0.050	9.78*10 ⁻⁴ , 0.045	2.23*10 ⁻⁴ , 0.038	-5.99*10 ⁻⁴ , 0.033	4.97*10 ⁻⁵ , 0.031
N=750 B, SD	3.05*10 ⁻⁴ , 0.060	5.43*10 ⁻⁴ , 0.048	3.90*10 ⁻⁴ , 0.040	-6.32*10 ⁻⁴ , 0.033	-7.74*10 ⁻⁴ , 0.029	-8.57*10 ⁻⁴ , 0.026	-9.20*10 ⁻⁴ , 0.023
N=1000 B, SD	-1.37*10 ⁻³ , 0.052	-2.43*10 ⁻³ , 0.041	-1.32*10 ⁻³ , 0.035	-8.16*10 ⁻⁴ , 0.030	-4.61*10 ⁻⁴ , 0.026	-7.97*10 ⁻⁴ , 0.023	-1.04*10 ⁻³ , 0.021

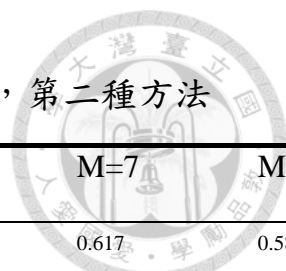
B:bias, 估計值跟真值之差; SD:估計值之標準差; M:標誌時間

表三 $S(t|T_M > m)$ 的真值及三組右受限資料之模擬結果，第一種方法

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
真值 M+1	0.829	0.781	0.735	0.693	0.653	0.616	0.582
N=500 B, SD	-5.14*10 ⁻⁴ , 0.015	-2.54*10 ⁻⁴ , 0.016	2.68*10 ⁻⁴ , 0.018	9.10*10 ⁻⁴ , 0.021	6.13*10 ⁻⁴ , 0.020	3.82*10 ⁻⁴ , 0.021	-1.76*10 ⁻⁴ , 0.023
N=750 B, SD	-5.81*10 ⁻⁴ , 0.012	1.97*10 ⁻⁴ , 0.014	1.54*10 ⁻⁴ , 0.015	1.58*10 ⁻⁴ , 0.016	4.07*10 ⁻⁴ , 0.018	-9.71*10 ⁻⁵ , 0.019	6.75*10 ⁻⁵ , 0.019
2N=1000 B, SD	-5.55*10 ⁻⁵ , 0.011	1.33*10 ⁻⁴ , 0.012	-1.52*10 ⁻⁴ , 0.013	-4.31*10 ⁻⁴ , 0.014	-7.80*10 ⁻⁴ , 0.015	-6.44*10 ⁻⁴ , 0.016	-5.19*10 ⁻⁴ , 0.017
真值 M+2	0.776	0.731	0.690	0.650	0.614	0.579	0.547
N=500 B, SD	-2.40*10 ⁻⁴ , 0.016	3.64*10 ⁻⁴ , 0.018	1.05*10 ⁻³ , 0.019	5.85*10 ⁻⁴ , 0.020	2.64*10 ⁻⁴ , 0.021	-1.17*10 ⁻⁴ , 0.022	-7.06*10 ⁻⁵ , 0.024
N=750 B, SD	3.43*10 ⁻⁴ , 0.014	1.48*10 ⁻⁴ , 0.015	4.77*10 ⁻⁵ , 0.016	3.82*10 ⁻⁴ , 0.018	-1.80*10 ⁻⁷ , 0.018	-4.13*10 ⁻⁵ , 0.019	1.00*10 ⁻⁴ , 0.020
N=1000 B, SD	1.34*10 ⁻⁴ , 0.012	-1.87*10 ⁻⁴ , 0.013	-3.98*10 ⁻⁴ , 0.014	-7.75*10 ⁻⁴ , 0.015	-5.95*10 ⁻⁴ , 0.016	-5.35*10 ⁻⁴ , 0.017	-5.32*10 ⁻⁴ , 0.018
真值 M+3	0.726	0.684	0.646	0.609	0.575	0.543	0.513
N=500 B, SD	4.04*10 ⁻⁴ , 0.018	1.12*10 ⁻³ , 0.019	7.31*10 ⁻⁴ , 0.020	2.38*10 ⁻⁴ , 0.021	-2.80*10 ⁻⁴ , 0.022	-4.18*10 ⁻⁵ , 0.024	4.14*10 ⁻⁶ , 0.025
N=750 B, SD	2.79*10 ⁻⁴ , 0.015	5.55*10 ⁻⁵ , 0.016	2.20*10 ⁻⁴ , 0.018	-6.63*10 ⁻⁵ , 0.018	7.32*10 ⁻⁵ , 0.019	-6.38*10 ⁻⁵ , 0.020	5.81*10 ⁻⁴ , 0.020
N=1000 B, SD	-2.01*10 ⁻⁴ , 0.013	-4.63*10 ⁻⁴ , 0.014	-7.75*10 ⁻⁴ , 0.015	-5.64*10 ⁻⁴ , 0.016	-4.56*10 ⁻⁴ , 0.017	-5.71*10 ⁻⁴ , 0.018	-2.52*10 ⁻⁴ , 0.018
真值 M+4	0.677	0.639	0.604	0.570	0.539	0.509	0.481
N=500 B, SD	1.18*10 ⁻³ , 0.019	7.73*10 ⁻⁴ , 0.021	4.20*10 ⁻⁴ , 0.021	-3.07*10 ⁻⁴ , 0.022	-2.08*10 ⁻⁴ , 0.024	2.44*10 ⁻⁵ , 0.025	-4.31*10 ⁻⁵ , 0.025
N=750 B, SD	1.46*10 ⁻⁴ , 0.017	2.48*10 ⁻⁴ , 0.018	-2.58*10 ⁻⁴ , 0.019	-5.24*10 ⁻⁶ , 0.019	3.83*10 ⁻⁵ , 0.020	4.05*10 ⁻⁴ , 0.020	7.43*10 ⁻⁴ , 0.021
N=1000 B, SD	-5.02*10 ⁻⁴ , 0.014	-8.48*10 ⁻⁴ , 0.015	-5.80*10 ⁻⁴ , 0.016	-4.05*10 ⁻⁴ , 0.017	-4.70*10 ⁻⁴ , 0.018	-2.96*10 ⁻⁴ , 0.018	-5.20*10 ⁻⁴ , 0.018
真值 M+5	0.632	0.597	0.564	0.533	0.504	0.476	0.450
N=500 B, SD	8.34*10 ⁻⁴ , 0.021	4.45*10 ⁻⁴ , 0.021	-1.18*10 ⁻⁴ , 0.022	-2.30*10 ⁻⁴ , 0.024	-1.51*10 ⁻⁴ , 0.024	-4.54*10 ⁻⁵ , 0.025	-3.50*10 ⁻⁴ , 0.025
N=750 B, SD	3.26*10 ⁻⁴ , 0.018	-2.26*10 ⁻⁴ , 0.019	-2.15*10 ⁻⁴ , 0.019	-4.73*10 ⁻⁵ , 0.019	4.95*10 ⁻⁴ , 0.020	5.49*10 ⁻⁴ , 0.020	9.22*10 ⁻⁴ , 0.021
N=1000 B, SD	-9.11*10 ⁻⁴ , 0.015	-6.50*10 ⁻⁴ , 0.016	-4.32*10 ⁻⁴ , 0.017	-4.04*10 ⁻⁴ , 0.018	-1.94*10 ⁻⁴ , 0.018	-5.69*10 ⁻⁴ , 0.018	-6.74*10 ⁻⁵ , 0.018
真值 M+6	0.590	0.558	0.527	0.498	0.471	0.446	0.422
N=500 B, SD	5.00*10 ⁻⁴ , 0.021	-1.08*10 ⁻⁴ , 0.022	-3.91*10 ⁻⁵ , 0.024	-1.53*10 ⁻⁴ , 0.025	-2.25*10 ⁻⁴ , 0.025	-3.60*10 ⁻⁴ , 0.025	-3.81*10 ⁻⁴ , 0.026
N=750 B, SD	-1.52*10 ⁻⁴ , 0.019	-1.94*10 ⁻⁴ , 0.019	-2.57*10 ⁻⁴ , 0.020	4.15*10 ⁻⁴ , 0.020	6.19*10 ⁻⁴ , 0.020	7.20*10 ⁻⁴ , 0.020	6.38*10 ⁻⁴ , 0.020
N=1000 B, SD	-7.20*10 ⁻⁴ , 0.016	-5.10*10 ⁻⁴ , 0.017	-4.45*10 ⁻⁴ , 0.018	-1.19*10 ⁻⁴ , 0.018	-4.63*10 ⁻⁴ , 0.018	-1.17*10 ⁻⁴ , 0.018	-2.76*10 ⁻⁴ , 0.018
真值 M+7	0.550	0.520	0.492	0.466	0.441	0.417	0.395
N=500 B, SD	-5.43*10 ⁻⁵ , 0.022	-4.45*10 ⁻⁵ , 0.024	2.54*10 ⁻⁵ , 0.024	-2.19*10 ⁻⁴ , 0.025	-5.41*10 ⁻⁴ , 0.025	-3.96*10 ⁻⁴ , 0.026	-1.17*10 ⁻⁴ , 0.026
N=750 B, SD	-1.16*10 ⁻⁴ , 0.019	-2.47*10 ⁻⁴ , 0.019	2.09*10 ⁻⁴ , 0.020	5.46*10 ⁻⁴ , 0.020	7.62*10 ⁻⁴ , 0.020	4.42*10 ⁻⁴ , 0.020	8.42*10 ⁻⁴ , 0.020
N=1000 B, SD	-5.80*10 ⁻⁴ , 0.017	-5.25*10 ⁻⁴ , 0.018	-1.70*10 ⁻⁴ , 0.018	-3.87*10 ⁻⁴ , 0.018	1.16*10 ⁻⁶ , 0.018	-3.22*10 ⁻⁴ , 0.018	4.78*10 ⁻⁶ , 0.018
真值 M+8	0.513	0.485	0.460	0.435	0.412	0.390	0.370
N=500 B, SD	-1.12*10 ⁻⁵ , 0.024	6.20*10 ⁻⁷ , 0.024	-5.31*10 ⁻⁵ , 0.025	-5.29*10 ⁻⁴ , 0.025	-5.60*10 ⁻⁴ , 0.025	-1.42*10 ⁻⁴ , 0.026	-2.86*10 ⁻⁶ , 0.026
N=750 B, SD	-1.81*10 ⁻⁴ , 0.019	2.02*10 ⁻⁴ , 0.020	3.57*10 ⁻⁴ , 0.020	6.97*10 ⁻⁴ , 0.020	4.69*10 ⁻⁴ , 0.020	6.57*10 ⁻⁴ , 0.020	5.62*10 ⁻⁴ , 0.020
N=1000 B, SD	-5.93*10 ⁻⁴ , 0.017	-2.60*10 ⁻⁴ , 0.018	-4.45*10 ⁻⁴ , 0.018	7.82*10 ⁻⁵ , 0.018	-1.94*10 ⁻⁴ , 0.018	-3.69*10 ⁻⁵ , 0.018	-3.48*10 ⁻⁵ , 0.019
真值 M+9	0.478	0.453	0.429	0.407	0.385	0.365	0.346
N=500 B, SD	1.74*10 ⁻⁵ , 0.024	-9.86*10 ⁻⁵ , 0.025	-3.77*10 ⁻⁴ , 0.025	-5.51*10 ⁻⁴ , 0.025	-2.82*10 ⁻⁴ , 0.025	-5.12*10 ⁻⁵ , 0.026	2.49*10 ⁻⁴ , 0.026
N=750 B, SD	2.49*10 ⁻⁴ , 0.020	3.44*10 ⁻⁴ , 0.020	5.23*10 ⁻⁴ , 0.020	4.11*10 ⁻⁴ , 0.020	6.85*10 ⁻⁴ , 0.020	3.85*10 ⁻⁴ , 0.020	6.99*10 ⁻⁴ , 0.021
N=1000 B, SD	-3.28*10 ⁻⁴ , 0.017	-5.45*10 ⁻⁴ , 0.018	1.26*10 ⁻⁵ , 0.018	-1.23*10 ⁻⁴ , 0.018	8.86*10 ⁻⁵ , 0.018	-6.71*10 ⁻⁵ , 0.019	9.75*10 ⁻⁵ , 0.019
真值 M+10	0.446	0.423	0.401	0.380	0.360	0.341	0.323
N=500 B, SD	-8.36*10 ⁻⁵ , 0.024	-4.43*10 ⁻⁴ , 0.024	-4.18*10 ⁻⁴ , 0.025	-2.73*10 ⁻⁴ , 0.025	-1.77*10 ⁻⁴ , 0.026	1.95*10 ⁻⁴ , 0.026	6.07*10 ⁻⁵ , 0.026
N=750 B, SD	3.77*10 ⁻⁴ , 0.020	5.06*10 ⁻⁴ , .020	2.57*10 ⁻⁴ , 0.020	6.31*10 ⁻⁴ , 0.020	4.03*10 ⁻⁴ , 0.020	5.22*10 ⁻⁴ , 0.020	3.62*10 ⁻⁴ , 0.021
N=1000 B, SD	-6.07*10 ⁻⁴ , 0.017	-1.05*10 ⁻⁴ , 0.018	-1.85*10 ⁻⁴ , 0.018	1.49*10 ⁻⁴ , 0.018	5.10*10 ⁻⁵ , 0.018	6.63*10 ⁻⁵ , 0.019	-1.22*10 ⁻⁴ , 0.019

B:bias，估計值跟真值之差；SD: 估計值之標準差；M:標誌時間

方法: $P(T_M > m) = 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M \leq T_D)$



表四 $S(t|T_M > m)$ 的真值及三組右受限資料之模擬結果，第二種方法

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
真值 M+1	0.830	0.781	0.736	0.694	0.654	0.617	0.582
N=500 B, SD	-5.20*10 ⁻⁵ , 0.017	1.84*10 ⁻⁴ , 0.018	6.94*10 ⁻⁴ , 0.020	1.32*10 ⁻³ , 0.021	1.02*10 ⁻³ , 0.022	7.66*10 ⁻⁴ , 0.022	1.86*10 ⁻⁴ , 0.023
N=750 B, SD	-1.78*10 ⁻⁴ , 0.014	5.84*10 ⁻⁴ , 0.015	5.19*10 ⁻⁴ , 0.017	5.05*10 ⁻⁴ , 0.018	7.34*10 ⁻⁴ , 0.019	2.11*10 ⁻⁴ , 0.019	3.60*10 ⁻⁴ , 0.020
N=1000 B, SD	-3.10*10 ⁻⁴ , 0.012	-1.18*10 ⁻⁴ , 0.014	-3.95*10 ⁻⁴ , 0.015	-6.73*10 ⁻⁴ , 0.016	-1.03*10 ⁻³ , 0.016	-8.72*10 ⁻⁴ , 0.017	-7.40*10 ⁻⁴ , 0.018
真值 M+2	0.777	0.732	0.690	0.651	0.614	0.580	0.547
N=500 B, SD	1.90*10 ⁻⁴ , 0.018	7.81*10 ⁻⁴ , 0.020	1.46*10 ⁻³ , 0.021	9.77*10 ⁻⁴ , 0.022	6.34*10 ⁻⁴ , 0.022	2.38*10 ⁻⁴ , 0.023	2.60*10 ⁻⁴ , 0.025
N=750 B, SD	7.18*10 ⁻⁴ , 0.016	5.07*10 ⁻⁴ , 0.017	3.87*10 ⁻⁴ , 0.018	7.03*10 ⁻⁴ , 0.019	3.01*10 ⁻⁴ , 0.019	2.45*10 ⁻⁴ , 0.020	3.71*10 ⁻⁴ , 0.020
N=1000 B, SD	-1.07*10 ⁻⁴ , 0.014	-4.25*10 ⁻⁴ , 0.015	-6.33*10 ⁻⁴ , 0.016	-1.00*10 ⁻³ , 0.016	-8.16*10 ⁻⁴ , 0.017	-7.51*10 ⁻⁴ , 0.018	-7.40*10 ⁻⁴ , 0.018
真值 M+3	0.726	0.685	0.646	0.610	0.576	0.544	0.513
N=500 B, SD	8.09*10 ⁻⁴ , 0.020	1.51*10 ⁻³ , 0.021	1.12*10 ⁻³ , 0.022	5.94*10 ⁻⁴ , 0.022	6.11*10 ⁻⁵ , 0.023	2.81*10 ⁻⁴ , 0.024	3.09*10 ⁻⁴ , 0.025
N=750 B, SD	6.25*10 ⁻⁴ , 0.017	3.88*10 ⁻⁴ , 0.018	5.32*10 ⁻⁴ , 0.019	2.29*10 ⁻⁴ , 0.019	3.53*10 ⁻⁴ , 0.020	2.00*10 ⁻⁴ , 0.020	8.31*10 ⁻⁴ , 0.020
N=1000 B, SD	-4.28*10 ⁻⁴ , 0.015	-6.92*10 ⁻⁴ , 0.016	-9.96*10 ⁻⁴ , 0.016	-7.80*10 ⁻⁴ , 0.017	-6.66*10 ⁻⁴ , 0.018	-7.76*10 ⁻⁴ , 0.018	-4.50*10 ⁻⁴ , 0.018
真值 M+4	0.678	0.640	0.604	0.571	0.539	0.509	0.481
N=500 B, SD	1.56*10 ⁻³ , 0.021	1.14*10 ⁻³ , 0.022	7.70*10 ⁻⁴ , 0.023	1.94*10 ⁻⁵ , 0.023	1.02*10 ⁻⁴ , 0.024	3.21*10 ⁻⁴ , 0.025	2.29*10 ⁻⁴ , 0.025
N=750 B, SD	4.67*10 ⁻⁴ , 0.018	5.54*10 ⁻⁴ , 0.019	2.90*10 ⁻⁵ , 0.019	2.68*10 ⁻⁴ , 0.019	2.95*10 ⁻⁴ , 0.020	6.48*10 ⁻⁴ , 0.020	9.78*10 ⁻⁴ , 0.021
N=1000 B, SD	-7.21*10 ⁻⁴ , 0.016	-1.06*10 ⁻³ , 0.016	-7.88*10 ⁻⁴ , 0.017	-6.10*10 ⁻⁴ , 0.018	-6.67*10 ⁻⁴ , 0.018	-4.90*10 ⁻⁴ , 0.018	-7.08*10 ⁻⁴ , 0.018
真值 M+5	0.633	0.598	0.565	0.534	0.504	0.477	0.451
N=500 B, SD	1.19*10 ⁻³ , 0.022	7.81*10 ⁻⁴ , 0.023	2.02*10 ⁻⁴ , 0.023	6.58*10 ⁻⁵ , 0.024	1.32*10 ⁻⁴ , 0.025	2.19*10 ⁻⁴ , 0.025	-9.99*10 ⁻⁵ , 0.025
N=750 B, SD	6.20*10 ⁻⁴ , 0.019	5.36*10 ⁻⁵ , 0.019	5.05*10 ⁻⁵ , 0.019	2.03*10 ⁻⁴ , 0.020	7.32*10 ⁻⁴ , 0.020	7.76*10 ⁻⁴ , 0.020	1.14*10 ⁻³ , 0.020
N=1000 B, SD	-1.12*10 ⁻³ , 0.016	-8.53*10 ⁻⁴ , 0.017	-6.30*10 ⁻⁴ , 0.018	-5.97*10 ⁻⁴ , 0.018	-3.80*10 ⁻⁴ , 0.018	-7.52*10 ⁻⁴ , 0.018	-2.42*10 ⁻⁴ , 0.018
真值 M+6	0.590	0.558	0.527	0.499	0.472	0.446	0.422
N=500 B, SD	8.27*10 ⁻⁴ , 0.022	2.00*10 ⁻⁴ , 0.023	2.51*10 ⁻⁴ , 0.024	1.17*10 ⁻⁴ , 0.025	2.69*10 ⁻⁵ , 0.025	-1.17*10 ⁻⁴ , 0.025	-1.58*10 ⁻⁴ , 0.025
N=750 B, SD	1.17*10 ⁻⁴ , 0.019	6.54*10 ⁻⁵ , 0.019	-1.32*10 ⁻⁵ , 0.020	6.46*10 ⁻⁴ , 0.020	8.39*10 ⁻⁴ , 0.020	9.32*10 ⁻⁴ , 0.020	8.40*10 ⁻⁴ , 0.020
N=1000 B, SD	-9.12*10 ⁻⁴ , 0.017	-7.03*10 ⁻⁴ , 0.018	-6.31*10 ⁻⁴ , 0.018	-3.01*10 ⁻⁴ , 0.018	-6.39*10 ⁻⁴ , 0.018	-2.88*10 ⁻⁴ , 0.018	-4.39*10 ⁻⁴ , 0.018
真值 M+7	0.551	0.521	0.492	0.466	0.441	0.417	0.395
N=500 B, SD	2.44*10 ⁻⁴ , 0.023	2.33*10 ⁻⁴ , 0.024	2.90*10 ⁻⁴ , 0.024	2.01*10 ⁻⁵ , 0.025	-3.01*10 ⁻⁴ , 0.025	-1.81*10 ⁻⁴ , 0.025	7.88*10 ⁻⁵ , 0.025
N=750 B, SD	1.31*10 ⁻⁴ , 0.019	-1.02*10 ⁻⁵ , 0.020	4.33*10 ⁻⁴ , 0.020	7.60*10 ⁻⁴ , 0.020	9.67*10 ⁻⁴ , 0.020	6.37*10 ⁻⁴ , 0.020	1.03*10 ⁻³ , 0.020
N=1000 B, SD	-7.62*10 ⁻⁴ , 0.018	-7.05*10 ⁻⁴ , 0.018	-3.44*10 ⁻⁴ , 0.018	-5.59*10 ⁻⁴ , 0.018	-1.62*10 ⁻⁴ , 0.018	-4.81*10 ⁻⁴ , 0.018	-1.50*10 ⁻⁴ , 0.018
真值 M+8	0.513	0.486	0.460	0.435	0.412	0.390	0.370
N=500 B, SD	2.58*10 ⁻⁴ , 0.024	2.54*10 ⁻⁴ , 0.024	1.81*10 ⁻⁴ , 0.025	-3.11*10 ⁻⁴ , 0.024	-3.56*10 ⁻⁴ , 0.025	4.73*10 ⁻⁵ , 0.025	1.67*10 ⁻⁴ , 0.025
N=750 B, SD	4.52*10 ⁻⁴ , 0.019	4.21*10 ⁻⁴ , 0.020	5.64*10 ⁻⁴ , 0.020	8.96*10 ⁻⁴ , 0.020	6.58*10 ⁻⁴ , 0.020	8.36*10 ⁻⁴ , 0.020	7.29*10 ⁻⁴ , 0.020
N=1000 B, SD	-7.64*10 ⁻⁴ , 0.018	-4.29*10 ⁻⁴ , 0.018	-6.10*10 ⁻⁴ , 0.018	-8.12*10 ⁻⁵ , 0.018	-3.46*10 ⁻⁴ , 0.018	-1.88*10 ⁻⁴ , 0.018	-4.84*10 ⁻⁴ , 0.018
真值 M+9	0.479	0.453	0.429	0.407	0.385	0.365	0.346
N=500 B, SD	2.63*10 ⁻⁴ , 0.024	1.24*10 ⁻⁴ , 0.024	-1.63*10 ⁻⁴ , 0.024	-3.58*10 ⁻⁴ , 0.024	-1.02*10 ⁻⁴ , 0.025	1.12*10 ⁻⁴ , 0.025	4.01*10 ⁻⁴ , 0.025
N=750 B, SD	4.58*10 ⁻⁴ , 0.020	5.47*10 ⁻⁴ , 0.020	7.17*10 ⁻⁴ , 0.020	5.94*10 ⁻⁴ , 0.020	8.58*10 ⁻⁴ , 0.020	5.47*10 ⁻⁴ , 0.020	8.48*10 ⁻⁴ , 0.020
N=1000 B, SD	-4.88*10 ⁻⁴ , 0.017	-7.06*10 ⁻⁴ , 0.018	-1.40*10 ⁻⁴ , 0.018	-2.71*10 ⁻⁴ , 0.018	-5.63*10 ⁻⁵ , 0.018	-2.13*10 ⁻⁴ , 0.018	-4.20*10 ⁻⁵ , 0.018
真值 M+10	0.447	0.423	0.401	0.380	0.360	0.341	0.324
N=500 B, SD	1.33*10 ⁻⁴ , 0.024	-2.40*10 ⁻⁴ , 0.024	-2.29*10 ⁻⁴ , 0.024	-1.04*10 ⁻⁴ , 0.024	-2.29*10 ⁻⁵ , 0.025	3.40*10 ⁻⁴ , 0.025	1.92*10 ⁻⁴ , 0.025
N=750 B, SD	5.70*10 ⁻⁴ , 0.020	6.95*10 ⁻⁴ , 0.020	4.34*10 ⁻⁴ , 0.020	7.98*10 ⁻⁴ , 0.019	5.59*10 ⁻⁴ , 0.019	6.65*10 ⁻⁴ , 0.020	5.00*10 ⁻⁴ , 0.020
N=1000 B, SD	-7.59*10 ⁻⁴ , 0.017	-2.54*10 ⁻⁴ , 0.018	-3.28*10 ⁻⁴ , 0.017	7.50*10 ⁻⁶ , 0.018	-8.89*10 ⁻⁵ , 0.018	-6.98*10 ⁻⁵ , 0.018	-2.54*10 ⁻⁴ , 0.018

B:bias，估計值跟真值之差；SD：估計值之標準差；M:標誌時間

方法： $P(T_M > m) = 1 - P(T_M \leq m)$



表五 甲狀腺乳突癌(不分年紀)的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及

95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m)$							
M+1, (est, sd)	0.864, 0.125	0.846, 0.100	0.881, 0.070	0.857, 0.077	0.868, 0.060	0.838, 0.062	0.824, 0.063
95%信賴區間	0.634~1.000	0.650~1.000	0.744~1.000	0.707~1.000	0.750~0.987	0.716~0.960	0.700~0.947
M+2, (est, sd)	0.834, 0.147	0.824, 0.114	0.862, 0.080	0.838, 0.086	0.844, 0.070	0.821, 0.068	0.799, 0.071
95%信賴區間	0.545~1.000	0.600~1.000	0.705~1.000	0.669~1.000	0.706~0.982	0.688~0.954	0.660~0.939
M+3, (est, sd)	0.814, 0.165	0.800, 0.127	0.846, 0.088	0.814, 0.096	0.83, 0.075	0.798, 0.077	0.768, 0.080
95%信賴區間	0.490~1.000	0.551~1.000	0.674~1.000	0.625~1.000	0.685~0.978	0.658~0.948	0.610~0.925
M+4, (est, sd)	0.792, 0.182	0.781, 0.138	0.829, 0.095	0.804, 0.101	0.815, 0.081	0.767, 0.086	0.751, 0.085
95%信賴區間	0.435~1.000	0.511~1.000	0.642~1.000	0.606~1.000	0.656~0.974	0.598~0.935	0.584~0.917
M+5, (est, sd)	0.776, 0.193	0.761, 0.147	0.820, 0.099	0.788, 0.107	0.794, 0.087	0.750, 0.091	0.729, 0.091
95%信賴區間	0.397~1.000	0.473~1.000	0.627~1.000	0.578~0.997	0.623~0.965	0.572~0.928	0.550~0.908
M+6, (est, sd)	0.763, 0.202	0.751, 0.151	0.806, 0.105	0.769, 0.113	0.782, 0.091	0.729, 0.097	0.712, 0.097
95%信賴區間	0.367~1.000	0.456~1.000	0.601~1.000	0.546~0.991	0.604~0.960	0.539~0.919	0.522~0.902
M+7, (est, sd)	0.758, 0.205	0.736, 0.157	0.789, 0.110	0.759, 0.117	0.766, 0.096	0.713, 0.102	0.669, 0.108
95%信賴區間	0.357~1.000	0.429~1.000	0.573~1.000	0.530~0.987	0.579~0.954	0.512~0.913	0.458~0.880
M+8, (est, sd)	0.751, 0.209	0.719, 0.163	0.781, 0.113	0.746, 0.121	0.752, 0.100	0.672, 0.113	0.624, 0.119
95%信賴區間	0.342~1.000	0.400~1.000	0.559~1.000	0.509~0.983	0.556~0.949	0.451~0.893	0.391~0.858
M+9, (est, sd)	0.742, 0.221	0.712, 0.165	0.770, 0.117	0.733, 0.126	0.715, 0.112	0.630, 0.124	0.570, 0.134
95%信賴區間	0.327~1.000	0.388~1.000	0.541~0.998	0.487~0.980	0.496~0.935	0.388~0.873	0.307~0.833
M+10, (est, sd)	0.737, 0.214	0.701, 0.168	0.758, 0.121	0.700, 0.136	0.678, 0.124	0.580, 0.138	0.533, 0.145
95%信賴區間	0.319~1.000	0.371~1.000	0.521~0.995	0.433~0.967	0.435~0.921	0.309~0.851	0.249~0.817
$S(t T_M > m)$							
M+1, (est, sd)	0.974, 0.004	0.968, 0.005	0.962, 0.006	0.956, 0.006	0.947, 0.007	0.939, 0.008	0.933, 0.009
95%信賴區間	0.966~0.982	0.958~0.977	0.951~0.973	0.944~0.968	0.933~0.962	0.922~0.955	0.915~0.950
M+2, (est, sd)	0.967, 0.005	0.963, 0.005	0.955, 0.006	0.948, 0.007	0.935, 0.008	0.932, 0.009	0.922, 0.010
95%信賴區間	0.957~0.977	0.952~0.973	0.942~0.967	0.934~0.962	0.918~0.952	0.915~0.950	0.902~0.942
M+3, (est, sd)	0.962, 0.005	0.955, 0.006	0.946, 0.007	0.935, 0.008	0.928, 0.009	0.922, 0.010	0.909, 0.012
95%信賴區間	0.951~0.972	0.944~0.967	0.932~0.960	0.919~0.951	0.909~0.946	0.902~0.942	0.886~0.931
M+4, (est, sd)	0.954, 0.006	0.947, 0.007	0.932, 0.008	0.927, 0.009	0.915, 0.011	0.908, 0.012	0.901, 0.012
95%信賴區間	0.942~0.966	0.933~0.960	0.915~0.948	0.909~0.945	0.894~0.936	0.886~0.931	0.877~0.925
M+5, (est, sd)	0.944, 0.007	0.932, 0.008	0.924, 0.009	0.914, 0.010	0.899, 0.012	0.900, 0.012	0.892, 0.014
95%信賴區間	0.931~0.958	0.917~0.948	0.906~0.942	0.894~0.934	0.875~0.923	0.876~0.925	0.865~0.919
M+6, (est, sd)	0.930, 0.008	0.924, 0.009	0.910, 0.010	0.897, 0.012	0.890, 0.013	0.891, 0.014	0.883, 0.014
95%信賴區間	0.914~0.945	0.907~0.941	0.890~0.931	0.874~0.920	0.864~0.915	0.864~0.918	0.856~0.912
M+7, (est, sd)	0.921, 0.009	0.911, 0.010	0.893, 0.012	0.887, 0.013	0.879, 0.015	0.882, 0.014	0.863, 0.017
95%信賴區間	0.904~0.939	0.892~0.930	0.870~0.917	0.863~0.912	0.850~0.908	0.854~0.911	0.830~0.897
M+8, (est, sd)	0.907, 0.010	0.894, 0.011	0.883, 0.013	0.876, 0.014	0.870, 0.016	0.861, 0.017	0.846, 0.019
95%信賴區間	0.888~0.927	0.872~0.916	0.858~0.908	0.849~0.903	0.839~0.900	0.827~0.895	0.808~0.884
M+9, (est, sd)	0.890, 0.012	0.884, 0.012	0.871, 0.014	0.866, 0.015	0.847, 0.018	0.843, 0.020	0.829, 0.021
95%信賴區間	0.867~0.913	0.861~0.907	0.844~0.899	0.837~0.895	0.811~0.883	0.805~0.882	0.788~0.870
M+10, (est, sd)	0.879, 0.012	0.872, 0.013	0.861, 0.015	0.842, 0.018	0.828, 0.021	0.825, 0.021	0.817, 0.024
95%信賴區間	0.855~0.904	0.846~0.898	0.831~0.892	0.808~0.877	0.787~0.868	0.783~0.867	0.770~0.864

est: 估計值; sd: 估計值之標準差; M:標誌時間



表六 甲狀腺乳突癌(不分年紀)的 $S(t|T_M \leq m, T_D > m)$ 及

$S(t|T_M > m, T_D > m)$ 的估計值及 95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m, T_D > m)$							
M+1, (est, sd)	0.972, 0.065	0.963, 0.054	0.982, 0.011	0.974, 0.005	0.979, 0.009	0.964, 0.017	0.980, 0.010
95%信賴區間	0.844~1.000	0.862~1.000	0.961~1.000	0.965~0.984	0.961~0.997	0.931~0.996	0.960~1.000
M+2, (est, sd)	0.939, 0.108	0.938, 0.113	0.961, 0.013	0.952, 0.010	0.952, 0.020	0.945, 0.023	0.952, 0.016
95%信賴區間	0.727~1.000	0.717~1.000	0.935~0.987	0.933~0.971	0.912~0.992	0.899~0.991	0.919~0.984
M+3, (est, sd)	0.916, 0.163	0.911, 0.156	0.943, 0.016	0.925, 0.021	0.938, 0.026	0.918, 0.028	0.914, 0.034
95%信賴區間	0.596~1.000	0.606~1.000	0.911~0.976	0.883~0.967	0.886~0.989	0.862~0.973	0.847~0.982
M+4, (est, sd)	0.891, 0.201	0.889, 0.199	0.924, 0.023	0.914, 0.026	0.919, 0.030	0.882, 0.043	0.894, 0.043
95%信賴區間	0.497~1.000	0.499~1.000	0.879~0.969	0.864~0.963	0.860~0.977	0.797~0.967	0.809~0.978
M+5, (est, sd)	0.874, 0.235	0.866, 0.214	0.915, 0.026	0.895, 0.029	0.895, 0.041	0.863, 0.051	0.868, 0.049
95%信賴區間	0.413~1.000	0.447~1.000	0.864~0.966	0.838~0.953	0.814~0.976	0.763~0.962	0.773~0.964
M+6, (est, sd)	0.859, 0.244	0.856, 0.222	0.899, 0.028	0.874, 0.038	0.882, 0.046	0.839, 0.057	0.847, 0.054
95%信賴區間	0.381~1.000	0.420~1.000	0.844~0.953	0.800~0.947	0.791~0.972	0.728~0.950	0.742~0.953
M+7, (est, sd)	0.853, 0.246	0.838, 0.229	0.880, 0.033	0.862, 0.043	0.864, 0.051	0.820, 0.060	0.796, 0.062
95%信賴區間	0.372~1.000	0.389~1.000	0.815~0.946	0.778~0.946	0.764~0.964	0.702~0.938	0.675~0.917
M+8, (est, sd)	0.845, 0.248	0.819, 0.232	0.871, 0.038	0.848, 0.047	0.848, 0.057	0.773, 0.071	0.743, 0.079
95%信賴區間	0.359~1.000	0.365~1.000	0.797~0.945	0.755~0.940	0.737~0.959	0.634~0.912	0.588~0.899
M+9, (est, sd)	0.835, 0.248	0.810, 0.232	0.858, 0.041	0.833, 0.052	0.806, 0.067	0.725, 0.088	0.679, 0.084
95%信賴區間	0.359~1.000	0.355~1.000	0.778~0.939	0.731~0.936	0.675~0.938	0.553~0.898	0.514~0.843
M+10, (est, sd)	0.830, 0.250	0.798, 0.234	0.846, 0.045	0.796, 0.064	0.764, 0.087	0.667, 0.093	0.634, 0.094
95%信賴區間	0.341~1.000	0.339~1.000	0.758~0.934	0.671~0.921	0.594~0.934	0.485~0.850	0.451~0.818
$S(t T_M > m, T_D > m)$							
M+1, (est, sd)	0.995, 0.001	0.993, 0.002	0.995, 0.002	0.993, 0.002	0.992, 0.003	0.989, 0.002	0.994, 0.003
95%信賴區間	0.993~0.997	0.989~0.998	0.991~1.000	0.988~0.998	0.986~0.998	0.984~0.994	0.988~1.000
M+2, (est, sd)	0.988, 0.003	0.988, 0.003	0.988, 0.003	0.985, 0.005	0.979, 0.004	0.983, 0.004	0.982, 0.006
95%信賴區間	0.982~0.993	0.982~0.995	0.981~0.994	0.976~0.993	0.971~0.987	0.975~0.991	0.972~0.993
M+3, (est, sd)	0.982, 0.004	0.981, 0.004	0.978, 0.005	0.971, 0.005	0.971, 0.007	0.972, 0.006	0.968, 0.009
95%信賴區間	0.975~0.989	0.973~0.988	0.968~0.989	0.961~0.980	0.957~0.986	0.960~0.984	0.950~0.987
M+4, (est, sd)	0.974, 0.004	0.971, 0.005	0.964, 0.005	0.963, 0.009	0.958, 0.011	0.957, 0.009	0.960, 0.011
95%信賴區間	0.966~0.982	0.962~0.981	0.954~0.974	0.946~0.980	0.936~0.980	0.939~0.976	0.939~0.981
M+5, (est, sd)	0.965, 0.005	0.957, 0.004	0.955, 0.009	0.949, 0.012	0.941, 0.014	0.949, 0.011	0.950, 0.013
95%信賴區間	0.954~0.975	0.948~0.966	0.937~0.974	0.925~0.974	0.913~0.969	0.928~0.970	0.925~0.976
M+6, (est, sd)	0.950, 0.005	0.949, 0.008	0.942, 0.013	0.932, 0.016	0.931, 0.017	0.939, 0.013	0.942, 0.015
95%信賴區間	0.940~0.959	0.933~0.965	0.917~0.966	0.900~0.963	0.898~0.965	0.913~0.965	0.913~0.971
M+7, (est, sd)	0.941, 0.008	0.935, 0.010	0.924, 0.017	0.922, 0.019	0.920, 0.021	0.930, 0.015	0.920, 0.016
95%信賴區間	0.925~0.958	0.915~0.955	0.891~0.957	0.885~0.958	0.879~0.961	0.902~0.959	0.889~0.950
M+8, (est, sd)	0.927, 0.010	0.918, 0.015	0.913, 0.019	0.910, 0.023	0.911, 0.023	0.908, 0.015	0.902, 0.022
95%信賴區間	0.906~0.947	0.888~0.947	0.876~0.951	0.866~0.954	0.865~0.956	0.878~0.938	0.858~0.945
M+9, (est, sd)	0.909, 0.015	0.907, 0.017	0.901, 0.023	0.900, 0.025	0.887, 0.024	0.889, 0.022	0.884, 0.025
95%信賴區間	0.879~0.938	0.873~0.941	0.856~0.947	0.850~0.949	0.841~0.933	0.846~0.931	0.835~0.932
M+10, (est, sd)	0.898, 0.017	0.895, 0.021	0.891, 0.026	0.875, 0.026	0.867, 0.029	0.870, 0.024	0.871, 0.026
95%信賴區間	0.864~0.932	0.854~0.936	0.840~0.942	0.823~0.926	0.810~0.923	0.822~0.917	0.819~0.923

est: 估計值; sd: 估計值之標準差; M:標誌時間

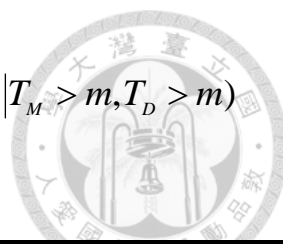


表七 甲狀腺乳突癌(40歲內)的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及

95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m)$							
M+1, (est, sd)	0.981, 0.061	0.983, 0.026	0.984, 0.017	0.977, 0.021	0.979, 0.018	0.974, 0.022	0.976, 0.018
95%信賴區間	0.862~1.000	0.932~1.000	0.950~1.000	0.935~1.000	0.944~1.000	0.930~1.000	0.941~1.000
M+2, (est, sd)	0.981, 0.061	0.983, 0.026	0.979, 0.021	0.977, 0.021	0.974, 0.022	0.974, 0.022	0.964, 0.027
95%信賴區間	0.862~1.000	0.932~1.000	0.938~1.000	0.935~1.000	0.930~1.000	0.930~1.000	0.912~1.000
M+3, (est, sd)	0.981, 0.061	0.979, 0.032	0.979, 0.021	0.972, 0.026	0.974, 0.022	0.961, 0.032	0.958, 0.032
95%信賴區間	0.862~1.000	0.917~1.000	0.938~1.000	0.920~1.000	0.930~1.000	0.898~1.000	0.896~1.000
M+4, (est, sd)	0.976, 0.070	0.979, 0.032	0.973, 0.025	0.972, 0.026	0.961, 0.032	0.954, 0.038	0.950, 0.042
95%信賴區間	0.839~1.000	0.917~1.000	0.924~1.000	0.920~1.000	0.898~1.000	0.881~1.000	0.868~1.000
M+5, (est, sd)	0.976, 0.070	0.975, 0.041	0.973, 0.025	0.958, 0.037	0.954, 0.038	0.947, 0.048	0.950, 0.042
95%信賴區間	0.839~1.000	0.894~1.000	0.924~1.000	0.885~1.000	0.881~1.000	0.852~1.000	0.868~1.000
M+6, (est, sd)	0.970, 0.084	0.975, 0.041	0.960, 0.036	0.950, 0.043	0.947, 0.048	0.947, 0.048	0.950, 0.042
95%信賴區間	0.806~1.000	0.894~1.000	0.890~1.000	0.865~1.000	0.852~1.000	0.852~1.000	0.868~1.000
M+7, (est, sd)	0.970, 0.084	0.964, 0.050	0.953, 0.042	0.942, 0.056	0.947, 0.048	0.947, 0.048	0.927, 0.051
95%信賴區間	0.806~1.000	0.866~1.000	0.871~1.000	0.832~1.000	0.852~1.000	0.852~1.000	0.827~1.000
M+8, (est, sd)	0.958, 0.102	0.958, 0.054	0.946, 0.055	0.942, 0.056	0.947, 0.048	0.924, 0.057	0.927, 0.051
95%信賴區間	0.758~1.000	0.851~1.000	0.839~1.000	0.832~1.000	0.852~1.000	0.812~1.000	0.827~1.000
M+9, (est, sd)	0.951, 0.112	0.952, 0.061	0.946, 0.055	0.942, 0.056	0.924, 0.057	0.924, 0.057	0.886, 0.081
95%信賴區間	0.731~1.000	0.833~1.000	0.839~1.000	0.832~1.000	0.812~1.000	0.812~1.000	0.727~1.000
M+10, (est, sd)	0.944, 0.121	0.952, 0.061	0.946, 0.055	0.919, 0.064	0.924, 0.057	0.881, 0.088	0.862, 0.087
95%信賴區間	0.707~1.000	0.833~1.000	0.839~1.000	0.794~1.000	0.812~1.000	0.709~1.000	0.691~1.000
$S(t T_M > m)$							
M+1, (est, sd)	0.992, 0.004	0.992, 0.004	0.992, 0.004	0.990, 0.004	0.990, 0.004	0.988, 0.005	0.933, 0.009
95%信賴區間	0.985~0.999	0.985~0.999	0.985~0.999	0.982~0.998	0.982~0.998	0.979~0.998	0.915~0.950
M+2, (est, sd)	0.992, 0.004	0.992, 0.004	0.990, 0.004	0.990, 0.004	0.988, 0.005	0.988, 0.005	0.922, 0.010
95%信賴區間	0.985~0.999	0.985~0.999	0.982~0.998	0.982~0.998	0.979~0.998	0.979~0.998	0.902~0.942
M+3, (est, sd)	0.992, 0.004	0.990, 0.004	0.990, 0.004	0.988, 0.005	0.988, 0.005	0.983, 0.007	0.909, 0.012
95%信賴區間	0.985~0.999	0.982~0.998	0.982~0.998	0.979~0.998	0.979~0.998	0.970~0.995	0.886~0.931
M+4, (est, sd)	0.990, 0.004	0.990, 0.004	0.988, 0.005	0.988, 0.005	0.983, 0.007	0.979, 0.008	0.901, 0.012
95%信賴區間	0.982~0.998	0.982~0.998	0.978~0.998	0.979~0.998	0.970~0.995	0.965~0.994	0.877~0.925
M+5, (est, sd)	0.990, 0.004	0.988, 0.005	0.988, 0.005	0.985, 0.007	0.979, 0.008	0.976, 0.008	0.892, 0.014
95%信賴區間	0.982~0.998	0.978~0.997	0.978~0.998	0.970~0.995	0.965~0.994	0.960~0.992	0.865~0.919
M+6, (est, sd)	0.988, 0.005	0.988, 0.005	0.982, 0.007	0.980, 0.007	0.976, 0.008	0.976, 0.008	0.883, 0.014
95%信賴區間	0.978~0.997	0.978~0.997	0.969~0.995	0.965~0.994	0.960~0.992	0.960~0.992	0.856~0.912
M+7, (est, sd)	0.988, 0.005	0.982, 0.007	0.979, 0.008	0.976, 0.008	0.976, 0.008	0.976, 0.008	0.863, 0.017
95%信賴區間	0.978~0.997	0.969~0.995	0.964~0.993	0.961~0.991	0.960~0.992	0.960~0.992	0.830~0.897
M+8, (est, sd)	0.982, 0.007	0.979, 0.008	0.976, 0.008	0.976, 0.008	0.976, 0.008	0.963, 0.012	0.846, 0.019
95%信賴區間	0.969~0.995	0.964~0.993	0.960~0.991	0.961~0.991	0.960~0.992	0.940~0.987	0.808~0.884
M+9, (est, sd)	0.979, 0.007	0.975, 0.008	0.976, 0.008	0.976, 0.008	0.963, 0.012	0.963, 0.012	0.829, 0.021
95%信賴區間	0.964~0.993	0.959~0.991	0.960~0.991	0.961~0.991	0.940~0.987	0.940~0.987	0.788~0.870
M+10, (est, sd)	0.975, 0.008	0.975, 0.008	0.976, 0.008	0.963, 0.012	0.963, 0.012	0.946, 0.016	0.817, 0.024
95%信賴區間	0.959~0.991	0.959~0.991	0.960~0.991	0.940~0.987	0.940~0.987	0.915~0.978	0.770~0.864

est: 估計值; sd: 估計值之標準差; M:標誌時間



表八 甲狀腺乳突癌(40歲內)的 $S(t|T_M \leq m, T_D > m)$ 及 $S(t|T_M > m, T_D > m)$

的估計值及 95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m, T_D > m)$							
M+1, (est, sd)	0.992, 0.047	1.000	1.000	0.995, 0.008	1.000	0.995, 0.009	1.000
95%信賴區間	0.900~1.000	1.000~1.000	1.000~1.000	0.980~1.000	1.000~1.000	0.977~1.000	1.000~1.000
M+2, (est, sd)	0.992, 0.047	1.000	0.995, 0.007	0.995, 0.008	0.995, 0.009	0.995, 0.009	0.988, 0.015
95%信賴區間	0.900~1.000	1.000~1.000	0.981~1.000	0.980~1.000	0.977~1.000	0.977~1.000	0.959~1.000
M+3, (est, sd)	0.992, 0.047	0.996, 0.010	0.995, 0.007	0.989, 0.014	0.995, 0.009	0.982, 0.021	0.981, 0.020
95%信賴區間	0.900~1.000	0.976~1.000	0.981~1.000	0.961~1.000	0.977~1.000	0.940~1.000	0.941~1.000
M+4, (est, sd)	0.987, 0.057	0.996, 0.010	0.990, 0.013	0.989, 0.014	0.982, 0.021	0.975, 0.027	0.973, 0.032
95%信賴區間	0.876~1.000	0.976~1.000	0.963~1.000	0.961~1.000	0.940~1.000	0.921~1.000	0.911~1.000
M+5, (est, sd)	0.987, 0.057	0.991, 0.026	0.990, 0.013	0.975, 0.027	0.975, 0.027	0.967, 0.039	0.973, 0.032
95%信賴區間	0.876~1.000	0.940~1.000	0.963~1.000	0.922~1.000	0.921~1.000	0.890~1.000	0.911~1.000
M+6, (est, sd)	0.981, 0.074	0.991, 0.026	0.976, 0.026	0.967, 0.034	0.967, 0.039	0.967, 0.039	0.973, 0.032
95%信賴區間	0.837~1.000	0.940~1.000	0.926~1.000	0.902~1.000	0.890~1.000	0.890~1.000	0.911~1.000
M+7, (est, sd)	0.981, 0.074	0.980, 0.038	0.969, 0.032	0.959, 0.048	0.967, 0.039	0.967, 0.039	0.950, 0.042
95%信賴區間	0.837~1.000	0.905~1.000	0.906~1.000	0.866~1.000	0.890~1.000	0.890~1.000	0.868~1.000
M+8, (est, sd)	0.969, 0.095	0.974, 0.044	0.962, 0.046	0.959, 0.048	0.967, 0.039	0.943, 0.048	0.950, 0.042
95%信賴區間	0.783~1.000	0.888~1.000	0.871~1.000	0.866~1.000	0.890~1.000	0.849~1.000	0.868~1.000
M+9, (est, sd)	0.962, 0.107	0.968, 0.052	0.962, 0.046	0.959, 0.048	0.943, 0.048	0.943, 0.048	0.908, 0.075
95%信賴區間	0.752~1.000	0.866~1.000	0.871~1.000	0.866~1.000	0.849~1.000	0.849~1.000	0.761~1.000
M+10, (est, sd)	0.955, 0.117	0.968, 0.052	0.962, 0.046	0.936, 0.056	0.943, 0.048	0.900, 0.082	0.883, 0.081
95%信賴區間	0.725~1.000	0.866~1.000	0.871~1.000	0.826~1.000	0.849~1.000	0.740~1.000	0.724~1.000
$S(t T_M > m, T_D > m)$							
M+1, (est, sd)	0.997, 0.002	1.000	1.000	0.998, 0.002	1.000	0.998, 0.002	1.000
95%信賴區間	0.993~1.000	1.000~1.000	1.000~1.000	0.994~1.000	1.000~1.000	0.993~1.000	1.000~1.000
M+2, (est, sd)	0.997, 0.002	1.000	0.998, 0.002	0.998, 0.002	0.998, 0.002	0.998, 0.002	0.994, 0.004
95%信賴區間	0.993~1.000	1.000~1.000	0.994~1.000	0.994~1.000	0.993~1.000	0.993~1.000	0.986~1.000
M+3, (est, sd)	0.997, 0.002	0.998, 0.002	0.998, 0.002	0.996, 0.003	0.998, 0.002	0.992, 0.005	0.991, 0.005
95%信賴區間	0.993~1.000	0.994~1.000	0.994~1.000	0.990~1.000	0.993~1.000	0.983~1.000	0.981~1.000
M+4, (est, sd)	0.995, 0.003	0.998, 0.002	0.996, 0.003	0.996, 0.003	0.992, 0.005	0.989, 0.006	0.988, 0.006
95%信賴區間	0.990~1.000	0.994~1.000	0.990~1.000	0.990~1.000	0.983~1.000	0.978~1.000	0.976~0.999
M+5, (est, sd)	0.995, 0.003	0.996, 0.003	0.996, 0.003	0.990, 0.005	0.989, 0.006	0.986, 0.006	0.988, 0.006
95%信賴區間	0.990~1.000	0.990~1.000	0.990~1.000	0.980~1.000	0.978~1.000	0.973~0.998	0.976~0.999
M+6, (est, sd)	0.993, 0.004	0.996, 0.003	0.990, 0.005	0.987, 0.006	0.986, 0.006	0.986, 0.006	0.988, 0.006
95%信賴區間	0.986~1.000	0.990~1.000	0.980~1.000	0.975~0.999	0.973~0.998	0.973~0.998	0.976~0.999
M+7, (est, sd)	0.993, 0.004	0.990, 0.005	0.987, 0.006	0.984, 0.007	0.986, 0.006	0.986, 0.006	0.975, 0.011
95%信賴區間	0.986~1.000	0.979~1.000	0.975~0.999	0.971~0.997	0.973~0.998	0.973~0.998	0.954~0.996
M+8, (est, sd)	0.987, 0.006	0.987, 0.006	0.984, 0.007	0.984, 0.007	0.986, 0.006	0.973, 0.011	0.975, 0.011
95%信賴區間	0.976~0.998	0.974~0.999	0.971~0.997	0.971~0.997	0.973~0.998	0.951~0.995	0.954~0.996
M+9, (est, sd)	0.984, 0.007	0.983, 0.007	0.984, 0.007	0.984, 0.007	0.973, 0.011	0.973, 0.011	0.958, 0.015
95%信賴區間	0.971~0.997	0.969~0.997	0.971~0.997	0.971~0.997	0.951~0.995	0.951~0.995	0.928~0.988
M+10, (est, sd)	0.980, 0.007	0.983, 0.007	0.984, 0.007	0.971, 0.011	0.973, 0.011	0.956, 0.016	0.948, 0.019
95%信賴區間	0.965~0.995	0.969~0.997	0.971~0.997	0.949~0.993	0.951~0.995	0.925~0.986	0.912~0.985

est: 估計值; sd: 估計值之標準差; M: 標誌時間



表九 甲狀腺乳突癌(40歲以上)的 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的估計值及

95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m)$							
M+1, (est, sd)	0.758, 0.183	0.748, 0.149	0.752, 0.130	0.716, 0.126	0.718, 0.100	0.686, 0.089	0.660, 0.088
95%信賴區間	0.400~1.000	0.455~1.000	0.498~1.000	0.469~0.964	0.521~0.914	0.512~0.859	0.489~0.832
M+2, (est, sd)	0.700, 0.221	0.710, 0.169	0.715, 0.145	0.675, 0.142	0.667, 0.115	0.651, 0.095	0.621, 0.095
95%信賴區間	0.266~1.000	0.378~1.000	0.432~0.998	0.397~0.954	0.443~0.892	0.464~0.839	0.435~0.808
M+3, (est, sd)	0.660, 0.247	0.673, 0.187	0.681, 0.159	0.632, 0.157	0.639, 0.120	0.612, 0.103	0.562, 0.107
95%信賴區間	0.176~1.000	0.306~1.000	0.370~0.992	0.325~0.939	0.403~0.875	0.409~0.814	0.353~0.771
M+4, (est, sd)	0.622, 0.270	0.639, 0.203	0.650, 0.168	0.611, 0.161	0.610, 0.126	0.552, 0.115	0.533, 0.111
95%信賴區間	0.092~1.000	0.241~1.000	0.320~0.980	0.295~0.927	0.363~0.857	0.327~0.777	0.315~0.750
M+5, (est, sd)	0.593, 0.288	0.611, 0.214	0.635, 0.171	0.587, 0.167	0.568, 0.133	0.523, 0.119	0.486, 0.117
95%信賴區間	0.029~1.000	0.192~1.000	0.297~0.969	0.261~0.913	0.307~0.829	0.289~0.757	0.257~0.715
M+6, (est, sd)	0.576, 0.297	0.596, 0.217	0.611, 0.176	0.554, 0.172	0.546, 0.136	0.477, 0.125	0.450, 0.122
95%信賴區間	0.000~1.000	0.170~1.000	0.267~0.956	0.217~0.891	0.279~0.813	0.232~0.722	0.211~0.688
M+7, (est, sd)	0.569, 0.298	0.577, 0.222	0.581, 0.180	0.538, 0.173	0.509, 0.141	0.441, 0.130	0.384, 0.127
95%信賴區間	0.000~1.000	0.142~1.000	0.228~0.935	0.199~0.878	0.233~0.785	0.187~0.795	0.135~0.632
M+8, (est, sd)	0.563, 0.298	0.554, 0.226	0.568, 0.180	0.512, 0.175	0.476, 0.144	0.377, 0.133	0.289, 0.136
95%信賴區間	0.000~1.000	0.111~1.000	0.214~0.922	0.168~0.858	0.194~0.759	0.116~0.639	0.022~0.556
M+9, (est, sd)	0.554, 0.297	0.544, 0.226	0.545, 0.181	0.486, 0.178	0.414, 0.150	0.287, 0.141	0.213, 0.144
95%信賴區間	0.000~1.000	0.102~0.987	0.191~0.900	0.137~0.835	0.121~0.707	0.012~0.563	0.000~0.495
M+10, (est, sd)	0.548, 0.296	0.528, 0.225	0.522, 0.181	0.439, 0.179	0.328, 0.159	0.215, 0.147	0.156, 0.145
95%信賴區間	0.000~1.000	0.086~0.970	0.166~0.877	0.088~0.790	0.016~0.641	0.000~0.503	0.000~0.440
$S(t T_M > m)$							
M+1, (est, sd)	0.959, 0.007	0.946, 0.009	0.938, 0.010	0.928, 0.011	0.914, 0.013	0.898, 0.014	0.887, 0.015
95%信賴區間	0.945~0.972	0.929~0.963	0.918~0.957	0.906~0.950	0.889~0.938	0.870~0.926	0.857~0.917
M+2, (est, sd)	0.945, 0.009	0.937, 0.010	0.926, 0.011	0.913, 0.012	0.893, 0.014	0.887, 0.015	0.872, 0.018
95%信賴區間	0.928~0.963	0.918~0.957	0.904~0.948	0.888~0.937	0.865~0.921	0.857~0.917	0.837~0.906
M+3, (est, sd)	0.936, 0.010	0.925, 0.011	0.909, 0.013	0.889, 0.014	0.879, 0.016	0.872, 0.017	0.849, 0.020
95%信賴區間	0.916~0.955	0.903~0.947	0.885~0.934	0.862~0.917	0.849~0.910	0.838~0.906	0.810~0.889
M+4, (est, sd)	0.922, 0.011	0.908, 0.012	0.884, 0.015	0.874, 0.016	0.861, 0.018	0.849, 0.020	0.837, 0.022
95%信賴區間	0.900~0.945	0.883~0.932	0.855~0.913	0.844~0.905	0.826~0.896	0.810~0.888	0.794~0.880
M+5, (est, sd)	0.904, 0.013	0.882, 0.015	0.869, 0.016	0.855, 0.018	0.833, 0.021	0.837, 0.022	0.820, 0.024
95%信賴區間	0.880~0.929	0.853~0.911	0.837~0.900	0.820~0.890	0.792~0.874	0.794~0.879	0.774~0.867
M+6, (est, sd)	0.878, 0.015	0.866, 0.016	0.849, 0.018	0.825, 0.021	0.819, 0.023	0.819, 0.024	0.804, 0.025
95%信賴區間	0.849~0.907	0.834~0.898	0.813~0.885	0.784~0.865	0.775~0.863	0.773~0.866	0.755~0.853
M+7, (est, sd)	0.861, 0.016	0.846, 0.018	0.818, 0.021	0.809, 0.023	0.799, 0.024	0.803, 0.025	0.775, 0.030
95%信賴區間	0.829~0.893	0.810~0.882	0.777~0.860	0.765~0.853	0.751~0.846	0.754~0.852	0.717~0.833
M+8, (est, sd)	0.840, 0.019	0.815, 0.021	0.802, 0.023	0.787, 0.024	0.781, 0.026	0.773, 0.029	0.743, 0.033
95%信賴區間	0.803~0.877	0.774~0.856	0.757~0.847	0.739~0.834	0.729~0.832	0.715~0.830	0.677~0.808
M+9, (est, sd)	0.808, 0.022	0.798, 0.023	0.779, 0.025	0.767, 0.026	0.749, 0.031	0.739, 0.033	0.722, 0.036
95%信賴區間	0.765~0.851	0.754~0.843	0.731~0.828	0.716~0.818	0.689~0.809	0.674~0.803	0.652~0.793
M+10, (est, sd)	0.792, 0.023	0.775, 0.025	0.759, 0.027	0.732, 0.031	0.712, 0.034	0.717, 0.036	0.705, 0.041
95%信賴區間	0.746~0.838	0.726~0.823	0.706~0.812	0.671~0.793	0.645~0.778	0.647~0.787	0.625~0.786

est: 估計值; sd: 估計值之標準差; M:標誌時間

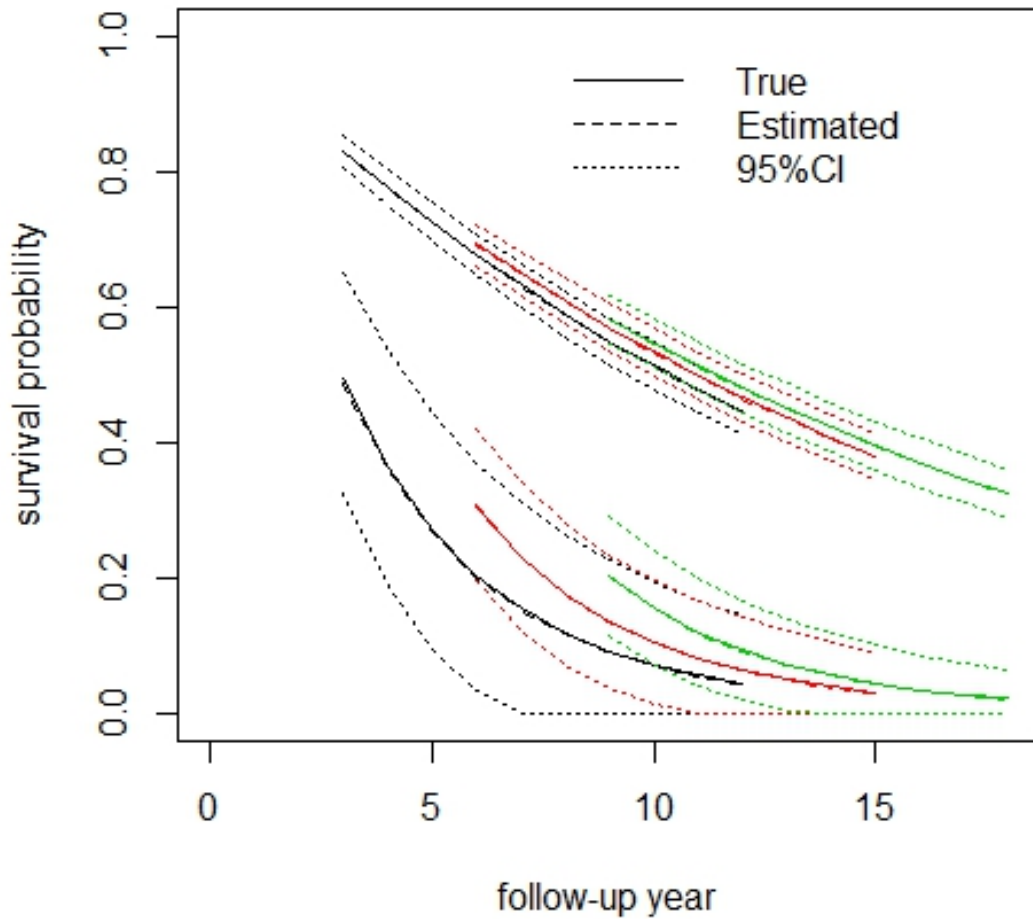


表十 甲狀腺乳突癌(40歲以上)的 $S(t|T_M \leq m, T_D > m)$ 及

$S(t|T_M > m, T_D > m)$ 的估計值及 95%信賴區間

標誌時間 存活時間	M=2	M=3	M=4	M=5	M=6	M=7	M=8
$S(t T_M \leq m, T_D > m)$							
M+1, (est, sd)	0.952, 0.073	0.927, 0.076	0.955, 0.048	0.943, 0.049	0.944, 0.039	0.918, 0.042	0.950, 0.029
95%信賴區間	0.810~1.000	0.778~1.000	0.861~1.000	0.848~1.000	0.868~1.000	0.835~1.000	0.894~1.000
M+2, (est, sd)	0.879, 0.159	0.880, 0.118	0.908, 0.84	0.889, 0.087	0.878, 0.072	0.871, 0.058	0.894, 0.049
95%信賴區間	0.568~1.000	0.650~1.000	0.744~1.000	0.718~1.000	0.737~1.000	0.758~0.984	0.798~0.991
M+3, (est, sd)	0.829, 0.215	0.834, 0.155	0.865, 0.118	0.832, 0.125	0.841, 0.086	0.819, 0.075	0.809, 0.079
95%信賴區間	0.407~1.000	0.531~1.000	0.633~1.000	0.588~1.000	0.673~1.000	0.671~0.966	0.654~0.963
M+4, (est, sd)	0.781, 0.268	0.792, 0.188	0.825, 0.143	0.805, 0.137	0.803, 0.116	0.739, 0.100	0.767, 0.090
95%信賴區間	0.256~1.000	0.423~1.000	0.546~1.000	0.537~1.000	0.608~0.997	0.542~0.935	0.590~0.943
M+5, (est, sd)	0.745, 0.306	0.757, 0.211	0.804, 0.151	0.773, 0.152	0.747, 0.116	0.700, 0.110	0.700, 0.106
95%信賴區間	0.145~1.000	0.343~1.000	0.508~1.000	0.475~1.000	0.520~0.975	0.484~0.915	0.492~0.908
M+6, (est, sd)	0.723, 0.327	0.739, 0.219	0.776, 0.163	0.730, 0.167	0.718, 0.124	0.638, 0.123	0.647, 0.121
95%信賴區間	0.083~1.000	0.310~1.000	0.456~1.000	0.403~1.000	0.475~0.961	0.396~0.880	0.409~0.885
M+7, (est, sd)	0.715, 0.330	0.715, 0.229	0.739, 0.176	0.709, 0.171	0.669, 0.136	0.590, 0.135	0.552, 0.139
95%信賴區間	0.068~1.000	0.266~1.000	0.394~1.000	0.374~1.000	0.403~0.936	0.326~0.855	0.280~0.824
M+8, (est, sd)	0.707, 0.332	0.686, 0.240	0.721, 0.177	0.675, 0.179	0.627, 0.146	0.505, 0.146	0.417, 0.165
95%信賴區間	0.056~1.000	0.217~1.000	0.374~1.000	0.325~1.000	0.341~0.912	0.218~0.792	0.093~0.740
M+9, (est, sd)	0.695, 0.331	0.675, 0.240	0.693, 0.180	0.641, 0.187	0.545, 0.161	0.385, 0.165	0.307, 0.185
95%信賴區間	0.046~1.000	0.205~1.000	0.340~1.000	0.274~1.000	0.230~0.859	0.061~0.708	0.000~0.669
M+10, (est, sd)	0.689, 0.331	0.654, 0.240	0.663, 0.184	0.578, 0.194	0.431, 0.184	0.288, 0.180	0.224, 0.192
95%信賴區間	0.040~1.000	0.184~1.000	0.303~1.000	0.198~0.959	0.071~0.792	0.000~0.641	0.000~0.600
$S(t T_M > m, T_D > m)$							
M+1, (est, sd)	0.993, 0.003	0.987, 0.004	0.991, 0.003	0.988, 0.004	0.985, 0.005	0.981, 0.006	0.988, 0.005
95%信賴區間	0.987~0.998	0.979~0.995	0.984~0.998	0.980~0.997	0.975~0.995	0.970~0.993	0.978~0.998
M+2, (est, sd)	0.979, 0.003	0.978, 0.006	0.978, 0.006	0.972, 0.007	0.962, 0.009	0.969, 0.008	0.971, 0.009
95%信賴區間	0.968~0.990	0.967~0.989	0.967~0.990	0.959~0.986	0.945~0.979	0.954~0.985	0.953~0.989
M+3, (est, sd)	0.969, 0.007	0.965, 0.008	0.961, 0.008	0.947, 0.010	0.948, 0.011	0.953, 0.011	0.946, 0.013
95%信賴區間	0.956~0.983	0.950~0.980	0.945~0.977	0.928~0.967	0.926~0.969	0.931~0.974	0.921~0.971
M+4, (est, sd)	0.955, 0.008	0.947, 0.009	0.934, 0.012	0.931, 0.012	0.928, 0.014	0.928, 0.014	0.932, 0.015
95%信賴區間	0.939~0.972	0.928~0.965	0.912~0.957	0.907~0.956	0.901~0.955	0.900~0.956	0.902~0.962
M+5, (est, sd)	0.937, 0.010	0.920, 0.012	0.918, 0.014	0.911, 0.015	0.898, 0.017	0.914, 0.017	0.914, 0.018
95%信賴區間	0.917~0.957	0.896~0.944	0.8911~0.944	0.881~0.940	0.864~0.932	0.881~0.947	0.878~0.949
M+6, (est, sd)	0.909, 0.013	0.904, 0.014	0.897, 0.016	0.879, 0.019	0.883, 0.020	0.895, 0.020	0.896, 0.020
95%信賴區間	0.884~0.935	0.876~0.932	0.865~0.928	0.842~0.915	0.844~0.921	0.857~0.934	0.856~0.935
M+7, (est, sd)	0.892, 0.015	0.882, 0.017	0.865, 0.019	0.862, 0.021	0.861, 0.022	0.877, 0.021	0.863, 0.027
95%信賴區間	0.863~0.921	0.850~0.915	0.827~0.902	0.821~0.903	0.817~0.904	0.836~0.919	0.811~0.916
M+8, (est, sd)	0.870, 0.017	0.850, 0.020	0.848, 0.022	0.838, 0.023	0.841, 0.025	0.845, 0.027	0.827, 0.032
95%信賴區間	0.836~0.904	0.811~0.888	0.805~0.890	0.792~0.883	0.793~0.890	0.791~0.898	0.764~0.890
M+9, (est, sd)	0.837, 0.021	0.833, 0.022	0.823, 0.024	0.817, 0.025	0.807, 0.030	0.807, 0.032	0.805, 0.035
95%信賴區間	0.796~0.878	0.790~0.876	0.777~0.870	0.767~0.867	0.749~0.866	0.744~0.870	0.736~0.873
M+10, (est, sd)	0.820, 0.023	0.808, 0.024	0.802, 0.027	0.780, 0.031	0.767, 0.034	0.783, 0.035	0.785, 0.042
95%信賴區間	0.775~0.864	0.761~0.855	0.750~0.854	0.719~0.841	0.700~0.833	0.714~0.853	0.704~0.867

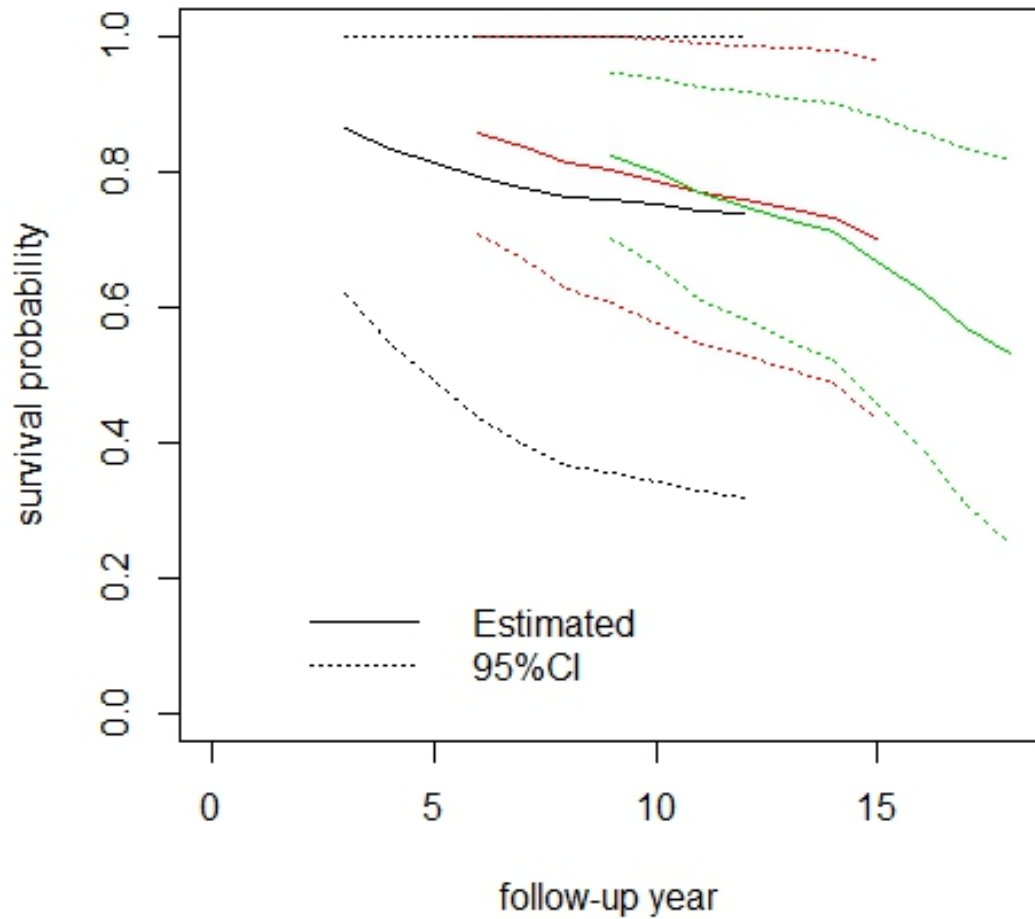
est: 估計值; sd: 估計值之標準差; M:標誌時間



圖三 模擬資料的標誌時間為 2(黑線)、5(紅線)、8(綠線)年之

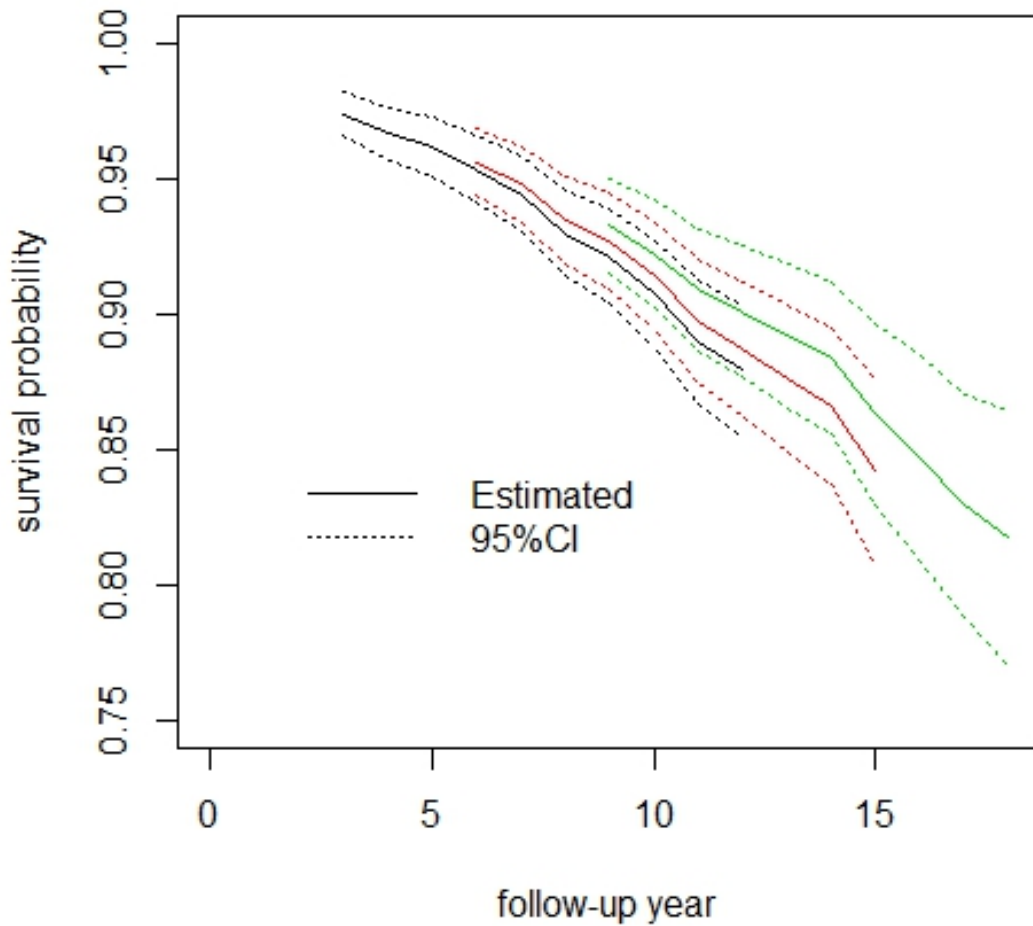
$S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 及真值、估計值及 95%信賴區間。

True: 真值; Estimated: 估計值; 95% CI: 估計值的 95%信賴區間。



圖四 甲狀腺乳突癌患者的標誌時間為不大於2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m)$ 估計值及95%信賴區間。

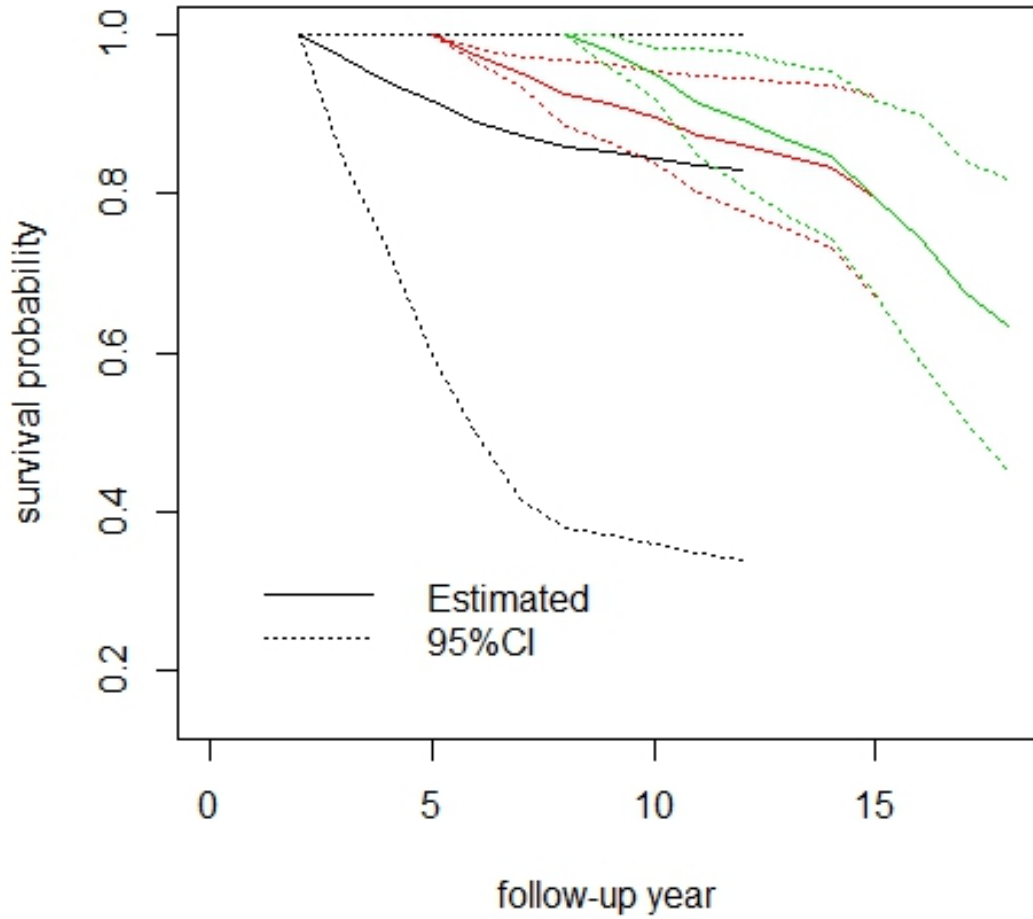
Estimated: 估計值; 95% CI: 估計值的95%信賴區間。



圖五 甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、

8(綠線)年之 $S(t|T_M > m)$ 估計值及 95%信賴區間。

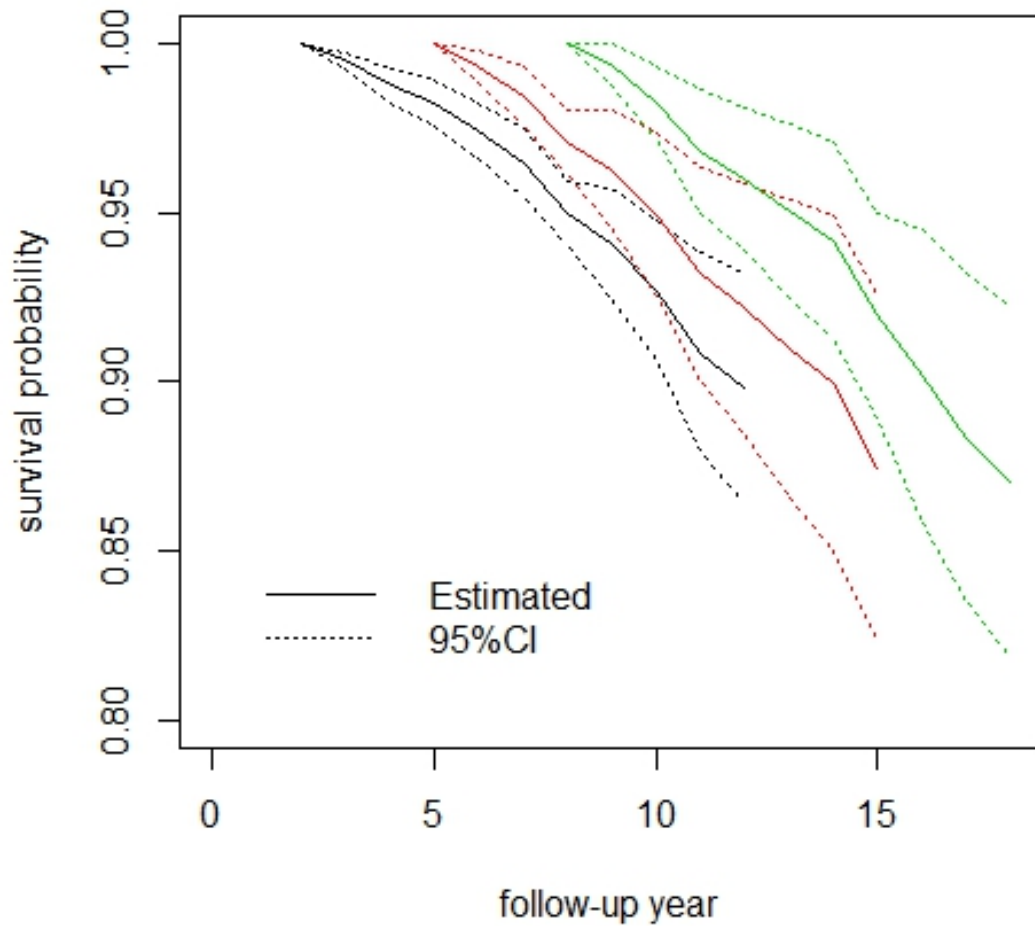
Estimated: 估計值; 95% CI: 估計值的 95%信賴區間。



圖六 甲狀腺乳突癌患者的標誌時間為不大於2(黑線)、5(紅線)、

8(綠線)年之 $S(t|T_M \leq m, T_D > m)$ 估計值及95%信賴區間。

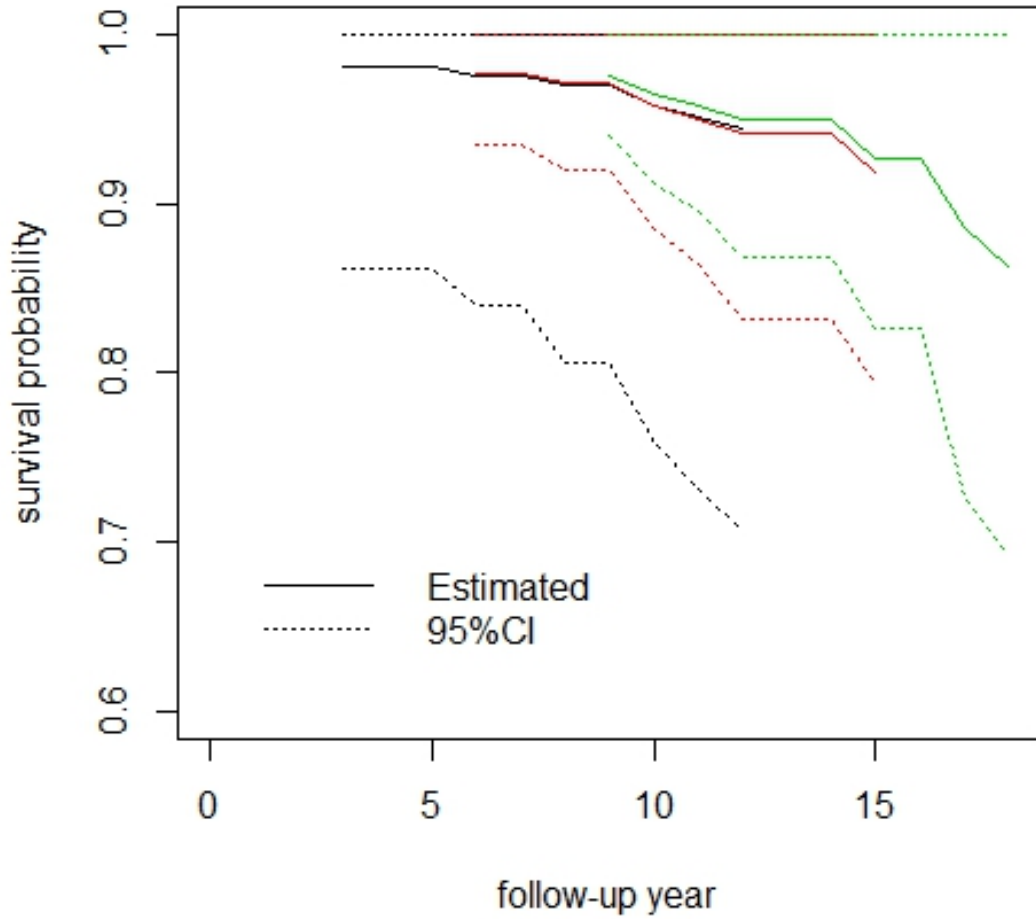
Estimated: 估計值;95% CI: 估計值的95%信賴區間。



圖七 甲狀腺乳突癌患者的標誌時間為大於2(黑線)、5(紅線)、8(綠

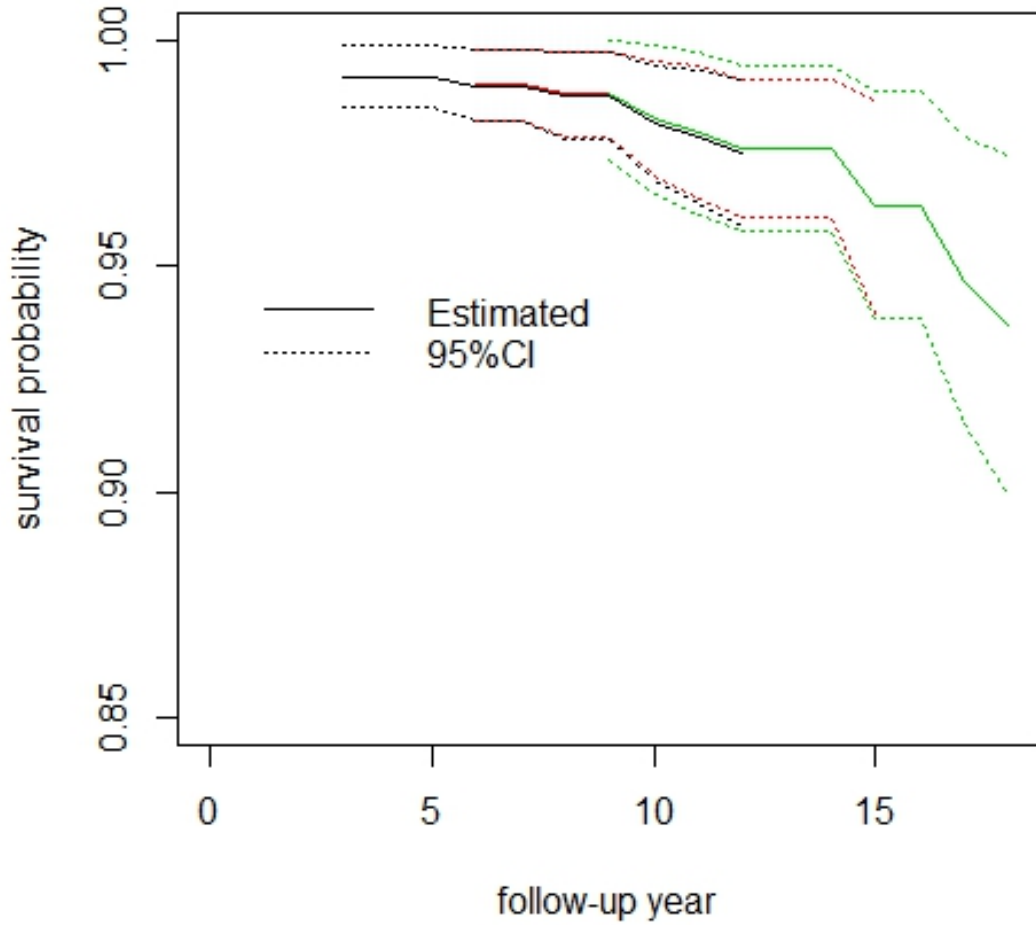
線)年之 $S(t|T_M > m, T_D > m)$ 估計值及 95%信賴區間。

Estimated: 估計值; 95% CI: 估計值的 95%信賴區間。

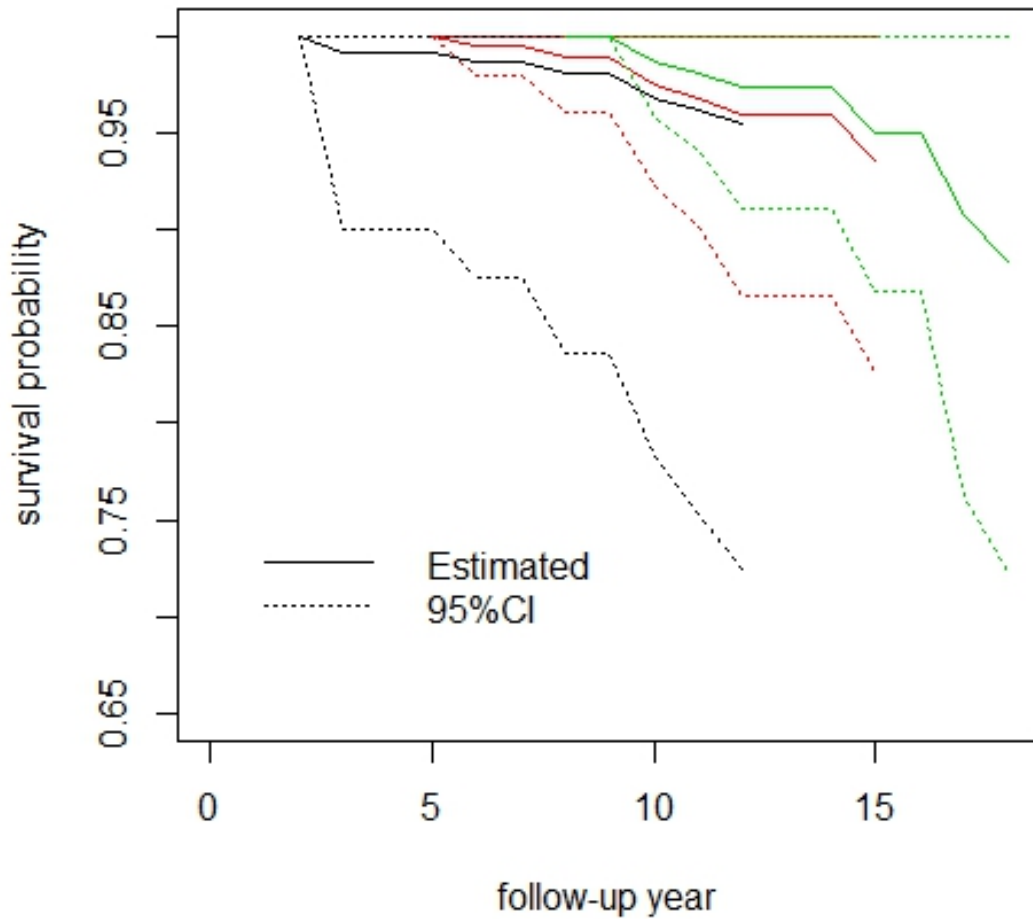


圖八 甲狀腺乳突癌患者(40歲內)的標誌時間為不大於2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m)$ 估計值及95%信賴區間。

Estimated: 估計值; 95% CI: 估計值的95%信賴區間。

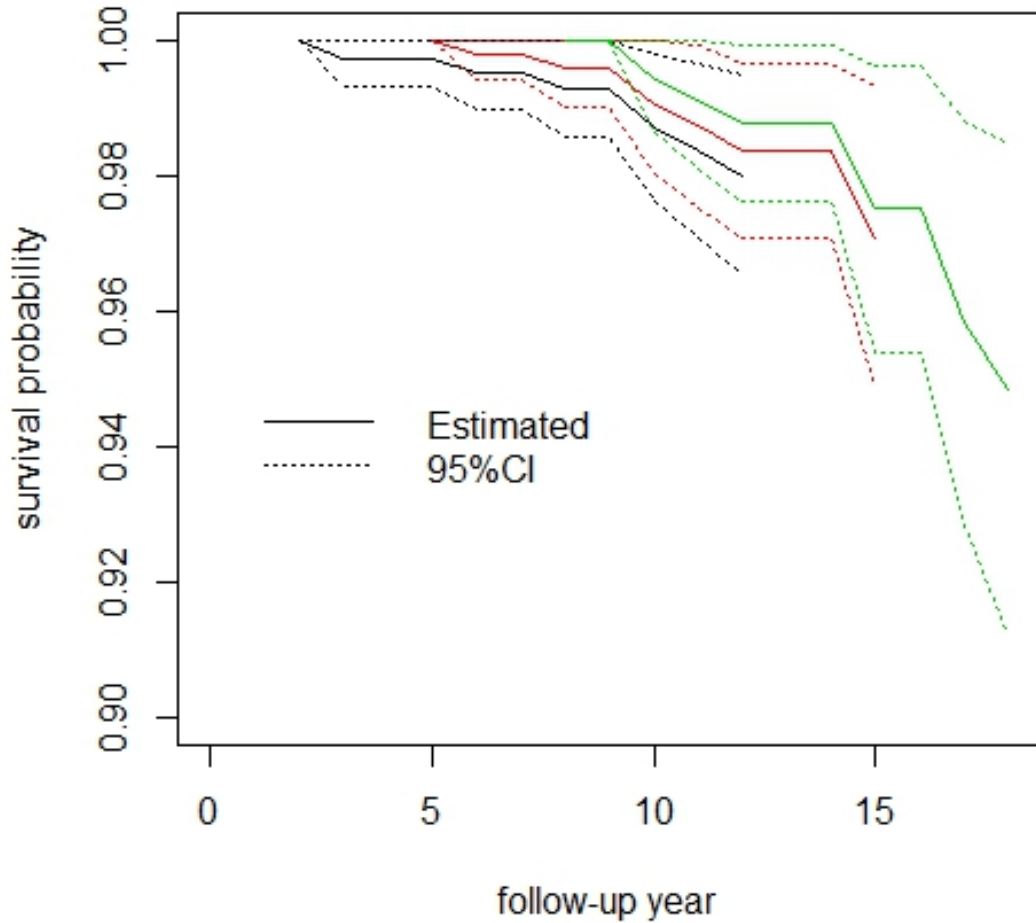


圖九 甲狀腺乳突癌患者(40歲內)的標誌時間為大於2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M > m)$ 估計值及95%信賴區間。
 Estimated: 估計值;95% CI: 估計值的95%信賴區間。



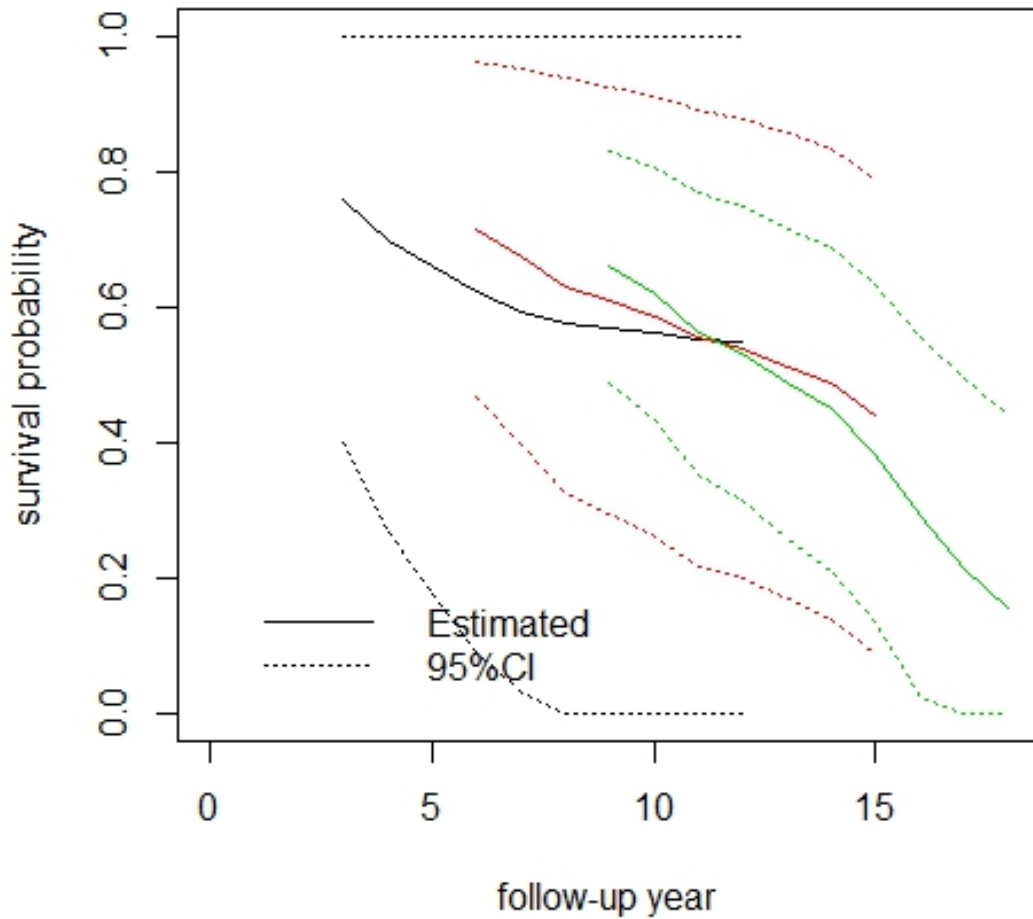
圖十 甲狀腺乳突癌患者(40歲內)的標誌時間為不大於2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m, T_D > m)$ 估計值及95%信賴區間。

Estimated: 估計值; 95% CI: 估計值的95%信賴區間。



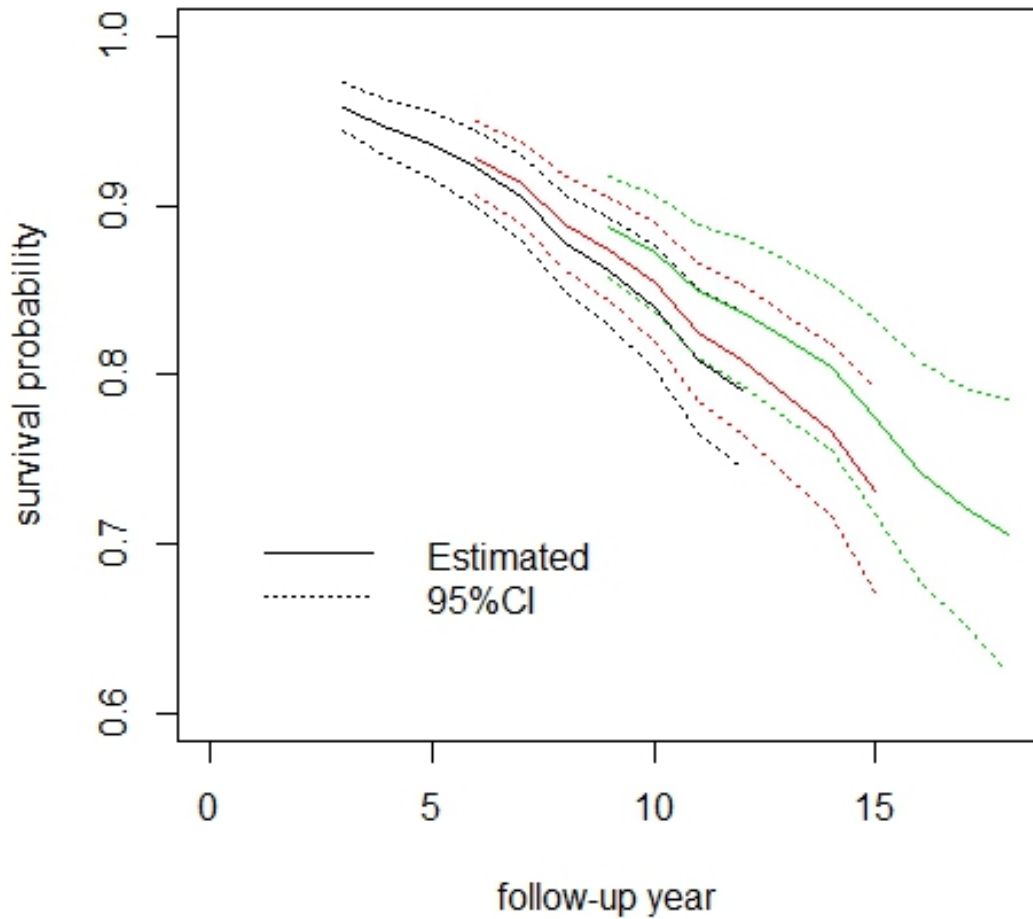
圖十一 甲狀腺乳突癌患者(40歲內)的標誌時間為大於2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M > m, T_D > m)$ 估計值及95%信賴區間。

Estimated: 估計值; 95% CI: 估計值的95%信賴區間。



圖十二 40 歲以上之甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m)$ 估計值及 95%信賴區間。

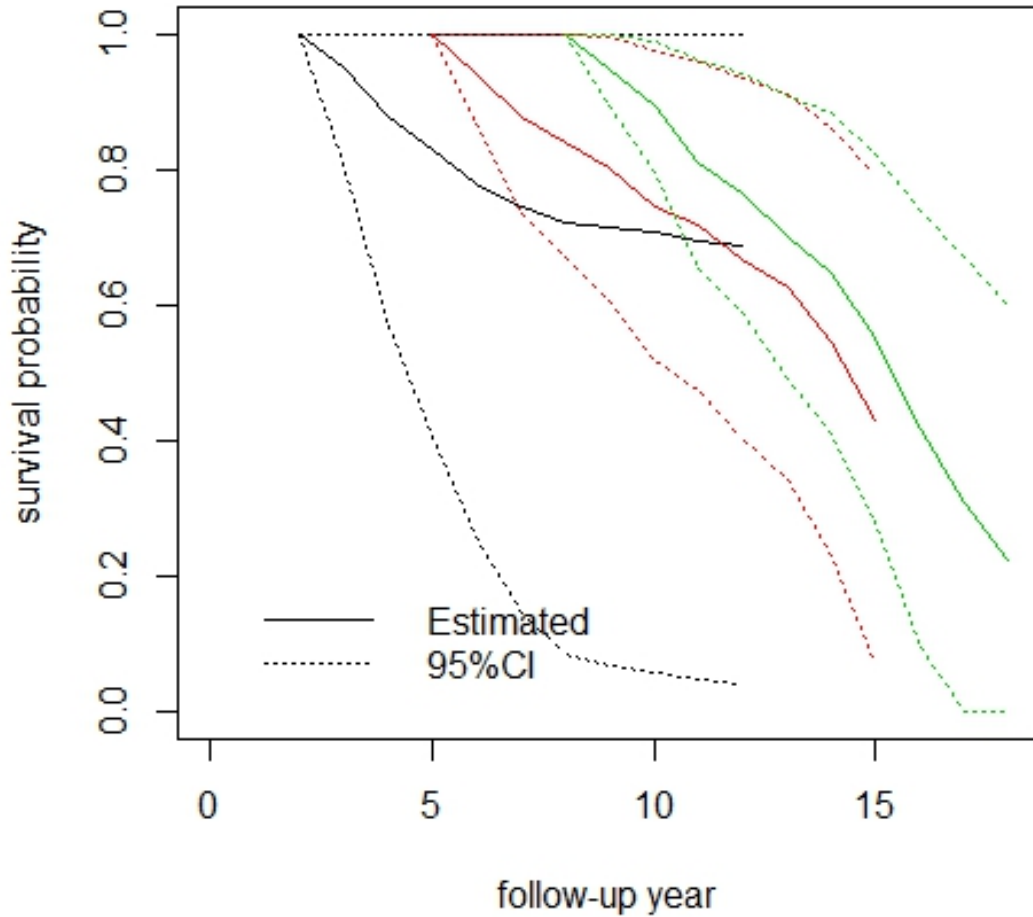
Estimated: 估計值; 95% CI: 估計值的 95%信賴區間。



圖十三 40 歲以上之甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、

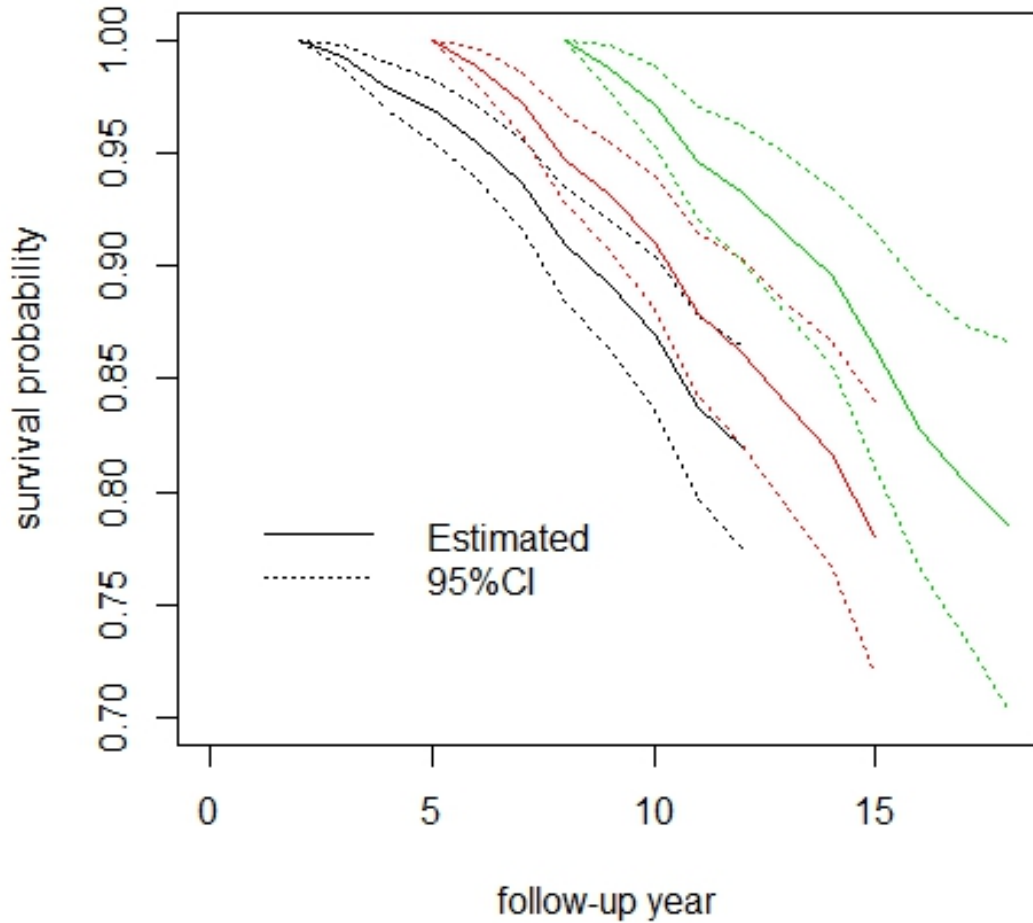
5(紅線)、8(綠線)年之 $S(t|T_M > m)$ 估計值及 95% 信賴區間。

Estimated: 估計值; 95% CI: 估計值的 95% 信賴區間。



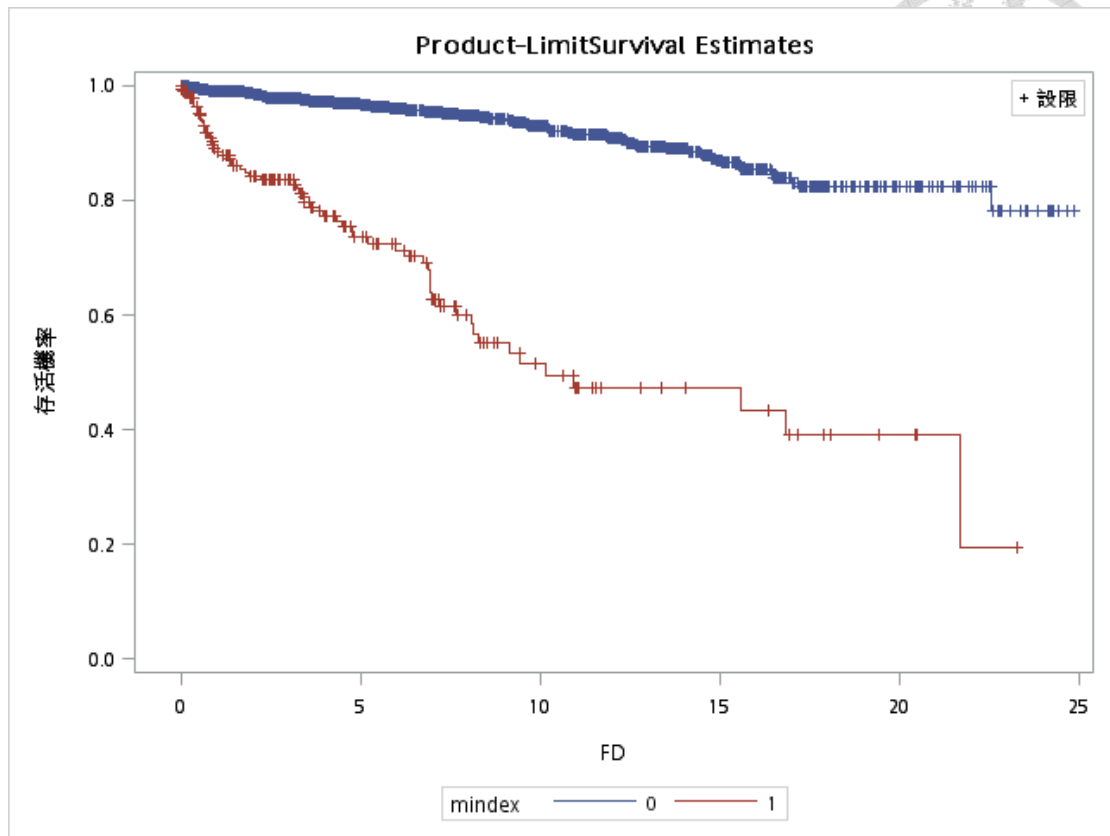
圖十四 40 歲以上之甲狀腺乳突癌患者的標誌時間為不大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M \leq m, T_D > m)$ 估計值及 95% 信賴區間。

Estimated: 估計值; 95% CI: 估計值的 95% 信賴區間。

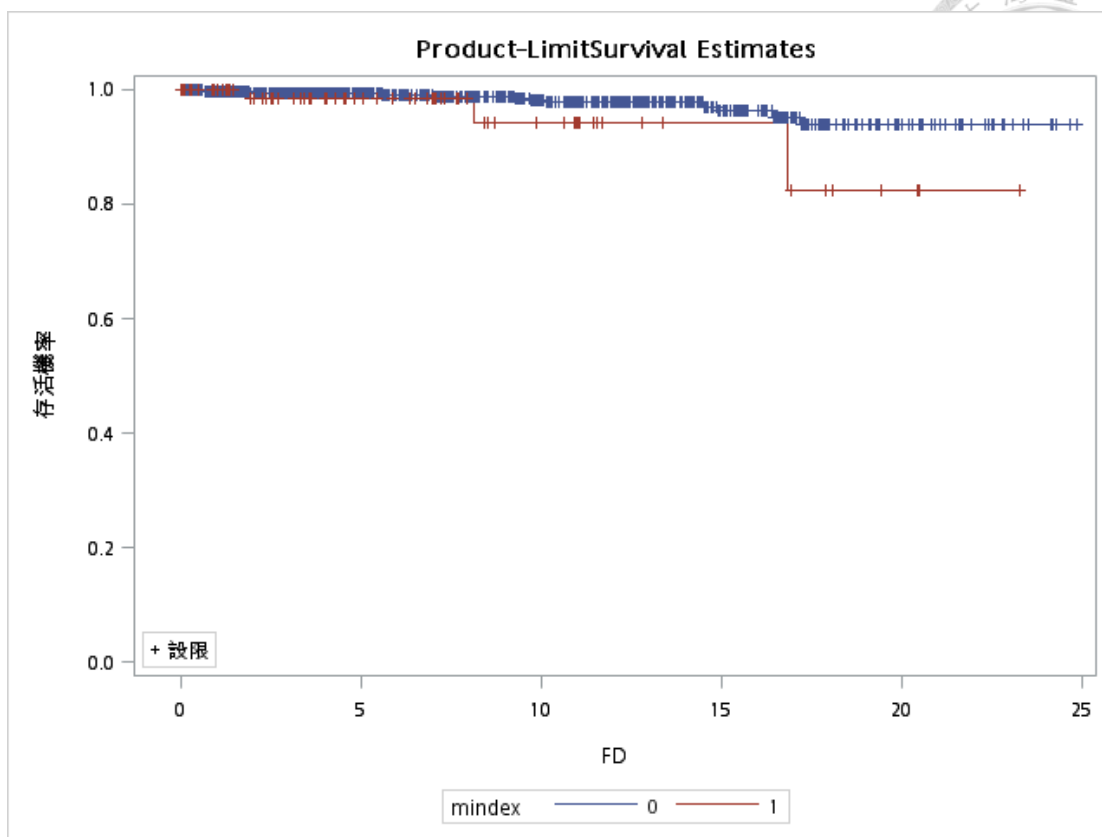


圖十五 40 歲以上之甲狀腺乳突癌患者的標誌時間為大於 2(黑線)、5(紅線)、8(綠線)年之 $S(t|T_M > m, T_D > m)$ 估計值及 95% 信賴區間。

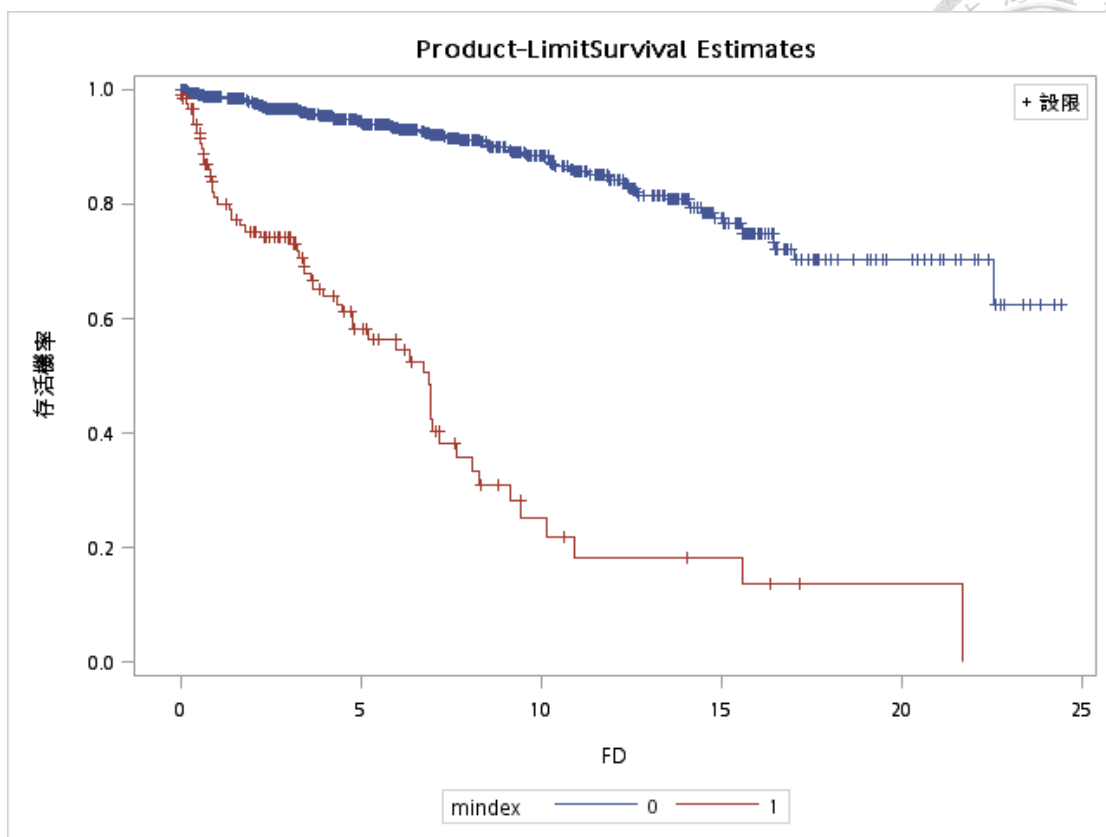
Estimated: 估計值; 95% CI: 估計值的 95% 信賴區間。



圖十六 甲狀腺乳突癌患者的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier 估計量)。



圖十七 甲狀腺乳突癌患者(40歲內)的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier估計量)。



圖十八 甲狀腺乳突癌患者(40歲以上)的未發生標誌事件(藍線)及發生標誌事件(紅線)的存活機率(Kaplan-Meier估計量)。

參考文獻



Al-Khalidi, H. R., Hong, Y., Fleming, T. R., and Therneau, T. M. (2011).

Insights on the robust variance estimator under recurrent-events model.

Biometrics **67**, 1564-1572.

Anderson, J. R., Cain, K. C., and Gelber, R. D. (1983). Analysis of survival by tumor response. *Journal of Clinical Oncology* **1**, 710-719.

Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine* **31**, 3946-3958.

Aydin, Z., Altunbasak, Y., and Borodovsky, M. (2006). Protein secondary structure prediction for a single-sequence using hidden semi-Markov models. *BMC Bioinformatics* **7**, 178.

Bartolomeo, N., Trerotoli, P., and Serio, G. (2011). Progression of liver cirrhosis to HCC: an application of hidden Markov model. *BMC Medical Research Methodology* **11**, 38.

Boberg, K. M., Rocca, G., Egeland, T., *et al.* (2002). Time-dependent Cox



regression model is superior in prediction of prognosis in primary sclerosing cholangitis. *Hepatology* **35**, 652-657.

Bodger, K., Bowering, K., Sarkar, S., Thompson, E., and Pearson, M. G.

(2011). All-cause mortality after first ERCP in England: clinically guided analysis of hospital episode statistics with linkage to registry of death.

Gastrointestinal Endoscopy **74**, 825-833.

Chang, S. H. (2000). A two-sample comparison for multiple ordered event data. *Biometrics* **56**, 183-189.

Chang, S. H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis* **10**, 175-190.

Chen, Y. H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis* **18**, 36-57.

Choi, S., Lagakos, S. W., Schooley, R. T., and Volberding, P. A. (1993).

CD4+ lymphocytes are an incomplete surrogate marker for clinical progression in persons with asymptomatic HIV infection taking zidovudine.



Annals of Internal Medicine **118**, 674-680.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease epidemiology. *Biometrika* **65**, 141-151.

Cook, R. J., and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16**, 911-924.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B* **34**, 187-220.

Day, R., Bryant, J., and Lefkopoulou, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika* **84**, 45-56.

Ding, A. A., Shi, G., Wang, W., and Hsieh, J.-J. (2009). Marginal regression analysis for semi-competing risks data under dependent censoring. *Scandinavian Journal of Statistics* **36**, 481-500.

Fan, J., Hsu, L., and Prentice, R. L. (2000). Dependence estimation over a finite bivariate failure time region. *Lifetime Data Analysis* **6**, 343-355.



Fine, J. P., and Gray. R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* **94**, 496-509.

Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907-919.

Fischer, M., Harvald, B., and Hauge, M. (1969). A Danish twin study of schizophrenia. *The British Journal of Psychiatry* **115**, 981-990.

Flandre, P., and O'Quigley, J. (1995). A two-stage procedure for survival studies with surrogate endpoints. *Biometrics* **51**, 969-976.

Fu, T.-C., Su, D.-H., and Chang, S.-H. (2015). Serial association analyses of recurrent gap time data via Kendall's tau. *Biostatistics*. **17**, 188-202.

Gail, M. H. (1972). Does Cardiac Transplantation Prolong Life? A reassessment. *Annals of Internal Medicine* **76**, 815– 871.

Ghosh, D. (2009). On assessing surrogacy in a single trial setting using a semicompeting risks paradigm. *Biometrics* **65**, 521-529.

Ghosh, D., and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554-562.



Ghosh, D., and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663-688.

Ghosh, D., Taylor, J. M., and Sargent, D. J. (2012). Meta-analysis for surrogacy: accelerated failure time models and semicompeting risks modeling. *Biometrics* **68**, 226-232.

Hartmann, O., Schuetz, P., Albrich, W. C., Anker, S. D., Mueller, B., and Schmidt, T. (2012). Time-dependent Cox regression: serial measurement of the cardiovascular biomarker proadrenomedullin improves survival prediction in patients with lower respiratory tract infection. *International Journal of Cardiology* **161**, 166-173.

Hauge, M., Harvald, B., Fischer, M., *et al.* (1968). The Danish twin register. *Acta Geneticae Medicae et Gemellologiae*. **17**, 315-331.

Hougaard, P. (2000). Analysis of Multivariate Survival Data. *Springer-Verlag New York, Inc.*

Hsieh, J.-J., Wang, W., and Ding, A. A. (2008). Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B* **70**, 3-20.



Huang, C. Y., and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99**, 1153-1165.

Huang, X., and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics* **63**, 389-397.

Kalbfleisch, J. D., and Prentice, R. L. (eds) (2002). *The statistical analysis of failure time data*, 2nd edition. Hoboken, New Jersey: Wiley.

Kang, M., and Lagakos, S. W. (2007). Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics* **8**, 252-264.

Kucukalic-Selimovic, E., Alagic, J., Valjevac, A., Hadzovic-Dzuvo, A., Begic, A., and Beslic, N. (2012). The value of serum thyroglobulin levels and whole body (I-131) scintigraphy in the follow-up of the thyroid cancer patients after thyroidectomy. *Collegium Antropologicum* **36 Suppl 2**, 67-71.

Lagakos, S. W. (1976). A stochastic model for censored-survival data in the presence of an auxiliary variable. *Biometrics* **32**, 551-559.



Lagakos, S. W. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* **33**, 399-404.

Lakhal, L., Rivest, L. P., and Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics* **64**, 180-188.

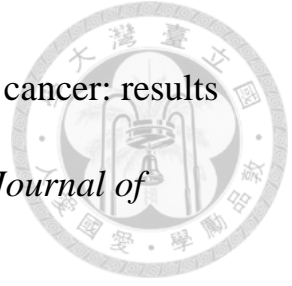
Lancaster, A., and Intrator, O. (1998). Panel data with survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association* **93**, 46-53.

Li, Q. H., and Lagakos, S. W. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16**, 925-940.

Liao, C. S., Yang, K. C., Yen, M. F., and Chen, T. H. (2005). Time-varying predictors for clinical surveillance of small hepatocellular carcinoma. *Cancer Journal* **11**, 226-233.

Lim, H. J., and Zhang, X. (2011). Additive and multiplicative hazards modeling for recurrent event data analysis. *BMC Medical Research Methodology* **11**, 101.

Liu, C. Y., Wu, C. Y., Lin, J. T., Lee, Y. C., Yen, A. M., and Chen, T. H.



(2006). Multistate and multifactorial progression of gastric cancer: results from community-based mass screening for gastric cancer. *Journal of Medical Screening* **13 Suppl 1**, S2-5.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747-756.

Mauguen, A., Rachet, B., Mathoulin-Pelissier, S., Macgrogan, G., Laurent, A., and Rondeau, V. (2013). Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine* **32**, 5366-5380. .

Meira-Machado, L., de Una-Alvarez, J., Cadarso-Suarez, C., and Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* **18**, 195-222.

Nan, B., Lin, X., Lisabeth, L. D., and Harlow, S. D. (2006). Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *Journal of the American Statistical Association* **101**, 65-77.

Neri, A., Marrelli, D., Rossi, S., *et al.* (2007). Breast cancer local



recurrence: risk factors and prognostic relevance of early time to recurrence.

World Journal of Surgery **31**, 36-45.

Oakes, D. (1982a). A concordance test for independence in the presence of censoring. *Biometrics* **38**, 451-455.

Oakes, D. (1982b). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* **73**, 414-422.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487.

Othus, M., and Li, Y. (2010). A Gaussian Copula Model for Multivariate Survival Data. *Statistics in Biosciences* **2**, 154-179.

Peng, L., and Fine, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics* **63**, 96-108.

Pintilie, M. (2007). Analysing and interpreting competing risk data. *Statistics in Medicine* **26**, 1360-1367.

Pollard, T. C., Villar, R. N., Norton, M. R., *et al.* (2010). Genetic influences in the aetiology of femoroacetabular impingement: a sibling



study. *The Journal of Bone and Joint Surgery. British volume* **92**, 209-216.

Porta, N., Calle, M. L., Malats, N., and Gomez, G. (2012). A dynamic model for the risk of bladder cancer progression. *Statistics in Medicine* **31**, 287-300. .

Sargent, D. J., Wieand, H. S., Haller, D. G., *et al.* (2005). Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* **23**, 8664-8670.

Schneider, L. S. (2013). Alzheimer disease pharmacologic treatment and treatment research. *Continuum (Minneap Minn)* **19**, 339-357.

Seifert, M., Gohr, A., Strickert, M., and Grosse, I. (2012). Parsimonious higher-order hidden Markov models for improved array-CGH analysis with applications to *Arabidopsis thaliana*. *PLoS Computational Biology* **8**, e1002286.

Shih, J. H., and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384-1399.



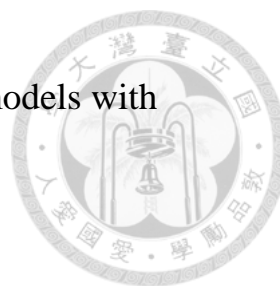
Slutske, W. S., Cho, S. B., Piasecki, T. M., and Martin, N. G. (2013). Genetic overlap between personality and risk for disordered gambling: evidence from a national community-based Australian twin study. *Journal of Abnormal Psychology* **122**, 250-255.

Spiekerman, C. F., and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *The Journal of the American Statistical Association* **93**, 1164-1175.

Su, D. H., Chang, S. H., and Chang, T. C. (2015). The impact of locoregional recurrences and distant metastases on the survival of patients with papillary thyroid carcinoma. *Clinical Endocrinology* **82**, 286-294.

Sweeting, M. J., Farewell, V. T., and De Angelis, D. (2010). Multi-state Markov models for disease progression in the presence of informative examination times: an application to hepatitis C. *Statistics in Medicine* **29**, 1161-1174.

Tan, Y. D., Fornage, M., George, V., and Xu, H. (2007). Parent-child pair design for detecting gene-environment interactions in complex diseases. *Human Genetics* **121**, 745-757.



Titman, A. C., and Sharples, L. D. (2010). Semi-Markov models with phase-type sojourn distributions. *Biometrics* **66**, 742-752.

Vefring, H. K., Wee, L., Jugessur, A., Gjessing, H. K., Nilsen, S. T., and Lie, R. T. (2010). Maternal angiotensinogen (AGT) haplotypes, fetal renin (REN) haplotypes and risk of preeclampsia; estimation of gene-gene interaction from family-triad data. *BMC Medical Genetics* **11**, 90.

Verburg, F. A., Mader, U., Tanase, K., *et al.* (2013). Life expectancy is reduced in differentiated thyroid cancer patients ≥ 45 years old with extensive local tumor invasion, lateral lymph node, or distant metastases at diagnosis and normal in all other DTC patients. *The Journal of Clinical Endocrinology and Metabolism* **98**, 172-180.

Wang, M. C., and Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association* **94**, 146-153.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B* **65**, 257-273.



Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.

Whitmore, G. A., Crowder, M. J., and Lawless, J. F. (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Analysis* **4**, 229-251.

Williams, M. E., Lacson, E., Jr., Wang, W., Lazarus, J. M., and Hakim, R. (2010). Glycemic control and extended hemodialysis survival in patients with diabetes mellitus: comparative results of traditional and time-dependent Cox model analyses. *Clinical Journal of the American Society of Nephrology* **5**, 1595-1601.

Wu, H. M., Yen, M. F., and Chen, T. H. (2004). SAS macro program for non-homogeneous Markov process in modeling multi-state disease progression. *Computer Methods and Programs in Biomedicine* **75**, 95-105.

Xu, R., and O'Quigley, J. (2000). Proportional hazards estimate of the conditional survival function. *Journal of the Royal Statistical Society*,



Series B **62**, 667-680.

Ye, Y., Kalbfleisch, J. D., and Schaubel, D. E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63**, 78-87.

Zhang, W., Fan, J., and Sun, Y. (2009). A semiparametric model for cluster data. *Annals of Statistics* **37**, 2377-2408.

Zhang, X., Zhang, M. J., and Fine, J. (2011). A proportional hazards regression model for the subdistribution with right-censored and left-truncated competing risks data. *Statistics in Medicine* **30**, 1933-1951.

Zhao, X., Liu, L., Liu, Y., and Xu, W. (2012). Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal* **54**, 585-599.

Zoppini, G., Targher, G., Chonchol, M., *et al.* (2012). Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. *Clinical Journal of the American Society of Nephrology* **7**, 401-408.



附錄



附錄 A 甲狀腺癌預後因子研究之台大醫院研究倫理委員會通過函

發文方式：紙本遞送

檔 號：

保存年限：

國立臺灣大學醫學院附設醫院 函

地址：10002台北市中山南路7號

承辦人：顏鈺芳

電話：02-23123456分機63596

傳真：02-2395-1950

電子信箱：ani ta0314@ntuh.gov.tw

受文者：內科部張天鈞醫師

發文日期：中華民國98年12月17日

發文字號：校附醫倫字第0983704047號

速別：普通件

密等及解密條件或保密期限：普通

附件：如文

主旨：有關 台端所主持之「甲狀腺癌患者之預後因子/The prognostic factors of thyroid cancer」（本院案號：200911068R）純學術臨床試驗案，符合快速審查條件及研究倫理規範，通過本院研究倫理委員會審查，同意核備，並提第159研究倫理委員會報備追認，請 查照。

說明：

- 一、本臨床試驗核准之有效期限為1年，計畫主持人應於到期前的1個月提出持續審查申請表，本案需經持續審查，方可繼續執行。
- 二、本會同意之臨床試驗計畫書之版本日期為：20091202。
- 三、臨床試驗執行期間，請確實依據「藥品優良臨床試驗準則」之相關規定辦理；為符合「藥品優良臨床試驗準則」結案查核作業，請計畫主持醫師保存所有文件備查。
- 四、依據國際醫學雜誌編輯委員會(The International Committee of Medical Journal Editors, ICMJE)之投稿規定(http://www.icmje.org/clin_trialup.htm)，臨床試驗研究計畫投稿者，需於招募第一位受試者參與試驗前，將通過研究倫理委員會審核之臨床試驗計畫資料登錄於臨床試驗公開網站，完成登錄作業後，國際醫學雜誌編輯委員會(



裝

訂

線

ICMJE)才會接受研究結果之發表。WHO對臨床試驗研究計畫之定義為任何對受試者或特定族群進行一個或多個與健康有關的介入措施(如藥物、外科處置、器材、行為治療、飲食介入及照護過程改變)以評估對健康的效益之計畫，非屬上述臨床試驗計畫，請計畫主持人自行決定是否登錄。

五、本院已向美國國家衛生研究院(National Institutes of Health, NIH) ClinicalTrials.gov 網站- Protocol Registration System (PRS [https:// register.clinicaltrials.gov/](https://register.clinicaltrials.gov/))申請本院專用帳號，供本院計畫主持人(PI)登錄所主持之臨床試驗研究計畫，登入網頁之帳號及密碼如下列：

(一)Organization : NTaiwanUH

(二)User Name : NTUH

(三>Password : NTUH99

六、隨函檢附臨床研究重要訊息通知單，請依計畫需要辦理相關事宜。

正本：內科部張天鈞醫師
副本：研究倫理委員會

院長 陳明豐

發文方式：紙本遞送

檔 號：

保存年限：

國立臺灣大學醫學院附設醫院 函

地址：100臺北市中山南路7號

承辦人：李桃森

電話：02-2312-3456轉66590

傳真：02-2395-1950

電子信箱：ntuhrec@ntuh.gov.tw

受文者：內科部張天鈞醫師

發文日期：中華民國99年11月11日

發文字號：校附醫倫字第0993704533號

速別：普通件

密等及解密條件或保密期限：普通

附件：

主旨：有關 台端所送之「甲狀腺癌患者之預後因子」（本院案號：200911068R）純學術臨床試驗計畫案持續審查報告及申請展延試驗期限，業經本院C研究倫理委員會審查，同意核備，並提第11次會議報備追認，請 查照。

說明：

- 一、同意展延試驗期限至2012年6月30日。
- 二、本臨床試驗核准之有效期限至2011年12月16日，計畫主持人應於到期前的1個月提出持續審查申請表，本案需經持續審查通過後，方可繼續執行。

正本：內科部張天鈞醫師

副本：研究倫理委員會

院長 陳明豐



附錄 B 無共變數下，條件存活機率 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的推導

(B · 1) 推導(3-7)式

$$\begin{aligned}
 & P(T_M \leq m) \\
 &= P(T_M \leq m|T_M \leq T_D)P(T_M \leq T_D) + P(T_M \leq m|T_M > T_D)P(T_M > T_D) \\
 &= P(T_M \leq m, T_D > m|T_M \leq T_D)P(T_M \leq T_D) + P(T_M \leq m, T_D \leq m|T_M \leq T_D)P(T_M \leq T_D) \\
 &+ P(T_M \leq m|T_M > T_D)P(T_M > T_D);
 \end{aligned}$$

或

$$\begin{aligned}
 & P(T_M \leq m) \\
 &= P(T_M \leq m, T_M \leq T_D) + P(T_M \leq m, T_M > T_D) \\
 &= P(T_M \leq m, T_D > m) + P(T_M \leq m, T_D \leq m) \\
 &= P(T_M \leq m, T_D > m, T_M \leq T_D) + P(T_M \leq m, T_D \leq m, T_M \leq T_D) + P(T_M \leq m, T_M > T_D);
 \end{aligned}$$

從(3-7)式來看， $P(T_M \leq m)$ 可依路徑分成兩部分，或依 T_D 的值分成兩部分，而 $P(T_M \leq m, T_D \leq m)$ 再依路徑又可分成兩部分，如此 $P(T_M \leq m)$ 分解成三部分，這是為了估計 $P(T_M \leq m)$ 時，這三部分是不同的。

(B · 2) 推導(3-10)式

$$\begin{aligned}
 & S(t|T_M \leq m) \\
 &= \frac{p(T_D > t, T_M \leq m)}{P(T_M \leq m)} \\
 &= \frac{\int_t^\infty P(T_M \leq m, T_D = u)du}{\int_0^m P(T_M \leq m, T_D = u)du + \int_m^\infty P(T_M \leq m, T_D = u)du} \\
 &= \frac{\int_t^\infty P(T_M \leq m|T_D = u)f(u)du}{\int_0^m P(T_M \leq m|T_D = u)f(u)du + \int_m^\infty P(T_M \leq m|T_D = u)f(u)du}
 \end{aligned}$$



$$\begin{aligned}
&= \frac{\int_t^\infty P(T_M \leq m | T_D = u) dF(u)}{\int_0^m P(T_M \leq m | T_D = u) dF(u) + \int_m^\infty P(T_M \leq m | T_D = u) dF(u)} \\
&= \frac{\int_t^\infty P(T_M \leq m | T_D = u) dF(u)}{\int_0^m P(T_M \leq u | T_D = u) dF(u) + \int_0^m P(u < T_M \leq m | T_D = u) dF(u) + \int_m^\infty P(T_M \leq m | T_D = u) dF(u)}
\end{aligned}$$

(B · 3) 推導 $g(r|T_D = u)$

$$\begin{aligned}
&g(r|T_D = u) \\
&= \frac{h_0(u) \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] S(u|T_M = r) g(r)}{\int_0^u h_0(u) \exp[\beta_1 + \beta_2 k(v)] S(u|T_M = v) g(v) dv + \int_u^\infty h_0(u) S(u|T_M = v) g(v) dv} \\
&= \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r|T_D \geq u) S(u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) S(u) dv + \int_u^\infty g(v|T_D \geq u) S(u) dv} \\
&= \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + \int_u^\infty g(v|T_D \geq u) dv} \\
&= \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)}
\end{aligned}$$

透過 Cox 模式的導入到分子、分母中而消除了基礎風險函數 $h_0(\cdot)$

而免除了其計算，而透過貝氏定理的機率反轉而使得原來無法估計的邊際機率函數 $g(r)$ 轉成可估計的條件式機率函數 $g(r|T_D \geq u)$ ，使得條件機率 $g(r|T_D = u)$ 轉為 β 和 $g(r|T_D \geq u)$ 的函數。根據不同的 T_M 的範圍將 $g(r|T_D = u)$ 積分，便可得到以下四式 $P(T_M \leq m|T_D = u)$ 、

$P(T_M \leq u|T_D = u)$ 、 $P(u < T_M \leq m|T_D = u)$ 及 $P(T_M > m|T_D = u)$ 以做為推導條件機率之 $S(t|T_M \leq m)$ 和 $S(t|T_M > m)$ 之準備。

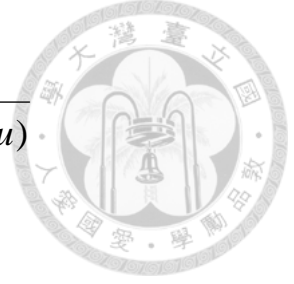


$$\begin{aligned}
P(T_M \leq m | T_D = u) &= \int_0^m P(T_M = r | T_D = u) dr \\
&= \frac{\int_0^m \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r | T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} \\
&= \frac{\int_0^m \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)}
\end{aligned}$$

$$\begin{aligned}
P(T_M \leq u | T_D = u) &= \int_0^u P(T_M = r | T_D = u) dr \\
&= \frac{\int_0^u \exp[\beta_1 + \beta_2 k(r)] g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)}
\end{aligned}$$

$$\begin{aligned}
P(u < T_M \leq m | T_D = u) &= \int_u^m P(T_M = r | T_D = u) dr \\
&= \frac{\int_u^m g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)}
\end{aligned}$$

$$\begin{aligned}
P(T_M > m | T_D = u) &= \int_m^\infty P(T_M = r | T_D = u) dr \\
&= \int_m^u P(T_M = r | T_D = u) dr + \int_u^\infty P(T_M = r | T_D = u) dr \\
&= \int_m^u \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r | T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dr \\
&\quad + \int_u^\infty \frac{\exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] g(r | T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dr \\
&= \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)}
\end{aligned}$$



$$\begin{aligned}
& + \frac{\int_u^\infty g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} \\
& = \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr + 1 - G(u|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)}
\end{aligned}$$

(B · 4) 推導 $S(t|T_M \leq m)$

根據上述(B · 3)， $S(t|T_M \leq m)$ 可以表示如下：

$$\begin{aligned}
S(t|T_M \leq m) &= \frac{P(T_M \leq m, T_D \geq t_D)}{P(T_M \leq m)} \\
&= \frac{\int_t^\infty P(T_M \leq m|T_D = u) dF(u)}{\int_0^m P(T_M \leq m|T_D = u) dF(u) + \int_m^\infty P(T_M \leq m|T_D = u) dF(u)} \\
&= \frac{\int_t^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)}{\int_0^m \left(\frac{\int_0^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) + \int_u^m g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} \right) dF(u) + \int_m^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)} \\
&= \frac{\int_t^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} dF(u)}{\int_0^m \left(\frac{\int_0^u \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, T_D \geq u) + \int_u^m P(T_M = r, T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} \right) dF(u) + \int_m^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} dF(u)}
\end{aligned}$$

如為完整資料，我們可利用經驗估計式來估計下列式子及

$\hat{S}(u)$ ，表示如下：

$$\hat{G}(u|T_D \geq u) = \frac{\sum_{i=1}^n I(T_{D_i} \geq u, T_{M_i} \leq u)}{\sum_{i=1}^n I(T_{D_i} \geq u)},$$

$$\hat{S}(u) = \frac{1}{n} \sum_{i=1}^n I(T_{D_i} > u),$$

$$\text{及 } d\hat{F}(u) = \hat{S}(u^-) - \hat{S}(u).$$



由上可知， $\hat{S}(t|T_M \leq m)$ 為 $\frac{\hat{A}_1^*}{\hat{A}_2^* + \hat{B}_1^*}$ 的型式， \hat{A}_1^* 為估計

$\hat{P}(T_D \geq t, T_M \leq m)$ 的式子，為圖二中的區域 1 之面積機率， \hat{A}_2^* 為估計

$\hat{P}(T_D \leq m, T_M \leq m)$ 的式子，為圖二中的區域 3+4 面積機率， \hat{B}_1^* 為估計

$\hat{P}(T_D \geq m, T_M \leq m)$ 的式子，為圖二中的區域 1+2 面積機率。 \hat{A}_1^* ， \hat{A}_2^* 和

\hat{B}_1^* 可表示為以下的式子：

$$\begin{aligned} \hat{A}_1^* &= \hat{P}(T_D \geq t, T_M \leq m) \\ &= \sum_{i=1}^n I(T_{D_i} \geq t) \frac{\sum_{j=1}^n \hat{A}_{1ij}^*}{\sum_{j=1}^n [\hat{A}_{2ij}^* + \hat{B}_{1ij}^*]} d\hat{F}(T_{D_i}), \end{aligned}$$

其中

$$\hat{A}_{1ij}^* = \left\{ \exp(\hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq m, T_{D_j} \geq T_{D_i}) \right\},$$

$$\hat{A}_{2ij}^* = \left\{ \exp(\hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) \right\} I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}),$$

$$\hat{B}_{1ij}^* = I(T_{D_j} \geq T_{D_i}, T_{M_j} > T_{D_i}).$$

另外，

$$\begin{aligned} \hat{A}_2^* &= \hat{P}(T_D \leq m, T_M \leq m) \\ &= \sum_{i=1}^n I(T_{D_i} \leq m) \frac{\sum_{j=1}^n \hat{A}_{3ij}^*}{\sum_{j=1}^n \{\hat{A}_{2ij}^* + \hat{B}_{1ij}^*\}} d\hat{F}(T_{D_i}), \end{aligned}$$

其中

$$\hat{A}_{3ij}^* = \left\{ \exp(\hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) \right\} I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}) + I(T_{D_i} < T_{M_j} \leq m, T_{D_j} \geq T_{D_i}),$$

和

$$\hat{B}_1^* = \hat{P}(T_D \geq m, T_M \leq m)$$



$$= \sum_{i=1}^n I(T_{D_i} \geq m) \frac{\sum_{j=1}^n \hat{A}_{1ij}^*}{\sum_{j=1}^n \{\hat{A}_{2ij}^* + \hat{B}_{1ij}^*\}} d\hat{F}(T_{D_i}) \circ$$

(B · 5) 推導 $S(t|T_M > m)$

要推導此式前，要先推導此式前，需先推導其分母，即 $P(T_M > m)$ ，

表示如下：

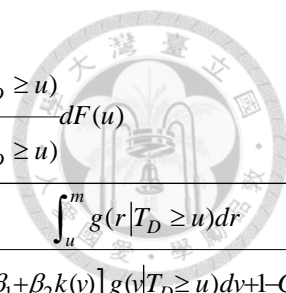
$$\begin{aligned} & P(T_M > m) \\ &= P(T_M > m|T_M \leq T_D)P(T_M \leq T_D) + P(T_M > m|T_M > T_D)P(T_M > T_D) \\ &= P(T_M > m|T_M \leq T_D)P(T_M \leq T_D) + P(T_M > T_D) - P(T_M \leq m|T_M > T_D)P(T_M > T_D) \\ &= P(T_M > m, T_M \leq T_D) + 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M > T_D) \\ &= 1 - P(T_M \leq m) \circ \end{aligned}$$

即 $P(T_M > m)$ 有兩種表示方法，第一種

$P(T_M > m, T_M \leq T_D) + 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M > T_D)$ ，即模擬方法中的

第一種方法，表示如下：

$$\begin{aligned} S(t|T_M > m) &= \frac{P(T_M > m, T_D > t)}{P(T_M > m)} \\ &= \frac{P(T_M > m, T_D > t)}{P(T_M > m, T_M \leq T_D) + 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M > T_D)} \\ &= \frac{\int_t^\infty P(T_M > m|T_D = u)dF(u)}{\int_m^\infty P(u \geq T_M > m|T_D = u)dF(u) + 1 - P(T_M \leq T_D) - \int_0^m P(u < T_M \leq m|T_D = u)dF(u)} \end{aligned}$$



$$\begin{aligned}
 & \frac{\int_t^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr + 1 - G(u|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)}{\int_m^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u) + 1 - P(T_M \leq T_D) - \int_0^m \frac{\int_u^m g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)} \\
 & \frac{\int_t^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, T_D \geq u) dr + P(T_M > u, T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} dF(u)}{\int_m^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} dF(u) + 1 - P(T_M \leq T_D) - \int_0^m \frac{\int_u^m P(T_M = r, T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] P(T_M = v, T_D \geq u) dv + P(T_M > u, T_D \geq u)} dF(u)}
 \end{aligned}$$

而第二種方法是 $1 - P(T_M \leq m)$ ，為模擬方法中的第二種方法。

$$S(t|T_M > m) = \frac{P(T_M > m, T_D > t)}{1 - P(T_M \leq m)}$$

$$\begin{aligned}
 & \frac{\int_t^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr + 1 - G(u|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)}{1 - \left(\int_0^m \frac{\left(\int_0^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr + \int_u^m g(r|T_D \geq u) dr \right)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u) + \int_m^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u) \right)}
 \end{aligned}$$

計算 $S(t|T_M > m)$ 和 $S(t|T_M \leq m)$ 的方法很類似， $T_M > m$ 的訊息必須運用 (3-8) 式，並運用 (B·3) 中的 $g(r|T_D = u)$ 在適當的 T_M 範圍對 $g(r|T_D = u)$ 積分便可得到計算條件存活機率 $S(t|T_M > m)$ 中分子和分母所需要的 $P(T_M > t|T_D = u)$ 、 $P(u < T_M \leq m|T_D = u)$ 及 $P(T_M > m|T_D = u)$ 等三式， $S(t|T_M > m)$ 的式子如下：

$$\begin{aligned}
 & S(t|T_M > m) \\
 & = \frac{P(T_M > m, T_D > t)}{P(T_M > m|T_M \leq T_D)P(T_M \leq T_D) + 1 - P(T_M \leq T_D) - P(T_M \leq m, T_M > T_D)}
 \end{aligned}$$

$$= \frac{\int_t^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r | T_D \geq u) dr + 1 - G(u | T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dF(u)}{\int_m^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dF(u) + 1 - P(T_M \leq T_D) - \int_0^m \frac{\int_u^m g(r | T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v | T_D \geq u) dv + 1 - G(u | T_D \geq u)} dF(u)}$$

在完整資料中， $P(T_M \leq T_D)$ 可用下式估計

$$\hat{P}(T_M \leq T_D) = \frac{1}{n} \sum_{i=1}^n I(T_{M_i} \leq T_{D_i}) .$$

利用上述式子，我們可得到 $\hat{S}(t | T_M > m)$ 。其估計式為

$$\hat{S}(t | T_M > m) = \frac{\hat{A}_1^{**}}{\hat{A}_2^{**} + \hat{B}_1^{**}} , \text{ 在此 } \hat{A}_1^{**} \text{ 為計算 } \hat{P}(T_M > m, T_D \geq t) , \text{ 為圖二的區域}$$

5+6 面積機率， \hat{A}_2^{**} 為計算 $\hat{P}(T_M > m, T_M \leq T_D)$ ，為圖二的區域 5+7 面積機

率， \hat{B}_1^{**} 為計算 $\hat{P}(T_M > m, T_M > T_D)$ ，為圖二的區域 6+8 面積機率， \hat{A}_1^{**} 、 \hat{A}_2^{**}

和 \hat{B}_1^{**} 可表示為

$$\begin{aligned} \hat{A}_1^{**} &= \hat{P}(T_M > m, T_D \geq t) \\ &= \sum_{i=1}^n I(T_{D_i} \geq t) \frac{\sum_{j=1}^n [\hat{A}_{1ij}^{**} + \hat{B}_{1ij}^{**}]}{\sum_{j=1}^n \{\hat{A}_{2ij}^{**} + \hat{B}_{1ij}^{**}\}} d\hat{F}(T_{D_i}) , \end{aligned}$$

其中

$$\hat{A}_{1ij}^{**} = \exp(\hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(m \leq T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}) ,$$

$$\hat{A}_{2ij}^{**} = \exp(\hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}) ,$$

$$\hat{B}_{1ij}^{**} = I(T_{D_j} \geq T_{D_i}, T_{M_j} > T_{D_i}) ,$$

另外，

$$\hat{A}_2^{**} = \hat{P}(T_M > m, T_M \leq T_D)$$



$$= \sum_{i=1}^n I(T_{D_i} > m) \frac{\sum_{j=1}^n [\hat{A}_{1ij}^{**}]}{\sum_{j=1}^n \{\hat{A}_{2ij}^{**} + \hat{B}_{1ij}^{**}\}} d\hat{F}(T_{D_i}),$$

和

$$\begin{aligned} \hat{B}_1^{**} &= \hat{P}(T_M > m, T_M > T_D) \\ &= 1 - \frac{1}{n} \sum_{i=1}^n I(T_{M_i} \leq T_{D_i}) - \sum_{i=1}^n I(T_{D_i} \leq m) \frac{\sum_{j=1}^n \hat{A}_{4ij}^*}{\sum_{j=1}^n \{\hat{A}_{2ij}^* + \hat{B}_{1ij}^*\}} d\hat{F}(T_{D_i}), \end{aligned}$$

其中

$$\hat{A}_{4ij}^* = I(T_{D_i} < T_{M_j} \leq m, T_{D_j} \geq T_{D_i}).$$

另因 $P(T_M > m) = 1 - P(T_M \leq m)$ ，故可表示如下：

$$\begin{aligned} S(t|T_M > m) &= \frac{P(T_M > m, T_D > t)}{1 - P(T_M \leq m)} \\ &= \frac{\int_t^\infty \frac{\int_m^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr + 1 - G(u|T_D \geq u)}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u)}{1 - \left(\int_0^m \frac{\int_0^u \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) + \int_u^m g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u) + \int_m^\infty \frac{\int_0^m \exp[\beta_1 + \beta_2 k(r)] g(r|T_D \geq u) dr}{\int_0^u \exp[\beta_1 + \beta_2 k(v)] g(v|T_D \geq u) dv + 1 - G(u|T_D \geq u)} dF(u) \right)}. \end{aligned}$$

故分母可運用 $1 - \hat{P}(T_M \leq m)$ 計算，以 $1 - (\hat{A}_2^* + \hat{B}_1^*)$ 來估計



附錄 C 條件存活機率 $S(t|T_M \leq m)$ 及 $S(t|T_M > m)$ 的推導

(C · 1) 推導 $S(t|T_M \leq m)$

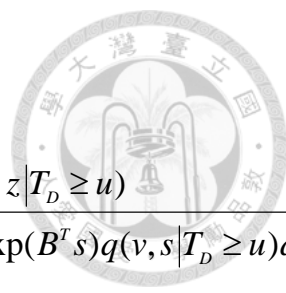
$$\begin{aligned}
 S(t|T_M \leq m) &= \frac{P(T_D > t, T_M \leq m)}{P(T_M \leq m)} \\
 &= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, T_D = u, Z = z) dz du}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, T_D = u, Z = z) dz du + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, T_D = u, Z = z) dz du} \\
 &= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) f(u) dz du}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) f(u) dz du + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) f(u) dz du} \\
 &= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) f(u) dz du}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) f(u) dz du + \int_m^\infty \int_{Z \in H_T} (P(T_M \leq m, Z = z | T_D = u) f(u) dz du} \\
 &= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) dz dF(u)}{\int_0^m \int_{Z \in H_T} (P(T_M \leq m, Z = z | T_D = u) + P(u < T_M \leq m, Z = z | T_D = u)) dz dF(u) + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) dz dF(u)}
 \end{aligned}$$

(C · 2) 推導 $q(r, z|T_D = u)$ 及後續 $S(t|T_M \leq m)$

推導 $q(r, z|T_D = u)$ ，為推導之關鍵式，表示如下：

$$\begin{aligned}
 q(r, z|T_D = u) &= \frac{h_0(u) \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] S(u|T_M = r, Z = z) q(r, z)}{\int_0^u \int_{Z \in H_T} h_0(v) \exp[B^T s + \beta_1 + \beta_2 k(v)] S(u|T_M = v, Z = s) q(v, s) ds dv + \int_u^\infty \int_{Z \in H_T} h_0(v) \exp(B^T s) S(u|T_M = v, Z = s) q(v, s) ds dv} \\
 &= \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z|T_D \geq u) S(u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s|T_D \geq u) S(u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s|T_D \geq u) S(u) ds dv} \\
 &= \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z|T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s|T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s|T_D \geq u) ds dv}
 \end{aligned}$$

根據 $q(r, z|T_D = u)$ ，可推導下列式子：



$$\begin{aligned}
P(T_M \leq m, Z = z | T_D = u) &= \int_0^m P(T_M = r, Z = z | T_D = u) dr \\
&= \int_0^m \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dr \\
&= \frac{\int_0^m \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv}
\end{aligned}$$

$$\begin{aligned}
P(T_M \leq u, Z = z | T_D = u) &= \int_0^u P(T_M = r, Z = z | T_D = u) dr \\
&= \int_0^u \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dr \\
&= \frac{\int_0^u \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv}
\end{aligned}$$

$$\begin{aligned}
P(T_M > m, Z = z | T_D = u) &= \int_m^\infty P(T_M = r, Z = z | T_D = u) dr \\
&= \int_m^u P(T_M = r, Z = z | T_D = u) dr + \int_u^\infty P(T_M = r, Z = z | T_D = u) dr \\
&= \int_m^u \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dr \\
&\quad + \int_u^\infty \frac{\exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u)}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dr \\
&= \frac{\int_m^u \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} \\
&\quad + \frac{\int_u^\infty \exp(B^T z) q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv}
\end{aligned}$$

$$= \frac{\int_m^u \exp[\exp(B^T z) + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr + \int_u^\infty \exp(B^T z) q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv}$$

根據上述式子，可推導 $S(t | T_M \leq m)$ 如下：

$$S(t | T_M \leq m)$$

$$= \frac{\int_t^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) dz dF(u)}{\int_0^m \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) dz dF(u) + \int_m^\infty \int_{Z \in H_T} P(T_M \leq m, Z = z | T_D = u) dz dF(u)}$$

其中上式分子為

$$P(T_M \leq m, T_D \geq t) = \int_t^\infty \int_{Z \in H_T} \frac{\int_0^m \exp[B^T z + \beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u),$$

而其分母分為兩部份，表示如下：

$$P(T_D \leq m, T_M \leq m)$$

$$= \int_0^m \int_{Z \in H_T} \frac{\int_0^u \exp[B^T z + \beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr + \int_u^m \exp(B^T z) q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u),$$

$$P(T_D > m, T_M \leq m)$$

$$= \int_m^\infty \int_{Z \in H_T} \frac{\int_0^m \exp[B^T s + \beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u)$$

故從以上的推導中，可知 $\hat{S}(t | T_M \leq m)$ 為 $\frac{p\hat{C}_1}{p\hat{D}_1 + p\hat{C}_2}$ 的型式，其中

$p\hat{C}_1$ ， $p\hat{D}_1$ 和 $p\hat{C}_2$ 可表示為以下的式子：

$$p\hat{C}_1 = \hat{P}(T_M \leq m, T_D \geq t)$$

$$= \sum_{i=1}^n I(T_{D_i} \geq t) \frac{\sum_{j=1}^n p\hat{C}_{1ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}_e(T_{D_i}),$$

其中 $d\hat{F}_e(\cdot)$ 為經驗估計式及



$$\begin{aligned}
 p\hat{C}_{1ij} &= \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq m, T_{D_j} \geq T_{D_i}) , \\
 p\hat{C}_{2ij} &= \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}) , \\
 p\hat{D}_{1ij} &= \exp(\hat{B}^T Z_j) I(T_{M_j} \geq T_{D_i}, T_{D_j} \geq T_{D_i}) ,
 \end{aligned}$$

另外，

$$\begin{aligned}
 p\hat{D}_1 &= \hat{P}(T_M \leq m, T_D \leq m) \\
 &= \sum_{i=1}^n I(T_{D_i} < m) \frac{\sum_{j=1}^n p\hat{C}_{3ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}_e(T_{D_i}) ,
 \end{aligned}$$

其中

$$\begin{aligned}
 p\hat{C}_{3ij} &= \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}) \\
 &\quad + \exp(\hat{B}^T Z_j) I(T_{D_i} < T_{M_j} \leq m, T_{D_j} \geq T_{D_i}) ,
 \end{aligned}$$

和

$$\begin{aligned}
 p\hat{C}_2 &= \hat{P}(T_M \leq m, T_D \geq m) \\
 &= \sum_{i=1}^n I(T_{D_i} \geq m) \frac{\sum_{j=1}^n p\hat{C}_{1ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}_e(T_{D_i}) .
 \end{aligned}$$

(C · 3) 推導 $S(t|T_M > m, Z \in H)$

H_T 為共變數之全域，若 H 為其子合或部分範圍，則若要得到 $S(t|T_M > m, Z \in H)$ ，則先推導 $S(t|T_M > m, Z = z)$ 如下：

$$\begin{aligned}
 S(t|T_M > m, Z = z) &= \frac{P(T_M > m, T_D > t, Z = z)}{P(T_M > m, Z = z)} \\
 &= \frac{P(T_M > m, T_D > t, Z = z)}{P(T_M > m, Z = z | T_M \leq T_D) P(T_M \leq T_D) + P(T_M > T_D, T_M > m, Z = z)}
 \end{aligned}$$



$$\begin{aligned}
&= \frac{\int_t^\infty P(T_M > m, Z = z | T_D = u) dF(u)}{\int_m^\infty P(u \geq T_M > m, Z = z | T_D = u) dF(u) + P(T_M > T_D, T_M > m, Z = z)} \\
&\quad \int_t^\infty \left\{ \frac{\exp(B^T z) \int_m^u \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr + \exp(B^T z) \int_u^\infty q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) q(v, s | T_D \geq u) ds dv} \right\} dF(u) \\
&= \frac{\int_m^\infty \frac{\exp(B^T z) \int_m^u \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dF(u) + w(z) - P(T_M \leq T_D, Z = z) - P(T_M > T_D, T_M \leq m, Z = z)}{\int_m^\infty \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) q(v, s | T_D \geq u) ds dv}
\end{aligned}$$

則 $S(t | T_M > m, Z \in H)$ 表示如下:

$$S(t | T_M > m, Z \in H) = \frac{P(T_M > m, T_D > t, Z \in H)}{P(T_M > m, Z \in H)}$$

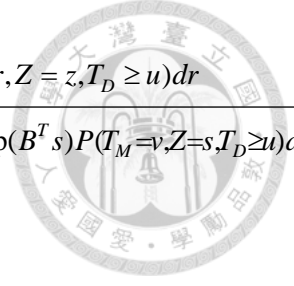
$$= \frac{\int_t^\infty \int_{Z \in H} P(T_M > m, Z = z | T_D = u) dz dF(u)}{\int_m^\infty \int_{Z \in H} P(u \geq T_M > m, Z = z | T_D = u) dz dF(u) + \int_{Z \in H} P(T_M > T_D, T_M > m, Z = z) dz}$$

其中分子表示如下:

$$\begin{aligned}
&\int_t^\infty \int_{Z \in H} P(T_M > m, Z = z | T_D = u) dz dF(u) \\
&= \int_t^\infty \int_{Z \in H} \left\{ \frac{\exp(B^T z) \int_m^u \exp[\beta_1 + \beta_2 k(r)] q(r, z | T_D \geq u) dr + \exp(B^T z) \int_u^\infty q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) q(v, s | T_D \geq u) ds dv} \right\} dz dF(u) \\
&= \int_t^\infty \int_{Z \in H} \left\{ \frac{\exp(B^T z) \int_m^u \exp[\beta_1 + \beta_2 k(r)] P(T_M = r, Z = z, T_D \geq u) dr + \exp(B^T z) \int_u^\infty q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] P(T_M = v, Z = s, T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) P(T_M = v, Z = s, T_D \geq u) ds dv} \right\} dz dF(u),
\end{aligned}$$

分母為兩部分，表示如下:

$$\begin{aligned}
&\int_m^\infty \int_{Z \in H} P(u \geq T_M > m, Z = z | T_D = u) dz dF(u) \\
&= \int_m^\infty \int_{Z \in H} \frac{\exp(B^T z) \int_m^u \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] q(r, z | T_D \geq u) dr}{\int_0^u \int_{Z \in H_t} \exp[B^T s + \beta_1 + \beta_2 k(v)] q(v, s | T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_t} \exp(B^T s) q(v, s | T_D \geq u) ds dv} dz dF(u)
\end{aligned}$$



$$\begin{aligned}
&= \int_m^\infty \int_{Z \in H} \frac{\exp(B^T z) \int_m^u \exp[\beta_1 I(r \leq u) + \beta_2 I(r \leq u)k(r)] P(T_M = r, Z = z, T_D \geq u) dr}{\int_0^u \int_{Z \in H_T} \exp[B^T s + \beta_1 + \beta_2 k(v)] P(T_M = v, Z = s, T_D \geq u) ds dv + \int_u^\infty \int_{Z \in H_T} \exp(B^T s) P(T_M = v, Z = s, T_D \geq u) ds dv} dz dF(u), \\
&\int_{Z \in H} P(T_M > T_D, T_M > m, Z = z) dz \\
&= \int_{Z \in H} (w(z) - P(T_M \leq T_D, Z = z) - P(T_M \leq m, T_M > T_D, Z = z)) dz \\
&= \int_{Z \in H} w(z) dz - \int_{Z \in H} P(T_M \leq T_D, Z = z) dz - \int_{Z \in H} P(T_M \leq m, T_M > T_D, Z = z) dz. \circ
\end{aligned}$$

上式中的 $\int_{Z \in H} P(T_M \leq T_D, Z = z) dz$ 在完整資料中可使用經驗估計

式，而在右受限資料中則可使用 IPCW 法估計，而

$\int_{Z \in H} P(T_M \leq m, T_M > T_D, Z = z) dz$ 則為 $\int_{Z \in H} P(T_M \leq m, Z = z) dz$ 的一部分，已在前述推導。

$S(t|T_M > m, Z \in H_T)$ 可視為 $S(t|T_M > m, Z \in H)$ 的特例，即 Z 的

範圍為全域，如此便可得到。可知 $\hat{S}(t|T_M > m)$ 為 $\frac{p\hat{C}_1^\#}{p\hat{C}_2^\# + p\hat{D}_1^\#}$ 的型式，其

中 $p\hat{C}_1^\#$ ， $p\hat{C}_2^\#$ 和 $p\hat{D}_1^\#$ 可表示為以下的式子：

$$\begin{aligned}
p\hat{C}_1^\# &= \hat{P}(T_M > m, T_D \geq t) \\
&= \sum_{i=1}^n I(T_{D_i} \geq t) \left[\frac{\sum_{j=1}^n p\hat{C}_{1ij}^{**} + \sum_{j=1}^n p\hat{D}_{1ij}^{**}}{\sum_{j=1}^n p\hat{C}_{2ij}^{**} + \sum_{j=1}^n p\hat{D}_{2ij}^{**}} \right] d\hat{F}_e(T_{D_i}),
\end{aligned}$$

其中

$$p\hat{C}_{1ij}^{**} = \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(m \leq T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}),$$

$$p\hat{D}_{1ij}^{**} = \exp(B^T Z_j) I(T_{D_j} \geq T_{D_i}, T_{M_j} \geq T_{D_i}),$$

$$p\hat{C}_{2ij}^{**} = \exp(\hat{B}^T Z_j + \hat{\beta}_1 + \hat{\beta}_2 k(T_{M_j})) I(T_{M_j} \leq T_{D_i}, T_{D_j} \geq T_{D_i}),$$

$$p\hat{D}_{2ij}^{**} = \exp(\hat{B}^T Z_j) I(T_{M_j} \geq T_{D_i}, T_{D_j} \geq T_{D_i}),$$



另外，

$$\begin{aligned}
p\hat{C}_2^\# &= \hat{P}(T_M > m, T_D \geq T_M) \\
&= \sum_{i=1}^n I(T_{D_i} \geq m) \left[\frac{\{\sum_{j=1}^n p\hat{C}_{1ij}^{**}\}}{\sum_{j=1}^n p\hat{C}_{2ij}^{**} + \sum_{j=1}^n p\hat{D}_{2ij}^{**}} \right] d\hat{F}_e(T_{D_i}), \\
p\hat{D}_1^\# &= \hat{P}(T_M > m, T_D < T_M) \\
&= 1 - \frac{1}{n} \sum_{i=1}^n I(T_{M_i} \leq T_{D_i}) - \sum_{i=1}^n I(T_{D_i} \leq m) \frac{\sum_{j=1}^n p\hat{C}_{4ij}}{\sum_{j=1}^n p\hat{C}_{2ij} + \sum_{j=1}^n p\hat{D}_{1ij}} d\hat{F}_e(T_{D_i}),
\end{aligned}$$

其中

$$p\hat{C}_{4ij} = I(T_{D_i} < T_{M_j} \leq m, T_{D_j} \geq T_{D_i}) \circ$$

$$\text{在上式中， } \hat{P}(T_M \leq T_D) = \frac{1}{n} \sum_{i=1}^n I(T_{M_i} \leq T_{D_i}) \circ$$

$\hat{P}(T_M > m)$ 除了用上述的 $p\hat{C}_2^\# + p\hat{D}_1^\#$ 估計外，還可用

$\hat{P}(T_M > m) = 1 - \hat{P}(T_M \leq m)$ 計算，故可用前節中的 $1 - (p\hat{D}_1 + p\hat{C}_2)$ 計算。

(C·4) 推導 $P(T_{D_i} > T_{M_i})$

在估計 $P(T_M > m)$ 時，會用到 $P(T_{D_i} > T_{M_i})$ ，但因此估計會受到右受

限時間的干擾，故必須用 IPCW 法估計，推導如下：

$$\begin{aligned}
P(\tilde{T}_{D_i} > \tilde{T}_{M_i}) &= P(C > \tilde{T}_{M_i}, T_{D_i} > T_{M_i}, \delta_{M_i} = 1) \\
&= C(\tilde{T}_{M_i}) P(T_{D_i} > T_{M_i}),
\end{aligned}$$

則 $P(T_{D_i} > T_{M_i}) = P(\tilde{T}_{D_i} > \tilde{T}_{M_i}) / C(\tilde{T}_{M_i})$ ，故

$$\hat{P}(T_D > T_M) = \frac{1}{n} \sum_{i=1}^n I(\tilde{T}_i > X_i, \delta_{M_i} = 1) / C(X_i) \circ$$



附錄 D 輔助定理的證明

輔助定理 1。若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則 $\tilde{h}(t | M(t), Z) = h(t | M^*(t), Z)$ 。

(證明)：根據 Cox 模式假設

$$\tilde{h}(t | M(t), Z) \equiv \tilde{h}(t | m(t), Z) ;$$

$$h(t | M^*(t), Z) \equiv h(t | m^*(t), Z) 。$$

以下分為二部分， $m(t) = 1$ 及 $m(t) = 0$ 來說明：

(1) $m(t) = 1$

$$\begin{aligned} \tilde{h}(t | m(t) = 1, Z) &= \lim_{\Delta t \rightarrow \infty} \frac{P(\tilde{T} \in [t, t + \Delta t), \delta_D = 1 | \tilde{T} \geq t, m(t) = 1, Z)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(\tilde{T} \in [t, t + \Delta t), \delta_D = 1, \tilde{T} \geq t, m(t) = 1, Z)}{\Delta t P(\tilde{T} \geq t, m(t) = 1, Z)} \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(T_D \in [t, t + \Delta t), T_D \geq t, C \geq t, T_M \leq t, T_D \geq T_M, Z)}{\Delta t P(T_D \geq t, C \geq t, T_M \leq t, T_D \geq T_M, Z)} , \end{aligned}$$

因為 $C \perp (T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ ，所以上式可表示為

$$\begin{aligned} &\tilde{h}(t | m(t) = 1, Z) \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(T_D \in [t, t + \Delta t), T_D \geq t, T_M \leq t, T_D \geq T_M, Z) P(C \geq t)}{\Delta t P(T_D \geq t, T_M \leq t, T_D \geq T_M, Z) P(C \geq t)} \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(T_D \in [t, t + \Delta t), T_D \geq t, T_M \leq t, T_D \geq T_M, Z)}{\Delta t P(T_D \geq t, T_M \leq t, T_D \geq T_M, Z)} \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(T_D \in [t, t + \Delta t) | T_D \geq t, T_M \leq t, T_D \geq T_M, Z)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow \infty} \frac{P(T_D \in [t, t + \Delta t) | m^*(t) = 1, Z)}{\Delta t} \end{aligned}$$



$$= h(t | m^*(t) = 1, Z) \circ$$

(2) $m(t) = 0$

同理，依上述方法可得

$$\tilde{h}(t | m(t) = 0, Z) = h(t | m^*(t) = 0, Z) \circ$$

根據這二式的結果，可得

$$\tilde{h}(t | m(t), Z) = h(t | m^*(t), Z) \circ$$

即 $\tilde{h}(t | M(t), Z) = h(t | M^*(t), Z) \circ$

輔助定理 2 若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則

$$P(X \leq t, \delta_M = 1 | \tilde{T} \geq t) = P(T_M \leq t | T_D \geq t) = G(t | T_D \geq t) ;$$

$$P(X > t, \delta_M = 1 | \tilde{T} \geq t) = P(T_M > t, T_D \geq T_M | T_D \geq t) \circ$$

(證明)

$$\begin{aligned} & P(X \leq t, \delta_M = 1 | \tilde{T} \geq t) \\ &= \frac{P(X \leq t, \delta_M = 1, \tilde{T} \geq t)}{P(\tilde{T} \geq t)} \\ &= \frac{P(T_M \leq t, T_D \geq t, C \geq t)}{P(T_D \geq t, C \geq t)} \circ \end{aligned}$$

因 $C \perp T_D$ 且 $C \perp T_M$ ，上式可表為

$$\begin{aligned} &= \frac{P(T_M \leq t, T_D \geq t) P(C \geq t)}{P(T_D \geq t) P(C \geq t)} \\ &= \frac{P(T_M \leq t, T_D \geq t)}{P(T_D \geq t)} \\ &= P(T_M \leq t | T_D \geq t) \end{aligned}$$



$$= G(t|T_D \geq t) \circ$$

同理，可証得

$$P(X > t, \delta_M = 1 | \tilde{T} \geq t) = P(T_M > t, T_D \geq T_M | T_D \geq t) \circ$$

輔助定理 3 若受限時間 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立時，則

$$q(X = r, \delta_M = 1, Z = z | \tilde{T} \geq u) = q(r, z | T_D \geq u) \circ$$

(証明)

$$\begin{aligned} & q(X = r, \delta_M = 1, Z = z | \tilde{T} \geq u) \\ &= \frac{P(X = r, \delta_M = 1, \tilde{T} \geq u)}{P(\tilde{T} \geq u)} \\ &= \frac{p(T_M = r, T_D \geq u, C \geq u, Z = z)}{P(T_D \geq u, C \geq u)} \circ \end{aligned}$$

因 C 和 $(T_M I(T_M < T_D), T_D, I(T_M < T_D), Z)$ 為獨立，上式可表為

$$\begin{aligned} & q(X = r, \delta_M = 1, Z = z | \tilde{T} \geq u) \\ &= \frac{p(T_M = r, T_D \geq u, Z = z)P(C \geq u)}{P(T_D \geq u)P(C \geq u)} \\ &= \frac{p(T_M = r, T_D \geq u, Z = z)}{P(T_D \geq u)} \\ &= q(r, Z | T_D \geq u) \circ \end{aligned}$$