國立臺灣大學公共衛生學院

流行病學與預防醫學研究所　生物醫學統計組

碩士論文

Division of Biostatistics, Graduate Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Master Thesis

探討半競爭風險資料下之不同特徵點比例風險模式動態預測表現

Comparison of dynamic prediction of different landmark proportional hazard model under semi-competing risks data

葉憲周

Hsien-Chou Yeh

指導教授：張淑惠　博士

Advisor: Shu-Hui Chang, Ph.D.

中華民國 108 年 8 月

August, 2019

doi:10.6342/NTU201903431

國立臺灣大學碩士學位論文
口試委員會審定書

論文中文題目

探討半競爭風險資料下之不同特徵點比例風險模式動態預測表現

論文英文題目

Comparison of dynamic prediction of different landmark proportional hazard model under semi-competing risks data

本論文係 葉憲周 君（學號 R06849031）在國立臺灣大學流行病學與預防醫學研究所完成之碩士學位論文，於民國 108 年 7 月 24 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）

蔡政安　　戴政

陳秀熙

doi:10.6342/NTU201903431

誌謝

在研究所兩年的學習裡，能夠如期完成碩士論文，需要感謝許多幫助過我的人，在此獻上我最大的敬意。

首先感謝指導教授張淑惠教授，老師從論文方向主題的思考、擬定、執行以及寫作的過程中都給我極大的幫助，在這兩年裡，總是不厭其煩地教導我在思考邏輯上的重要性以及實作過程中如何細心求證，除了學業上，老師平常也教導我許多做人的道理、做事的原則以及如何以不同的角度去看待這個世界。謝謝老師的耐心指導，讓我能順利完成學業。

此外也感謝研究室學長蘇登煌醫師，學長在論文方向上以及寫作上總能獨具慧眼地找出我許多該注意而未注意的小細節，在每週的討論中，也透過自身專業告訴我許多不同醫學知識，這都是我學習的目標。

還要感謝我的口試委員戴政教授、陳秀熙教授、蔡政安教授，謝謝老師們在我口試時給予我的提醒與指正，讓我知道自己需要加強的部分。

最後感謝我的朋友以及家人，謝謝爸爸、媽媽、妹妹在我求學生活中總是不斷地給予我支持以及鼓勵，提供我良好的環境，讓我能無後顧之憂地繼續往前邁進，安心順利的完成學業，達成我人生階段性的目標。

doi:10.6342/NTU201903431

中文摘要

在實際臨床應用上，準確地預測病患預後是一非常重要的議題。例如癌症病患的復發情況等生物標誌(biological marker)，為隨時間變動的資訊，為臨床介入治療以及預測未來風險有用的指標。本文在不同特徵時間點(landmark time)下建構一系列以 Cox 為基礎型式的特徵點模式(landmark model)，並加入隨時間變動標誌資訊進行未來存活機率之動態預測(dynamic prediction)。本文特別之處為，在不同特徵時間點下，建構以 Cox 為基礎型式之特徵點模式時，針對病患進入研究後至發生標誌事件之標誌時間(marker time)轉化為時間分段函數，更運用此隨時間改變之資訊納入模式，此想法對於在進行未來存活機率預測時，針對時間依賴性資訊處理會更為貼切。在模擬部分，考慮三種標誌時間及存活時間之間不同的相關性結構，並以文中所建構 Cox 為基礎型式的特徵點模式進行動態預測，比較其表現。最後文中也以大腸直腸癌及阿茲海默症實際資料為例，探討本文所考慮之特徵點模式其動態預測的表現。

關鍵字：動態預測、特徵點模式、多階段模式

doi:10.6342/NTU201903431

英文摘要

An important issue in clinical practice is to accurately predict the prognosis of patients in order to aid clinical decision-making. Biological markers, for example, recurrences in cancer patients, often serve as time-dependent information in the need of clinical intervention and the usefulness of the prediction of future risk. We consider several landmark Cox-type models at a sequence of landmark times to incorporate the time-dependent marker information for dynamic prediction of future survival probabilities. In particular, a piecewise function of the time to a marker included in the Cox-type model at each landmark time may be more adaptable to use the time-dependent marker information for predicting future survival probabilities. In simulation study, we consider three different correlation structures between the marker and survival times to assess the performance of dynamic prediction based on different landmark Cox-type models. Finally, we use colon cancer and dementia data to explore the dynamic prediction abilities under these landmark Cox-type models.

Key words: Dynamic prediction, landmark model, multi-state model

doi:10.6342/NTU201903431

目錄

口試委員審定書 i

誌謝 ii

中文摘要 iii

英文摘要 iv

圖目錄 vi

表目錄 ixx

第一章 研究動機與目的 1

第二章 文獻回顧 4

第一節 以標誌時間為預測因子的條件存活機率估計 4

第二節 標誌歷程之動態存活預測的統計分析 6

第三節 ROC 曲線 7

第四節 Brier 分數 8

第五節 以時間相依共變數比較聯合模式及特徵點模式預測結果 10

第六節 疾病進展資料下比較特徵點模式及聯合模式預測結果 11

第三章 研究方法 12

第四章 模擬 15

第一節 資料生成 15

第二節 模擬結果 18

doi:10.6342/NTU201903431

第五章 實例分析 ……29
第一節 大腸直腸癌研究 ……29
第二節 PAQUID 研究 ……35
第六章 討論與總結 ……42
第七章 參考文獻 ……44
附錄 ……48

doi:10.6342/NTU201903431

圖目錄

圖 1-1：半競爭風險資料路徑圖 1

圖 1-2：右設限半競爭風險資料 2

圖 3-1：將標誌時間 T_M 進行分段估計時，分段方式示意圖 13

圖 4-1：特徵時間點 $u = 4$，五種特徵點模式的預測存活曲線圖(時間依賴性 Cox 模式) 19

圖 4-2：特徵時間點 u 為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現(時間依賴性 Cox 模式) 20

圖 4-3：特徵時間點 u 為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數(時間依賴性 Cox 模式) 20

圖 4-4：在特徵時間點 u 為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(時間依賴性 Cox 模式) 21

圖 4-5：在特徵時間點 u 為 2 至 5 年下，已發生標誌事件個案五種模式預測 5 年後存活機率均方差(時間依賴性 Cox 模式) 21

圖 4-6：在特徵時間點 u =4，五種特徵點模式的預測存活曲線圖(對數常態) ..23

圖 4-7：特徵時間點 u 為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現(對數常態) 24

圖 4-8：特徵時間點 u 為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數(對數常態) 24

圖 4-9：在特徵時間點 u 為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(對數常態) 25

圖 4-10：在特徵時間點 u 為 2 至 5 年下，已發生標誌事件個案五種模式預測 5 年後存活機率均方差(對數常態) 25

doi:10.6342/NTU201903431

圖 4-11：在特徵時間點$u = 4$，五種特徵點模式的預測存活曲線圖(Clayton Copula) 26
圖 4-12：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現(Clayton Copula).....27
圖 4-13：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數(Clayton Copula).....27
圖 4-14：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(Clayton Copula).....28
圖 4-15：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(Clayton Copula).....28
圖 5-1：大腸直腸癌資料結構圖 30
圖 5-2：特徵時間點$u = 2$年下，以五種特徵點模式預測個案未來 7 年存活機率的存活曲線圖 31
圖 5-3：特徵時間點$u = 3$年下，以五種特徵點模式預測個案未來 6 年存活機率的存活曲線圖 31
圖 5-4：特徵時間點$u = 4$年下，以五種特徵點模式預測個案未來 5 年存活機率的存活曲線圖 32
圖 5-5：特徵時間點$u = 5$年下，以五種特徵點模式預測個案未來 4 年存活機率的存活曲線圖 32
圖 5-6：特徵時間點$u = 2$、3、4、5下，以五種特徵點模式預測未來 2 年存活機率 AUC 表現.....33
圖 5-7：特徵時間點$u = 2$、3、4、5下，以五種特徵點模式預測未來 2 年存活機率 Brier 分數表現 33
圖 5-8：特徵時間點$u = 2$、3、4、5下，未復發個案，五種特徵點模式預測未來 4 年存活機率均方差表現 34

doi:10.6342/NTU201903431

圖 5-9：特徵時間點$u = 2$、3、4、5下，已復發個案，五種特徵點模式預測未來 4 年存活機率均方差表現 34
圖 5-10：PAQUID 研究資料路徑圖 36
圖 5-11：特徵時間點$u = 10$年下，以五種特徵點模式預測個案未來 10 年存活機率的存活曲線圖 38
圖 5-12：特徵時間點$u = 11$年下，以五種特徵點模式預測個案未來 9 年存活機率的存活曲線圖 38
圖 5-13：特徵時間點$u = 12$年下，以五種特徵點模式預測個案未來 8 年存活機率的存活曲線圖 39
圖 5-14：特徵時間點$u = 13$年下，以五種特徵點模式預測個案未來 7 年存活機率的存活曲線圖 39
圖 5-15：特徵時間點$u = 10$、11、12、13下，以五種特徵點模式預測未來 4 年存活機率 AUC 表現.....40
圖 5-16：特徵時間點$u = 10$、11、12、13下，以五種特徵點模式預測未來 4 年存活機率 Brier 分數表現 40
圖 5-17：特徵時間點$u = 10$、11、12、13下，未被判定為阿茲海默症個案，五種特徵點模式預測未來 4 年存活機率均方差表現 41
圖 5-18：特徵時間點$u = 10$、11、12、13下，被判定為阿茲海默症個案，五種特徵點模式預測未來 4 年存活機率均方差表現 41

doi:10.6342/NTU201903431

表目錄

表 3-1：本研究建構之五種特徵點模式，表格中 Label 表示模式簡稱.....14

表 5-1：大腸直腸癌資料變數介紹.....30

表 5-2：PAQUID 資料變數介紹 37

doi:10.6342/NTU201903431

第一章 研究動機與目的

在近年醫學研究中，希望可以根據病患們的個別生物標誌(biomarkers)，建構一個「個人化醫療」的判斷工具，用以預測未來存活時間長短，並且隨著時間增加，將病患新的生物標誌資訊加入模型中，達到不斷更新對於未來時間點存活機率預測的目的。

舉例來說，Dancourt 等人 (2004)在針對大腸直腸癌病患的研究中，以病患在接受手術後有無發生復發事件當作一重要生物標誌放入模式中，評估有無復發對病患死亡風險的影響，在此例中，病患可能會先發生復發再死亡或未發生復發而直接死亡，這種情況下的資料結構稱為半競爭風險資料 (semi-competing data)，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) (圖 1-1)，而若考慮右設限情況，所有可能觀察到的資料情況如(圖 1-2)。

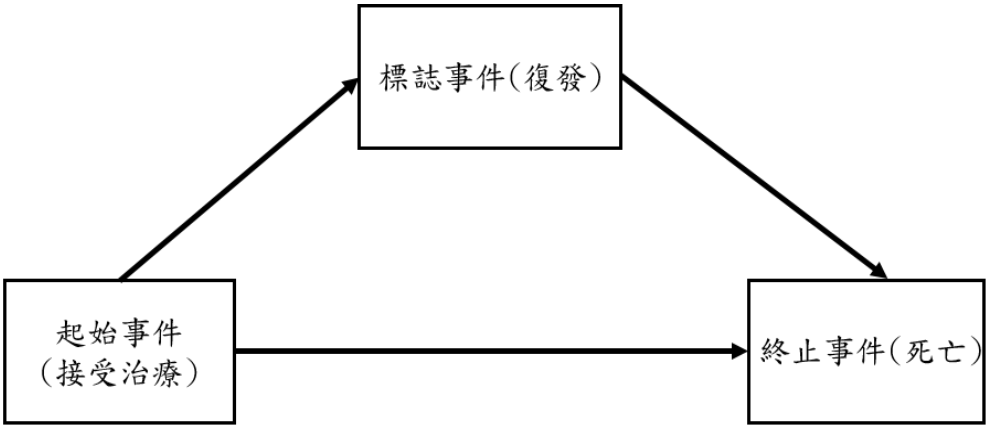


圖 1-1：半競爭風險資料路徑圖

doi:10.6342/NTU201903431

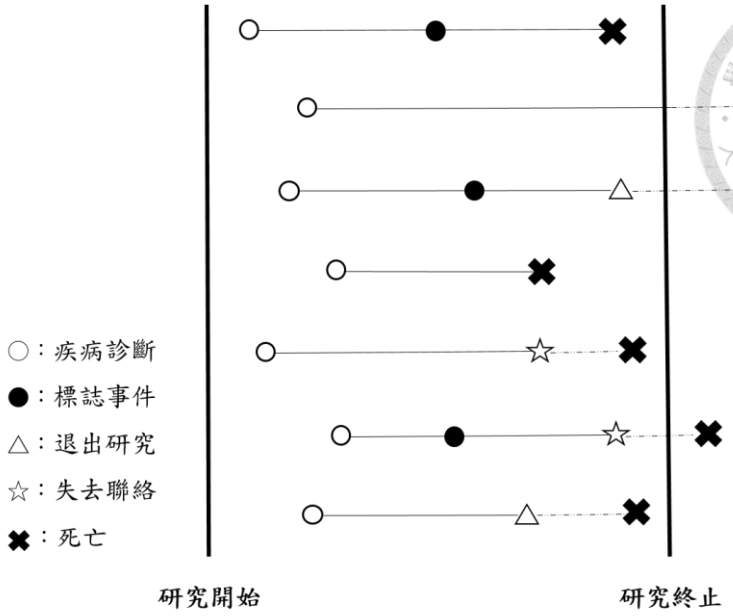


圖 1-2：右設限半競爭風險資料

如果在每次診療過程中能夠透過這些標誌事件訊息來了解病患未來存活的情形，則可以採取適當的醫療措施。由標誌事件所形成的歷程稱為標誌歷程 (marker process) (Whitmore, Crowder and Lawless, 1998)。雖然我們所關心的是病患在未來存活機率的表現，但是在疾病發展過程中，常面臨到病患在疾病進展過程中發生了某種標誌事件，對於其存活機率會不會有影響？或是此標誌事件發生的時間早晚對其存活機率的影響？而 Neri 等人 (2007)在乳癌的研究中發現，病患在接受手術後前 2 年發生復發的存活機率比術後 2 年之後發生復發的存活機率還低，因此復發越晚存活時間越久，由此可看出復發時間的長短也會對於病患未來存活時間造成影響。故 Peng (2010)在半競爭風險資料結構下，使用 Cox 模式，並以標誌事件及標誌時間長短為解釋變數，也就是用過去標誌歷史為預測因子，預測未來條件存活函數，比起只用是否發生標誌事件作為解釋變數，文中加入標誌時間的變項優勢為能夠討論標誌時間長短對於存活機率的影響。而在實際分析中也以過去 50、80、100、200 天前有無接受心臟病移植與等待時間存活的條件下，估計其未來時間的條件存活函數，藉以探討標誌時間長短對於死亡機率的影響。

doi:10.6342/NTU201903431

在實際應用上，建立對病患在未來某時間點的條件存活機率的預測模式，如何說服醫師以此預測方式是有一定準確度存在，足以作為診斷上判斷依據乃是重要的議題，因此本文研究的目的為建構不同的特徵點模式(landmark model) (van Houwelingen, 2007)後比較其預測出來的存活機率與實際存活情形之一致性。類似文章如 Suresh 等人(2017)，在半競爭風險資料結構下，以多階段聯合模式(multi-state joint model)與提出複合特徵點模式，對其後續存活機率預測進行比較，以 AUC (Area Under the Curve) (Hanley and McNeil, 1982)及 R-squared 進行模型準確度的評估，而在複合特徵點模式建構上，考慮了將不同特徵時間點(landmark time)下之資料堆疊後組成超級資料集(super dataset)，優點為可以保有較多資料進行估計，文中以此模式進行預測需要假設在每個特徵時間點下，所考慮的共變數對於未來存活機率的影響皆相同，但在真實情況中，標誌時間發生的早晚對於不同特徵時間點之條件存活機率的影響不會相同，因此本研究也希望除了進行模式比較外，擬設計較為彈性的特徵點模式，針對不同標誌時間之時間區間作分段(piecewise)估計，並以 AUC、Brier 分數(Graf et al., 1999)及均方差(Mean Square Error)探討評估這些模式所得預測存活機率之表現。Peng (2010)已建構出以標誌事件及標誌時間作為預測因子之特徵點模式，唯獨模式驗證部分需要做更進一步確認，因此在此基礎下，希望能夠將此模式最後驗證部分完成。本文也將延伸此模型，針對不同標誌時間時間區間作分段估計。在模擬部分，也在資料生成時考量標誌歷程與存活時間之間三種不同的相關性模式，最後藉由 AUC、Brier 分數及均方差判斷在給定這些預測因子下，模式所預測出來的存活機率與實際存活情形之一致性。最後，本文也以阿茲海默症(Dartigues. et al., 1992)及大腸直腸癌(Laurie. et al., 1989)等兩筆資料，進行實際分析，此兩筆資料差異為標誌事件發生長短不同以及標示時間及事件時間的 Kendall’s tau 相關係數不同，因此也可以藉著此兩筆資料比較本文建構之特徵點模式預測存活機率之表現。

doi:10.6342/NTU201903431

第二章 文獻回顧

本章第一節將回顧 Peng (2010)在半競爭風險資料結構下以標誌時間為預測因子之條件存活機率估計，第二節回顧 Su (2016)利用共變數和標誌歷程的分布，給予適當權重，預測未來終止事件發生機率，第三節及第四節則是回顧本文所使用到的評估準則，分別是 Heagerty and Zheng (2005)提出應用於存活模式中的 ROC 曲線以及 Brier 分數判斷模式預測力高低，第五節則是回顧 Rizopoulos 等人(2017)以時間相依共變數建構聯合模式、特徵點模式以及文中所提出混合特徵點模式(mixed landmark model)，此三種模式在預測能力上的比較。最後一節回顧 Suresh 等人(2017)探討以不同馬可夫模式生成半競爭風險資料後，比較特徵點模式及聯合模式之預測表現。

第一節 以標誌時間為預測因子的條件存活機率估計

在疾病發展過程中，復發對於死亡風險的影響在存活分析上扮演重要的角色。有些研究使用 Cox 模式，將復發當成固定時間下的指標函數作為解釋變數，來觀察其對死亡風險的影響。Yang (2008)在文章中以任意時間點為切點，判斷在此切點前後有無發生特定狀態之指標函數當作解釋變數放入 Cox 模式中，達到預測後續存活機率的目的。但這些方法只能判斷特定狀態對死亡的影響，並無法知道特定狀態時間長短對死亡的影響。所以，Peng (2010)在半競爭風險資料結構下，提出以標誌時間為預測因子，關心特定狀態時間長短對死亡的影響。此文章建立的模式有兩種：

1. 若只有在過去u時間的標誌狀態$m^*(u)$與在未來t時間點的死亡風險函數有關，考慮以下 Cox 模式

$$h\big(t\big|m^*(u)\big) = h_{0u}(t)exp(\beta m^*(u)I(u < t)) \qquad \textbf{(2-1)}$$

doi:10.6342/NTU201903431

$$m^*(u)=\begin{cases}1, \text{在過去}u\text{時間已發生標誌事件}\\0, \text{在過去}u\text{時間尚未發生標誌事件}\end{cases}$$

當在u時間還未發生標誌事件，即$m^*(u)=0$，由模式可得到在過去u時間還未發生標誌事件的條件下，未來t時間的死亡風險為基準風險函數$h_{0u}(t)$，而存活當在u時間已發生標誌事件，即$m^*(u)=1$，由模式可得到在過去u時間已發生標誌事件的條件下，未來t時間的死亡風險為$h_{0u}(t)exp(\beta)$。

2. 若過去u時間的標誌狀態$m^*(u)$及已有標誌時間與在未來t時間點的死亡風險函數有關時，考慮以下 Cox 模式

$$h\big(t\big|m^*(u)\big)=h_{0u}(t)exp((\beta_1 m^*(u)+\beta_2 T_M m^*(u))I(u<t)) \quad \textbf{(2-2)}$$

其中T_M表示病患從進入研究後至發生標誌事件時間的長度。當在u時間還未發生標誌事件，即$m^*(u)=0$，由模式可得到在過去u時間還未發生標誌事件的條件下，未來t時間的死亡風險為基準風險函數$h_{0u}(t)$，當在u時間已發生標誌事件，即$m^*(u)=1$，由模式可得到在過去u時間已發生標誌事件的條件下，未來t時間點的死亡風險為$h_{0u}(t)exp(\beta_1+\beta_2 T_M)$。

Peng (2010)以 Stanford Heart Transplant 資料為例，目的為探討在過去時間點$u=50$、80、100、200天前有無接受心臟移植的條件下，估計未來時間的條件存活函數，並以上述兩模式進行模式配適，觀察是否等待接受心臟移植的時間對於死亡有風險有影響。

Peng 在此篇文章最大之貢獻為將標誌時間納入模式中作為預測因子，考量到標誌時間長短對於未來存活時間影響不同。但若要在實際應用上提供醫生良好的判斷依據，需要進一步確認模式在預測能力上是否有達到一定的水準，所以，本研究希望可以比較這些模式所預測出來的存活機率與實際存活情形之一致性，並且延伸 Peng 本篇文章的想法，設計出更有彈性的模式，並了解其在實際應用層面上的表現。

doi:10.6342/NTU201903431

第二節 標誌歷程之動態存活預測的統計分析

在臨床上常面臨的問題是假如發生了某種標誌事件，對病患的預後會產生什麼影響？抑或是發生標誌事件時間的早晚是不是也會跟其往後存活機率有關，這些例子都是病患或是醫師想了解的問題。

類似以標誌歷程預測未來存活機率的方式，Lakhal 等人(2008)利用 Archimedean copula 的模式，主要著墨於標誌時間的估計以及未發生標誌事件時，未來發生終止事件機率，對於已發生標誌事件情形並未著墨。Porta 等人(2012)以多階段模式及 Cox 模式做預測，但在做標誌事件對終止事件預測時，並未對基礎共變數做處理。Maugeun 等人(2013)利用聯合脆弱模式及 Cox 模式做預測，優點在於利用脆弱性變數去除復發和死亡間相關性，但模式只能針對特定基礎共變數做預測。如果要針對基礎共變數及標誌歷程的變化做未來終止事件的預測必須分別模式化基礎共變數與各事件間的關係，這是有難度的。且時間依賴性 Cox 模式為一即時解釋模式，並無法達到用過去標誌歷程預測未來存活機率的目的。因此 Su (2016)採用時間依賴性 Cox 模式運用貝氏定理作機率的反轉及條件機率的運算來處理這問題，利用共變數及標誌歷程之條件式分布，在發生終止事件的時間點上給予適當的權重，便可在不同標誌歷程及共變數條件下，估計未來條件存活函數。文中除了將標誌歷程視為時間依賴性變數，進而建構其和終止事件的相關性，也可同時處理基礎共變數對其影響。

考慮以下兩種情境：

1. 考慮終止事件的發生，只跟標誌事件有關，Cox 模式為

$$h\big(t\big|M^*(t)\big) \equiv h(t|m^*(t), T_M) \equiv h_0(t)exp\{\beta_1 m^*(t) + \beta_2 m^*(t)k(T_M)\} \quad \textbf{(2-3)}$$

其中$k(T_M)$為標誌時間T_M的函數。

2. 考慮終止事件的發生，與共變數與標誌事件皆有關，Cox 模式為

doi:10.6342/NTU201903431

$$h(t|M^*(t),Z) \equiv h(t|m^*(t),T_M,Z)$$
$$\equiv h_0(t)exp\{B^TZ + \beta_1 m^*(t) + \beta_2 m^*(t)k(T_M)\} \quad \textbf{(2-4)}$$

在針對右設限資料，考慮基礎共變數情況下，推導$S(t|T_M \le m)$以及$S(t|T_M > m)$條件存活機率。當$t > m$時，利用標誌歷程預測未來存活函數：在推導$S(t|T_M \le m)$以及$S(t|T_M > m)$之前，需要先計算分母$P(T_M \le m)$以及$P(T_M > m)$部分，但在計算$P(T_M \le m)$時，$P(T_M \le m, T_M > T_D)$無法觀察到且$P(T_M \le m, T_M \le T_D)$部分若用無母數計算並無法充分運用到足夠訊息而需要大量資料才可運算，因此 Su 在文中透過模式假設即可充分運用訊息且可以估算$P(T_M \le m, T_M > T_D)$，解決了個案數過少而造成估計不穩定的問題。

本研究參考文中資料生成部分以時間依賴性 Cox 模式生成資料，將標誌歷程當時間依賴性變數，接著以不同特徵點模式探討預測出來的存活機率與實際存活情形之一致性。

第三節 ROC 曲線

除了建構模式外，我們還希望可以了解模式中預測因子與感興趣結果間關係的強弱，針對連續型變數可以計算兩者之間的相關係數做判斷，當感興趣結果為二元變數時，則可以藉由繪製 ROC 曲線後以曲線下面積(AUC)做判斷。但當感興趣結果為存活時間，由於個案可能會因為各項因素而未觀察到終止事件的發生，因此需要考量到設限的情況，故對於真陽性(sensitivity)以及真陰性(specificity)在計算上也不同。針對時間相依性真陽性及真陰性計算，Etzioni 等人(1999)以及 Slate and Turnbull (2000)提出 *Incident / static* 想法，針對病例組(case)，在每個事件發生的時間點t都計算一次真陽性，而控制組(control)真陰性的計算則是固定在單一個時間點t^*下做計算，但在真實狀況中，研究早期位於控制組之個案也可能會因為在後期發生事件而轉移到病例組中，因此對於真陰性計算也希望能隨著時間而改變。Heagerty and Zheng (2005)設計以 *Incident / dynamic* 方式計算的時間

doi:10.6342/NTU201903431

相依的真陽性以及真陰性。在此情況下之真陽性、真陰性定義為：

$$sensitivity(c,t): P(M > c|T_D = t) = P(M > c|dN^*(t) = 1)$$
$$specificity(c,t): P(M \le c|T_D > t) = P(M \le c|N^*(t) = 0) \quad \textbf{(2-5)}$$

根據此想法，在 ROC 曲線部分，可表示為

$$ROC_t(p) = TP_t\left\{[FP_t]^{-1}(p)\right\} \quad \textbf{(2-6)}$$

TP 為真陽性，$FP(1-specificity)$ 為偽陽性，$[FP_t]^{-1}(p) = inf_c\{c : FP_t(c) \le p\}$，$p$ 表示偽陽性率。

故針對在各事件發生時間點 t 之下 AUC(t)可定義為

$$AUC(t) = \int_0^1 ROC_t(p)\,dp$$
$$= \frac{\#\{j \in R(t) : x_j < x_i\} + 0.5 * \#\{j \in R(t), j \ne i; x_j = x_i\}}{Y(t) - 1} \quad \textbf{(2-7)}$$

其中 x_i & x_j 代表第 i & j 個案經由模型判斷後的預測後指標，藉由與時間相關之曲線下面積，可以觀察在給定標誌時間點 u 後，預測在未來 t 時間點時模式預測之準確度。

此方法優點為針對在各事件發生時間點 t 之下，存在於風險集合中之病患進行分類，更貼近風險模型的想法，且若要針對時間相依共變數進行診斷，此方法也更容易計算。故本研究在 ROC 曲線計算上採用 *Incident / dynamic* 方式。

第四節 Brier 分數

在評估模型預測能力上，直覺做法為給定病患共變數(X)情況下，評估病患實際存活時間與預測存活時間的差距，此評估準則估計方式表示為

$$\frac{1}{n}\sum_{i=1}^{n}\left(T_i - T(X_i)\right)^2 \quad \textbf{(2-8)}$$

在$\hat{T}(X_i)$估計上，通常使用存活曲線中，病患存活時間的中位數代表，但以此方

doi:10.6342/NTU201903431

式計算必然會造成估計不穩的現象。另一方式為預測該位病患在 t 時間點的存活情形，設立切點c，預測病患在t時間點存活機率($\hat{\pi}(t|X=j)$)，轉換為指標函數$\hat{Y}(X)=I(\hat{\pi}(t|X=j)>c)$後，與實際是否存活狀況做比較，表示方式為

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i \neq Y\left(X_i\right)\right) \quad \textbf{(2-9)}$$

此方法透過計算預測結果與真實狀況不一致的比例來評估模型預測能力，但此方法在研究一開始$t=0$時，對於病患在t時間點的真實狀態並無法得知，因此在判斷病患真實狀況$Y=I(T>t)$上會有困難。

Hilden 等人 (1978)、Spiegelhalter 等人 (1986)以及 Linnet 等人 (1989)在文章中提到，以「某時間點預測存活機率值」來替代「某時間點預測是否存活」更能真實反應模型預測力的表現，因此選擇藉由 Brier 分數，評估真實存活情況與預測存活機率值的離散程度，表示方式如下：

$$BS(t)=\frac{1}{n}\sum_{i=1}^{n}\left(I\left(T_i>t\right)-\pi\left(t|X_i\right)\right)^2 \quad \textbf{(2-10)}$$

但在實際觀察到的資料中，會遇到設限的問題，實際遇到狀況有以下三種：

1. $T_i \leq t$ and $\delta_i = 1$
2. $T_i > t$ and ($\delta_i = 1$ or $\delta_i = 0$)
3. $T_i \leq t$ and $\delta_i = 0$

其中$\delta = I(T < C)$為事件發生指標函數，在這三種情境下，第三種狀況是無法得知的，由於設限發生點在t之前，因此在此種情況下之 Brier 分數也無法計算。為了彌補由於設限情形發生而造成資料損失的狀況，對於以上每一種類別給予適當的權重作計算，更能符合真實情形。

首先定義設限分布$G(t)=P(C>t)$，對於第一種情況設限權重為$1/\hat{G}(T_i)$，第二種為$1/\hat{G}(t)$，因此 Brier 分數重新調整後為

$$BS^C(t)=\frac{1}{n}\sum_{i=1}^{n}\left\{\left(0-\pi\left(t|X_i\right)\right)^2 I\left(T_i \leq t,\delta_i=1\right)\left(\frac{1}{G\left(T_i\right)}\right)+\left(1-\pi\left(t|X_i\right)\right)^2 I\left(T_i>t\right)\left(\frac{1}{G(t)}\right)\right\} \quad \textbf{(2-11)}$$

doi:10.6342/NTU201903431

第五節 以時間相依共變數比較聯合模式及特徵點模式預測結果

Rizopoulos 等人(2017)以特徵點模式、聯合模式進行動態存活機率預測，並以主動脈瓣膜資料進行實際分析比較。以聯合模式進行預測需要針對個案重複觀測資訊(文中以主動脈梯度為例)進行模式之假設，大多以線性混合模式為基礎，此方法優勢為能夠使用到較多資訊進行預測，但遇到較大困難為計算複雜耗時且如果模式假設錯誤，對於預測結果則會造成影響。另一常見方法為特徵點模式，多以 Cox 模式為基礎，將解釋變數與存活時間連結起來，優勢為計算簡單快速，但遇到的困難則是在建立模式時會有損失資料風險且對於重複觀測資訊較無法有效利用。本篇文章中另外提出了混合特徵點模式，在聯合模式及特徵點模式間取得折衷的想法。針對在標誌時間點u仍處於風險集合中的個案，皆配適線性混合模式後再以 Cox 模式進行預測。

在實際分析上，本篇文章在預測上針對個案重複觀測資訊的處理提供了三種不同的想法，分別為直接以線性混合模式估計的結果代入、考慮模式估計完結果外也以短時間內之變化量作為參考依據，最後則是考慮在預測時間點t以前累積的資訊量做為參考依據。在預測結果上，三種模式 AUC 表現大多介於 0.5 至 0.6 之間，推測可能原因為該篇文章並未準確捕捉到病患主動脈梯度變化的資訊或是只考慮治療組別以及主動脈梯度兩個參數不足以提供良好預測能力。而本研究也承接此篇文章想法，除了進行各模式間比較外，也以不同想法建構特徵點模式。希望藉由標誌事件的發生以及針對標誌時間做較彈性化的處理後，建立各種特徵點模式進行未來存活機率的預測並做後續模式間的比較。

doi:10.6342/NTU201903431

第六節 疾病進展資料下比較特徵點模式及聯合模式預測結果

Suresh 等人(2017)以特徵點模式及聯合模式進行動態存活機率預測，並比較其預測的能力。以聯合模式進行預測，需要針對生物標誌進行模式的假設，並與存活模式連結，建構出聯合機率分布，以此模式進行存活機率的估計尚須考量到對於生物標誌模式假設的正確性，若生物標誌實際分布與模式差異很大，對於預測結果也會有很大影響，且此方法在計算上需要考慮到數值積分及計算上穩定性，故也較為複雜。進行動態預測還有另一種常用模式，稱為特徵點模式，van Houwelingen (2007) 以 Cox 模式針對在特徵時間點 u 仍然存活的個案預測其在未來 $u+w$ 時間點(w 為預測區間)存活機率。以特徵點模式進行預測可以免除對於生物標記的模式假設，如此以來可以減輕計算上的複雜度及參數估際上難度，但此方法限制為對於在每一個特徵時間點前已離開研究的個案則無法繼續納入模型估計中，因此會有損失資料的問題。在過去文獻中，也有針對比較特徵點模式與聯合模式在預測能力的表現，如 Cortese 等人 (2013)在競爭風險資料結構下，比較兩模式累積事件函數差異。Rizopoulos 等人 (2013)對於生物標誌嘗試以不同線性模式假設，再進行預測，發現普遍而言聯合模式之預測能力都較特徵點模式更佳。Suresh 等人則承接以上想法，以不同結構之馬可夫及半馬可夫模式生成資料，探討特徵點模式及聯合模式在這些不同資料結構下預測能力的表現。而本文也延續此想法，以 Su (2016)文中使用時間依賴性 Cox 模式生成資料外，也以其他不同相關性結構方式生成資料，並比較不同特徵點模式預測出來的存活機率與實際存活情形之一致性。

doi:10.6342/NTU201903431

第三章 研究方法

本文目標為在給定個案某特徵時間點u下的標誌事件歷史其後續存活機率之動態預測，以T_D表示為從疾病診斷到發生死亡的時間，為存活時間(survival time)，T_M表示從疾病診斷到發生標誌事件時間，稱為標誌時間(maker time)。此處定義其標誌事件$m^*(s)=I(T_M \le s, T_M < T_D)$ 以及其標誌歷史

$M^*(u)=\{m^*(s), 0 \le s \le u\}$，即估計以下條件存活函數：

$$\pi(u,t)=P\left(T_D \ge t \middle| T_D > u, M^*(u)\right),\ t > u \qquad \textbf{(3-1)}$$

又因為，此條件存活機率又可表示如下

$$\pi(u,t)=P\left(T_D > t \middle| T_D > u, M^*(u)\right)=exp\left\{-\int_u^t h_u\left(s \middle| M^*(u)\right)ds\right\} \qquad \textbf{(3-2)}$$

在此，$h_u\left(t \middle| M^*(u)\right)=\lim_{dt\to 0}\dfrac{P\left(T_D \in [t, t+dt) \middle| T_D \ge t, M^*(u)\right)}{dt},\ t > u$ **(3-3)**

為給定個案在某特徵時間點u的標誌事件歷史之條件風險函數。

本章節將介紹針對此條件風險函數$h_u\left(t \middle| M^*(u)\right)$考慮不同之特徵點 Cox 模式，以此得到$\pi(u,t)$ 的估計。首先，考慮最基本特徵點比例風險模式，當中只包含在特徵時間點u之前是否發生標誌事件之指標函數，模式表示如下：

$$h_u\left(t \middle| M^*(u)\right)=h_{0u}(t)exp\left(\beta_{1u}m^*(u)\right) \qquad \textbf{(3-4)}$$

在(3-4)的模式，並未使用到標誌時間的資訊，故本文接續考慮另外兩種模式：

除了以在u時間點前是否發生標誌事件作為預測因子外，還可以加入標誌時間T_M進行討論，本文在標誌時間點操作上考慮以下兩種方式：

1. 直接將標誌時間T_M當作共變數放入模式中，模式表示如下：

$$h_u\left(t \middle| M^*(u)\right)=h_{0u}(t)exp\left(\beta_{1u}\cdot m^*(u)+\beta_{2u}\cdot m^*(u)\cdot T_M\right) \qquad \textbf{(3-5)}$$

doi:10.6342/NTU201903431

2. 將標誌時間內做分段，將不同標誌時間區間分段估計，對於未來存活機率影響不同。模式表示如下：

$$h_u\left(t\middle|M^*(u)\right)=h_{0u}(t)exp\left(\sum_{j=1}^{J+1}\beta_{ju}\cdot I\left(T_M\in I_j\right)\cdot m^*(u)\cdot g\left(T_M\right)\right) \quad \textbf{(3-6)}$$

其中，$I_j=\left(t_{j-1},t_j\right],t_j\le u, j=1,2,...J+1$，表示標誌時間之分段區間，$t_j$ 表示標誌時間 T_M 的切點且 t_0，J 表示切點個數，示意圖如下，

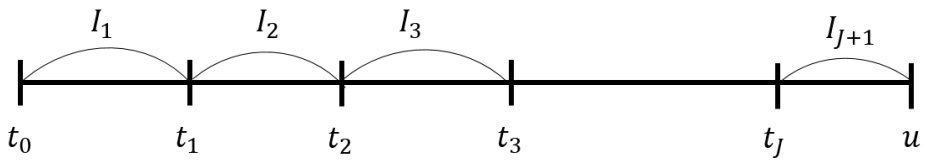


圖 3-1：將標誌時間 T_M 進行分段估計時，分段方式示意圖

而 $g(x)$ 為已知函數，例如，$g(x)=log(x)$，本文藉由調整 $g(x)$ 讓預測模式能更符合資料真實存活情形。此模式優點除了能描述在不同標誌時間區間發生標誌事件的個案對其往後存活機率有不同影響外，也可以免除 Suresh 等人 (2017) 以特徵點超級模式 (landmark super model)進行預測時，需以特徵時間點為分層依據，並假設每個特徵時間點下，模式中共變數對存活機率影響皆相同。

建構上述模式時，於第四章模擬部分觀察到，當將標誌時間作分段估計，若要針對已發生標誌事件個案存活機率作調整，在不同相關性資料結構情況下會影響未發生標誌事件之個案存活機率預測在均方差的表現，但我們仍希望建構的模式在預測上對於兩者皆同時達到一定的水準。故本文以標誌事件作分層依據，將未發生標誌事件與已發生標誌事件兩種情況分開估計，模式表示為：

$$\begin{aligned}&h_u\left(t\middle|M^*(u)\right)\\&=\left[h_{01u}\left(t,I\left(m^*(u)=0\right)\right)+h_{02u}\left(t,I\left(m^*(u)=1\right)\right)\right]\\&\qquad exp\left(\sum_{j=1}^{J+1}\beta_{ju}\cdot I\left(T_M\in I_j\right)\cdot m^*(u)\cdot g\left(T_M\right)\right)\end{aligned} \quad \textbf{(3-7)}$$

上述之模式均可以右設限半競爭風險資料建構其部分概似函數(partial likelihood)，估計出迴歸係數後估計出其條件存活函數 $\pi(u,t)$。

doi:10.6342/NTU201903431

本研究建構之模式整理如下表：

Model		Label
landmark model	$h\left(t\middle\vert M^{*}(u)\right)=h_{0u}(t)exp\left\{\beta_{1u}\cdot m^{*}(u)\right\}$	LM1
	$h\left(t\middle\vert M^{*}(u)\right)=h_{0u}(t)exp\left\{\beta_{1u}\cdot m^{*}(u)+\beta_{2u}\cdot m^{*}(u)\cdot T_{M}\right\}$	LM2
	$h\left(t\middle\vert M^{*}(u)\right)$ $=h_{0u}(t)exp\left\{\sum_{j=1}^{J+1}\beta_{ju}\cdot I\left(T_{M}\in I_{j}\right)\cdot m^{*}(u)\cdot g\left(T_{M}\right)\right\}$	LM3
stratified landmark model	$h\left(t\middle\vert M^{*}(u)\right)$ $=\left[h_{01u}\left(t,I\left(m^{*}(u)=0\right)\right)+h_{02u}\left(t,I\left(m^{*}(u)=1\right)\right)\right]exp\left\{\beta_{1}\cdot m^{*}(u)\cdot T_{M}\right\}$	LSM1
	$h\left(t\middle\vert M^{*}(u)\right)$ $=\left[h_{01u}\left(t,I\left(m^{*}(u)=0\right)\right)+h_{02u}\left(t,I\left(m^{*}(u)=1\right)\right)\right]\cdot$ $exp\left\{\sum_{j=1}^{J+1}\beta_{ju}\cdot I\left(T_{M}\in I_{j}\right)\cdot m^{*}(u)\cdot g\left(T_{M}\right)\right\}$	LSM2

表 3-1：本研究建構之五種特徵點模式，表格中 Label 表示模式簡稱

doi:10.6342/NTU201903431

第四章 模擬

為了瞭解估計量在右設限資料中是否仍然有穩健的估計，必須模擬生成右設限資料。在模擬部分，根據本文所使用之資料結構，必須生成標誌狀態、標誌時間、存活時間、右設限時間，以下分為兩小節說明，第一節介紹不同相關性結構下模擬資料生成步驟，第二節為模擬結果。

第一節 資料生成

- **時間依賴性 Cox 模式**

在產生資料時，主要生成標誌時間、存活時間及右設限時間，並根據生成的資料形成觀察資料，資料產生步驟如下：

1. 假設標誌時間T_M服從指數分布，$T_M \sim exp(\lambda_M)$，$\lambda_M = 0.05$，方法為以均勻分布產生變數(u_{T_M})，$T_M = -ln(u_{T_M})/\lambda_M$

2. 假設風險函數為

$$h\left(t\middle|m^*(t)\right) = h_0(t)\exp(\beta_1 m^*(t)) \qquad \textbf{(4-1)}$$

其中$m^*(t) = I(T_M < T_D, T_M \leq t)$，令$h_0(t) = exp(-2.5)$、$\beta_1 = 1.5$以及2.8。即在$t$時間點下，未發生標誌事件之風險函數為$h\left(t\middle|m^*(t)\right) = exp(-2.5)$，而發生標誌事件之風險函數為$h\left(t\middle|m^*(t)\right) = exp(-2.5)exp(\beta_1)$

3. 藉由$h\left(t\middle|m^*(t)\right)$求得不同標誌狀態下之存活函數$S\left(t\middle|m^*(t)\right)$

(1). 當t時間點下未發生標誌事件，存活函數$S(t|m^*(t) = 0)$可表示為

$$S\left(t\middle|m^*(t)=0\right) = exp\left(-\int_0^t h_0(s)ds\right) = exp\left(-exp(-2.5)t\right) \qquad \textbf{(4-2)}$$

(2). 當t時間點下已發生標誌事件，存活函數$S(t|m^*(t) = 1)$可表示為

doi:10.6342/NTU201903431

$$
\begin{aligned}
S\left(t\middle|m^*(t)=1\right) &= exp\left[-\int_0^t h_0(s)exp\left(\beta_1\cdot m^*(t)\right)ds\right] \\
&= exp\left[-\left(\int_0^{T_M} h_0(s)ds+\int_{T_M}^t h_0(s)exp\left(\beta_1\cdot m^*(s)\right)ds\right)\right] \\
&= exp\left[-\left(exp(-2.5)\left(\int_0^{T_M} ds+\int_{T_M}^t exp(\beta_1)ds\right)\right)\right] \\
&= exp\left[-\left(exp(-2.5)\left(T_M+exp(\beta_1)t-exp(\beta_1)T_M\right)\right)\right]
\end{aligned}
$$

4. 欲求存活時間T_D，定義 $F\left(t\middle|m^*(t)\right)=P\left(T\le t\middle|m^*(t)\right)=1-S\left(t\middle|m^*(t)\right)$，則

$$
F\left(t\middle|m^*(t)\right)=\begin{cases} 1-exp\left(-exp(-2.5)t\right), m^*(t)=0 \\ 1-exp\left[-\left(exp(-2.5)\left(T_M+exp(\beta_1)t-exp(\beta_1)T_M\right)\right)\right], m^*(t)=1 \end{cases}
$$

而 $u_{T_D}=F\left(t\middle|m^*(t)\right)\sim U(0,1)$

5. 求得存活時間t

$$
T_D=\begin{cases} \dfrac{-log\left(u_{T_D}\right)}{exp(-2.5)}, \text{若 } -log\left(u_{T_D}\right)<exp(-2.5)\cdot T_M \\ \dfrac{-log\left(u_{T_D}\right)-exp(-2.5)T_M+exp(-2.5)exp(\beta_1)T_M}{exp(-2.5)exp(\beta_1)}, \text{若其他} \end{cases}
$$

其中$u_{T_D}\sim U(0,1)$

6. 假設設限時間C產生方式為，先以均勻分布產生變數u_C，$u_C\sim U(0,1)$，當$\beta_1=1.5$ 時，$\mathrm{C}=-ln(u_C)/0.02$，當$\beta_1=2.8$ 時，$\mathrm{C}=-ln(u_C)/0.035$。

7. 當$\beta_1=1.5$時，資料中T_M之 25 百分位數約為 5.9，50 百分位數約為 14.4，平均值約為 20，而T_D之 25 百分位數約為 2.3，50 百分位數約為 5.3，平均值約為 8.8。當$\beta_2=2.8$時，資料中T_M之 25 百分位數約為 5.9，50 百分位數約為 14.4，平均值約為 20，而T_D之 25 百分位數約為 2.7，50 百分位數約為 6.1，平均值約為 8.5。標誌事件發生比例，即$P(T_M<T_D, T_D<C)$比例設定為 33%。

doi:10.6342/NTU201903431

- **以對數常態分佈(Lognormal)生成不同標誌時間及事件時間之間相關性資料**

1. 以雙變數常態分佈生成T_M以及T_D，$(T_M, T_D) \sim MVN(\begin{bmatrix}1.3\\0.9\end{bmatrix}, D)$，其中$D$為共變異數矩陣，$D = \begin{bmatrix}\sigma_{T_M}^2 & \rho\sigma_{T_M}\sigma_{T_D}\\ \rho\sigma_{T_M}\sigma_{T_D} & \sigma_{T_D}^2\end{bmatrix}$。
2. 生成T_M以及T_D後，取指數得到最後T_M以及T_D數值。其中在 Kendall's tau 相關係數為 0.6 之下，T_M之 25 百分位數約為 1.4，50 百分位數約為 3.6，平均值約為 9.7，而T_D之 25 百分位數約為 0.9，50 百分位數約為 2.4，平均值約為 6.7。在 Kendall's tau 相關係數為 0.8 之下，T_M之 25 百分位數約為 1.5，50 百分位數約為 4，平均值約為 9.7，而T_D之 25 百分位數約為 1.1，50 百分位數約為 2.6，平均值約為 6.7。標誌事件發生比例，即$P(T_M < T_D, T_D < C)$皆設定為 20%。設定T_M以及T_D在相關性 0.6 以及 0.8，T_M變異數為 1.95，T_D變異數為 2，共變異數根據不同相關性分別設定為 1.75 以及 1.88。
3. 假設設限時間C產生方式為，先以均勻分布產生變數u_C，$u_C \sim U(0,1)$，則$C = -ln(u_C)/0.05$，設限率為 20%。

- **以 Clayton Copula 方式生成不同標誌時間及事件時間之間相關性資料**

1. 產生兩獨立且服從均勻分布之變數u_1, v，利用條件機率$P(U_2 \le u_2 | U_1 = u_1)$求出$u_2$。
2. 本文給定 Kendall's tau(τ)為 0.6 以及 0.8，經由$\frac{\theta}{2+\theta} = \tau$求出參數$\theta = 3$及 8。
3. 產生$u_1, v \sim U(0,1)$。
4. 根據u_1, v產生$u_2 = \left[u_1^{-\theta} \cdot \left(v^{-\frac{\theta}{1+\theta}} - 1\right) + 1\right]^{-1/\theta}$。
5. 得到u_1, u_2後，分別以$\lambda_{u_1} = 0.1$以及$\lambda_{u_2} = 0.13$之指數分配做轉換，代表T_M以及T_D。在相關性 0.6 之下，T_M之 25 百分位數約為 3.3，50 百分位數約為

doi:10.6342/NTU201903431

7.2，平均值約為 10.1，而T_D之 25 百分位數約為 2.6，50 百分位數約為 5.9，平均值約為 8.2，$T_M < T_D$比例設定為 20%。。在相關性 0.8 之下，T_M之 25 百分位數約為 2.8，50 百分位數約為 7.2，平均值約為 9.8，而T_D之 25 百分位數約為 2.2，50 百分位數約為 5.5，平均值約為 7.7，標誌事件發生比例，即$P(T_M < T_D, T_D < C)$設定為 15%。

6. 假設設限時間C產生方式為，先以均勻分布產生變數u_C，$u_C \sim U(0,1)$，則$C = -ln(u_C)/0.04$，設限率為 21%。

第二節 模擬結果

本節將探討以三種不同相關性模式生成資料之預測表現，以及介紹模式調整的過程。本研究藉由給定不同相關係數於這些相關性模式下，希望以 LM3 模式進行預測時，能夠找出對於$g(T_M)$在調整上共同的方式，以利往後在做實際資料分析時，能有模式調整上依循的方向。除此之外，在根據上述三種不同資料生成方式之預測結果發現，在部份情況下，若要針對已發生標誌事件個案存活曲線做調整，對於未發生標誌事件個案未來存活機率預測與真實存活情況間不一致性增加的狀況出現。但我們仍希望可以盡可能同時兼顧這兩種情況在預測上面的表現。因此本文後續將 LM2 以及 LM3 模式以是否發生標誌事件做為分層依據，建立 LSM1 以及 LSM2 模式進行預測。下方皆呈現在特徵時間點 4 年之下，五個模式在預測未來 10 年內存活機率之存活曲線圖以及在預測未來 5 年存活機率之 AUC、Brier 分數以及均方差表現，其餘結果則呈現於附錄當中。

- **時間依賴性 Cox 模式**

產生 1000 組樣本數分別為 1000 等右設限資料，而真值計算為 5000 次樣本數為 5000 所組成的無設限資料所產生。1000 組樣本數為 1000 的右設限資料標誌事件之係數平均值為 1.591(變異數為 0.003)。

doi:10.6342/NTU201903431

下方呈現$\beta_1 = 1.5$時預測結果，在u時間點前是否發生標誌事件的存活機率真值及三組右設限資料在特徵時間點為 4 年下，五種模式預測未來存活機率表現，在 LM1 部分對於已發生標誌事件個案，其預測存活機率有略為高估情形，因此在 LM2 加入標誌時間進行預測，發現高估情形改善沒有特別明顯，後續以 LM3、LSM1、LSM2 進行預測，並以$log(T_M)^x$型式將存活機率高估部分做調整。從圖 4-1 結果看到，由於本文特徵點模式是由 Cox 模式為基礎所建構，本身與時間依賴性 Cox 模式結構相近，因此預測之存活曲線皆與真值接近，故本文在後續多考慮了另外兩種方式生成資料的方式。但在各項評估準則方面(圖 4-2 至圖 4-5)，觀察到由於 LSM1 以及 LSM2 將資料做分層估計，因此所使用到之資料量較少，預測表現也較其他三個特徵點模式不穩健。

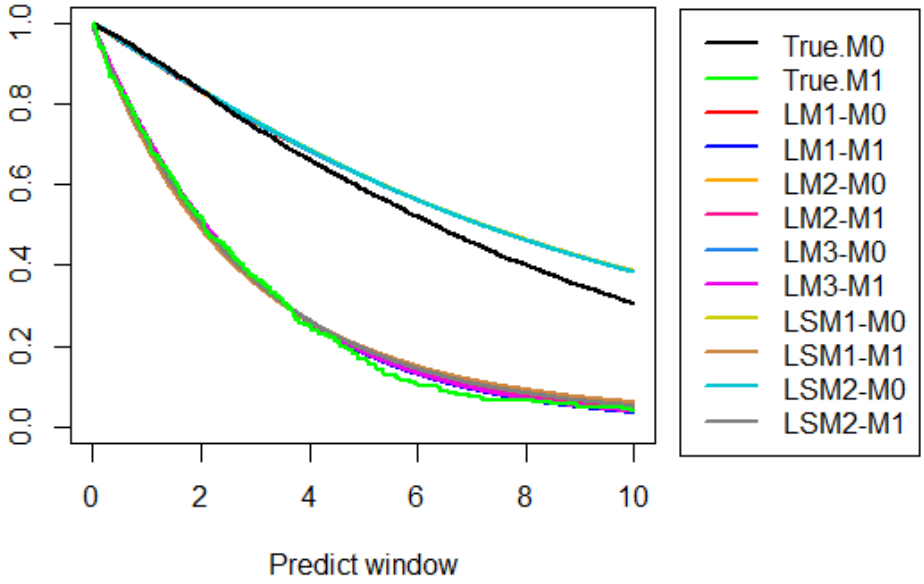


圖 4-1：特徵時間點$u = 4$，五種特徵點模式的預測存活曲線圖(時間依賴性 Cox 模式)

(e.g. LM1-M0 表示在特徵時間點 4 年前，未發生標誌事件個案未來 10 年存活曲線圖。True.M0 表示在不考慮設限情況下，特徵時間點 4 年前未發生標誌事件未來 10 年存活機率真值之曲線圖，條件存活機率表示為$\pi(u+w,u) = P\left(T_D > u+w \middle| u, M^*(u)\right)$，其中 w 表示預測區間(predict window))

doi:10.6342/NTU201903431

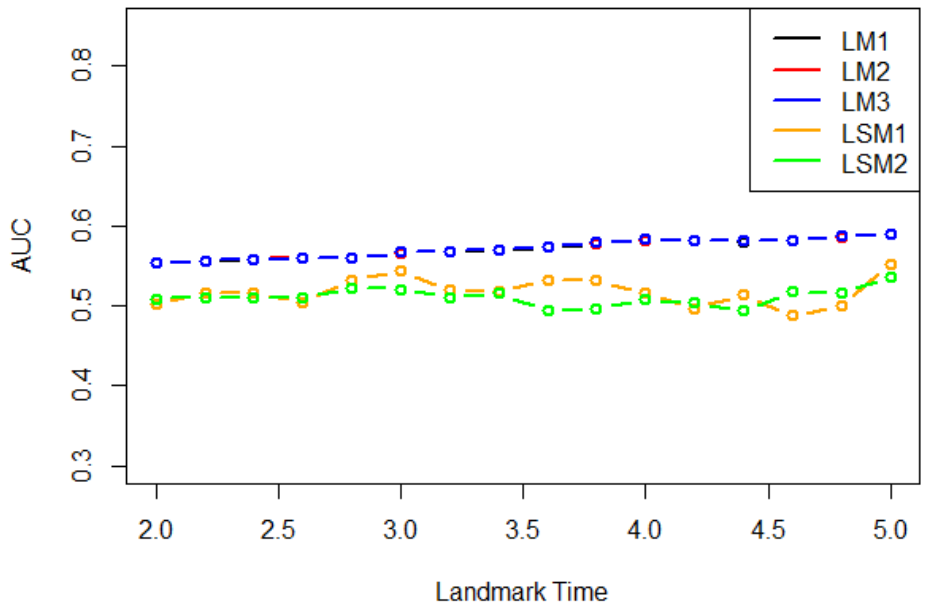


圖 4-2：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現 (時間依賴性 Cox 模式)

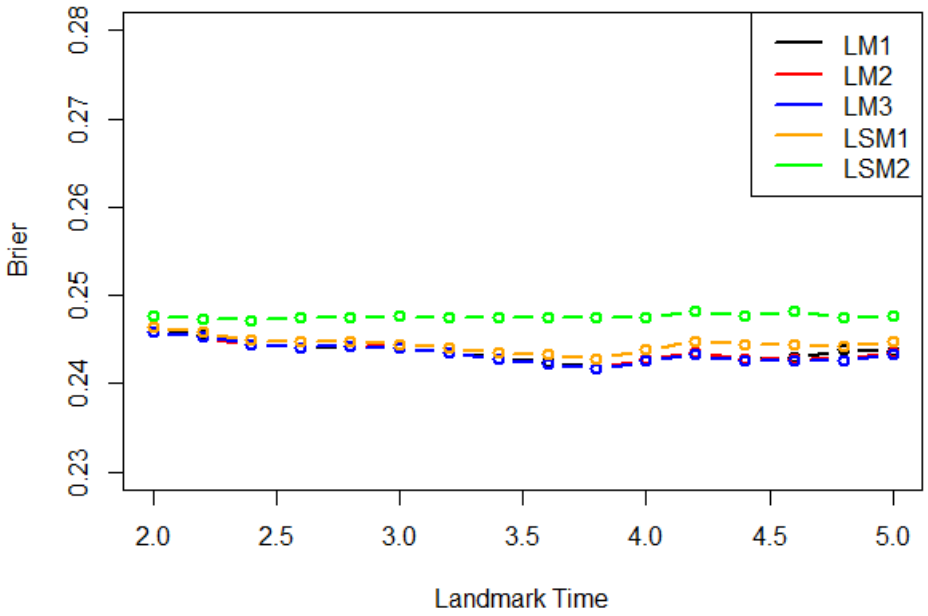


圖 4-3：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數 (時間依賴性 Cox 模式)

doi:10.6342/NTU201903431

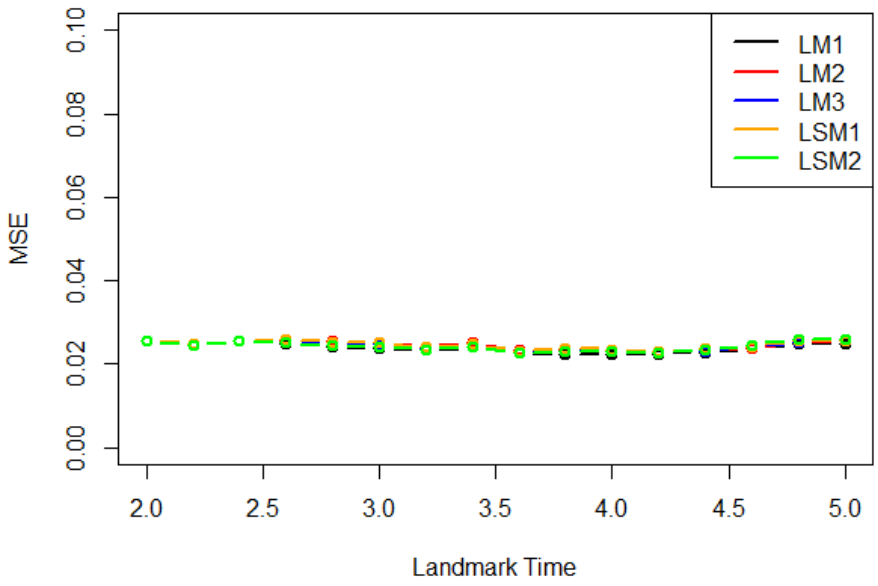


圖 4-4：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(時間依賴性 Cox 模式)

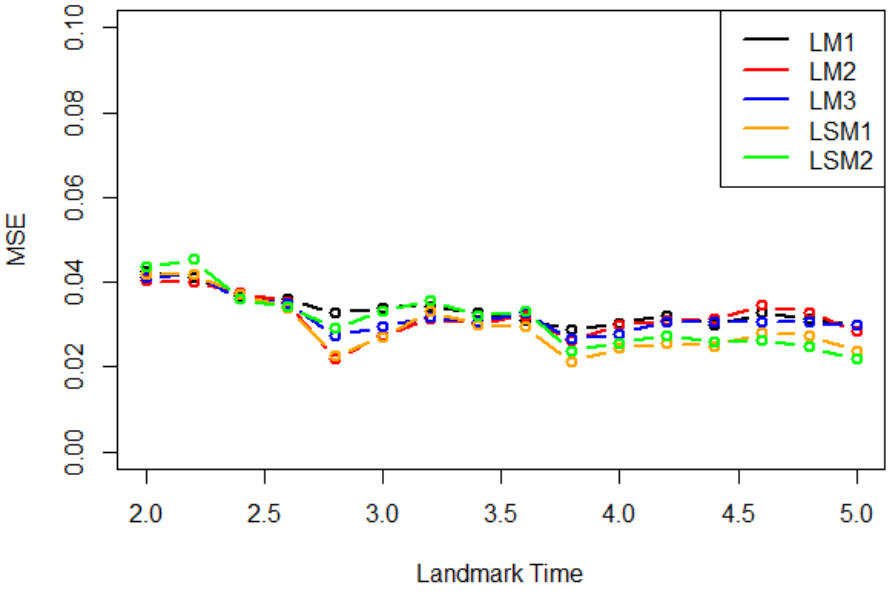


圖 4-5：在特徵時間點u為 2 至 5 年下，已發生標誌事件個案五種模式預測 5 年後存活機率均方差(時間依賴性 Cox 模式)

doi:10.6342/NTU201903431

- **對數常態分佈(Lognormal)**

產生 1000 組樣本數為 1000 的右設限資料，真值計算為 5000 次樣本數為 5000 所組成的無設限資料所產生。探討在樣本數 1000、相關係數 Kendall's tau = 0.6、0.8 下，在特徵時間點為 2、3、4、5 年時，以三種模式進行預測以及評估準則表現結果，下方呈現當 Kendall's tau = 0.6 下，特徵時間點為 4 年時，三種模式預測未來存活機率的表現，在 LM1 部分對於未發生標誌事件個案在預測表現上有高估情況出現，而對於已發生標誌事件個案預測之存活曲線與真值接近。接著加入標誌時間(LM2)進行預測後，對於兩年內存活機率反而出現些微低估而三年後存活機率高估的情況，故判斷在不同標誌時間區間下，對於未來存活機率影響是不同的，最後考量 LM3、LSM1、LSM2 作為預測模式，調整方式為：

(1). 在特徵時間點u在 2 至 4 年情況下，以每一年作為分段單位，將標誌時間作分段。當標誌時間位於 1 至 2 年區間時，以$(T_M)^x$函數作為轉換型式，調整低估部分，低估情形越嚴重，給予次方越高，對於高估部分則以$log(T_M)^x$型式調整，高估情形越嚴重，所給予的次方也越高。

(2). 而針對特徵時間點u在 5 年的情況下，由於使用特徵點模式作為預測模式會有損失資料的狀況，由於早期發生標誌事件之個案多無法活超過 5 年，因此若採用第(1)點分段方式，在標誌時間位於 1 至 2 年區間之估計將會不穩定，故分段方式為將標誌時間在 1 至 4 年間視為一個區間，4 至 5 年為另一個區間，以此方式建構預測模式。

根據圖 4-6 結果發現，將T_M時間作分段處理後，並沒有如想像中能夠將高估部分做調整，反而有高估更多情況，但若觀察附錄中特徵時間點$u = 2$、3的存活曲線圖，發現 LM2 的預測則沒有比其他模式佳，因此在這個情況下，可以推論對於較早期發生標誌事件的個案往後存活機率的預測可以考慮使用 LM2 以外的模式進行，而較晚發生的個案則可以選擇使用 LM2 以及 LSM1 預測。

doi:10.6342/NTU201903431

若從評估準則觀察(圖 4-7 到圖 4-10)，發現雖然在存活曲線部分模式 LM2 以及 LSM1 與真值最接近，但從圖 4-10 可以明顯看出，兩模式均方差較高，又均方差為偏差(bias)以及變異(variance)組成，故可以看出這兩個模式偏差雖然降低但付出的代價為模式較不穩健。

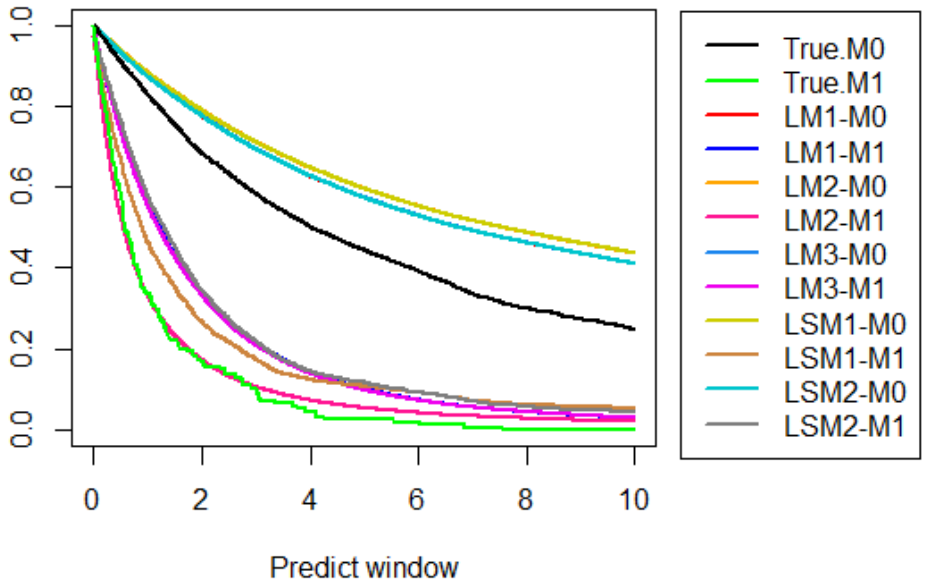


圖 4-6：在特徵時間點u =4，五種特徵點模式的預測存活曲線圖(對數常態)(e.g. LM1-M0 表示在特徵時間點 4 年前，未發生標誌事件個案未來 10 年存活曲線圖。True.M0 表示在不考慮設限情況下，特徵時間點 4 年前未發生標誌事件未來 10 年存活機率真值之曲線圖，條件存活機率表示為

$\pi(u+w,u)=P\left(T_D>u+w\middle|u,M^*(u)\right)$**，其中 w 表示預測區間(predict window))**

doi:10.6342/NTU201903431

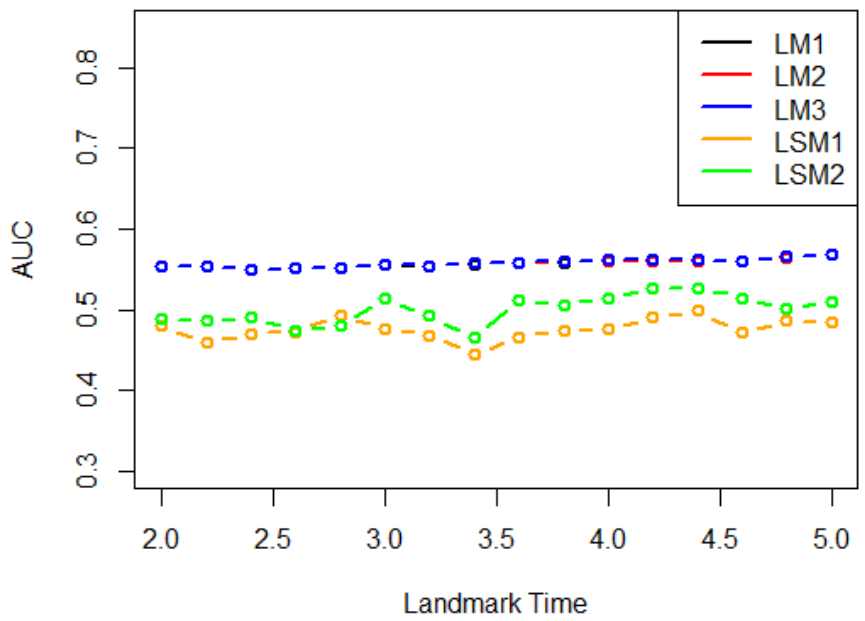


圖 4-7：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現 (對數常態)

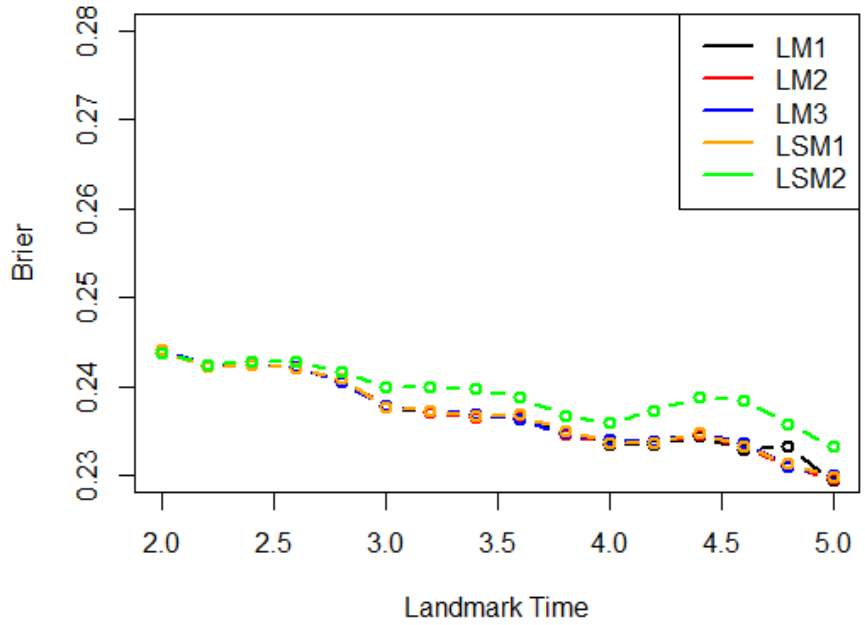


圖 4-8：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數 (對數常態)

doi:10.6342/NTU201903431

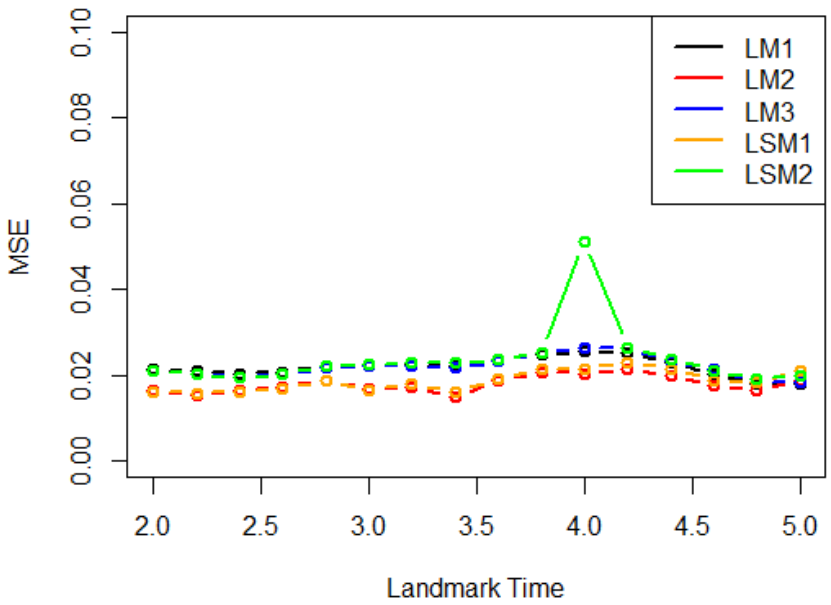


圖 4-9：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(對數常態)

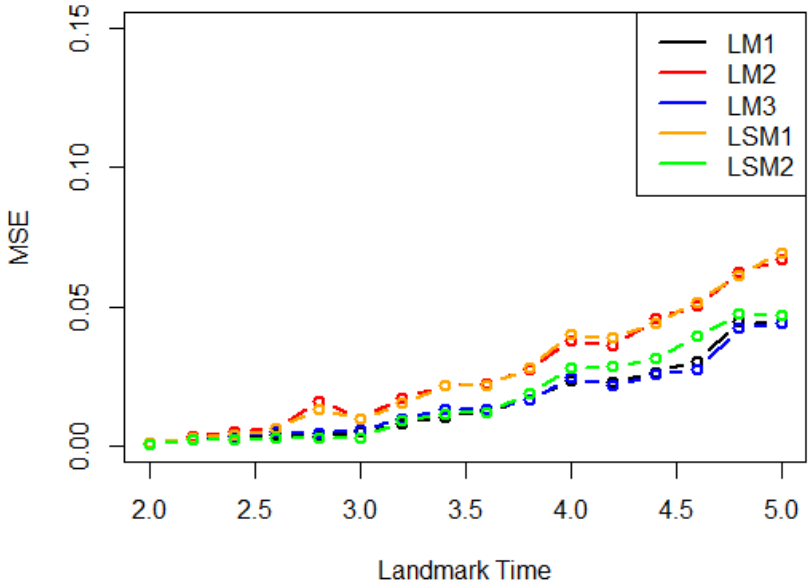


圖 4-10：在特徵時間點u為 2 至 5 年下，已發生標誌事件個案五種模式預測 5 年後存活機率均方差(對數常態)

doi:10.6342/NTU201903431

- **Clayton Copula**

產生 1000 組樣本數為 1000 的右設限資料，設限率約為 23%，真值計算為 5000 次樣本數 5000 所組成的無設限資料。考慮在 Kendall's tau 為 0.6 以及 0.8 下，模型預測表現。下方呈現在 Kendall's tau 為 0.6，特徵時間點為 4 年下，以三種模式預測未來存活情形的存活曲線圖，在 LM1 部分對於特徵時間點前已復發的個案而言，其預測的存活機率皆有高估的現象，而加入標誌時間以 LM2 進行預測，預測結果仍然高估，因此以 LM3、LSM1、LSM2 模式進行配適，此處將T_{Mu}轉換為$log(T_M)^x$，發現藉由調整次方項參數x的大小，確實可以改善低估的現象，且調整方式只需更動最後一段標誌時間區間x的大小，若低估越嚴重，x就給與越大的值，因此依循此規律，則有辦法調整預測的存活機率，不過在T_M與T_D相關性不高的情況下，在給定次方項至一定大小時即無法繼續調整。結果呈現如下，從圖 4-11 看出 LM2 以及 LSM1 預測結果較接近真值，但在評估準則方面(圖 4-12 至圖 4-15)，從圖 4-15 可以明顯看出，兩模式均方差較高，又均方差為偏差(bias)以及變異(variance)組成，故可以看出這兩個模式偏差雖然降低但付出的代價為模式較不穩健。

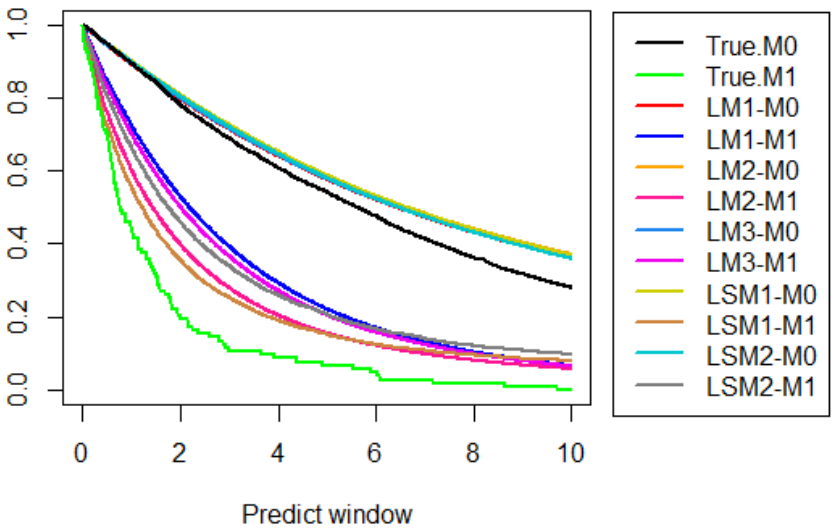


圖 4-11：在特徵時間點$u = 4$，五種特徵點模式的預測存活曲線圖(Clayton Copula)

doi:10.6342/NTU201903431

(e.g. LM1-M0 表示在特徵時間點 4 年前，未發生標誌事件個案未來 10 年存活曲線圖。True.M0 表示在不考慮設限情況下，特徵時間點 4 年前未發生標誌事件未來 10 年存活機率真值之曲線圖，條件存活機率表示為 $\pi(u+w,u)=P\left(T_D>u+w\middle|u,M^*(u)\right)$，其中 w 表示預測區間(predict window))

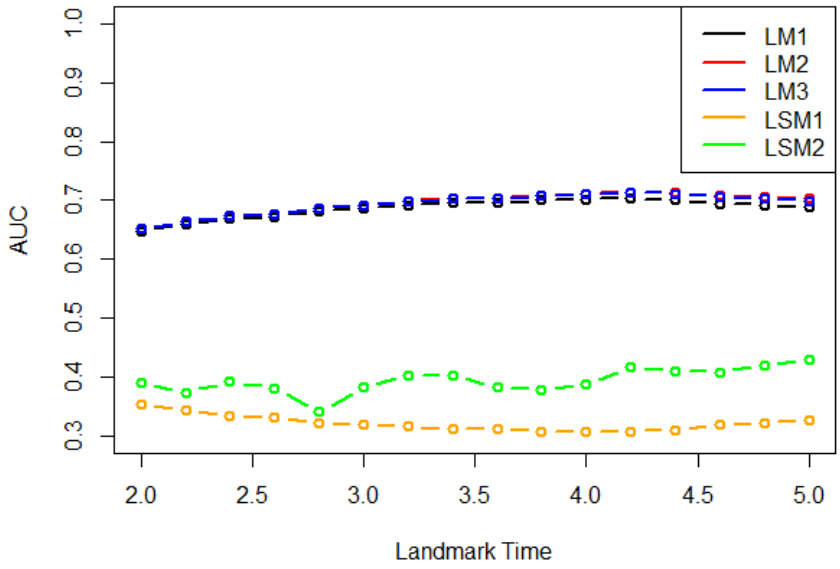


圖 4-12：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 AUC 表現(Clayton Copula)

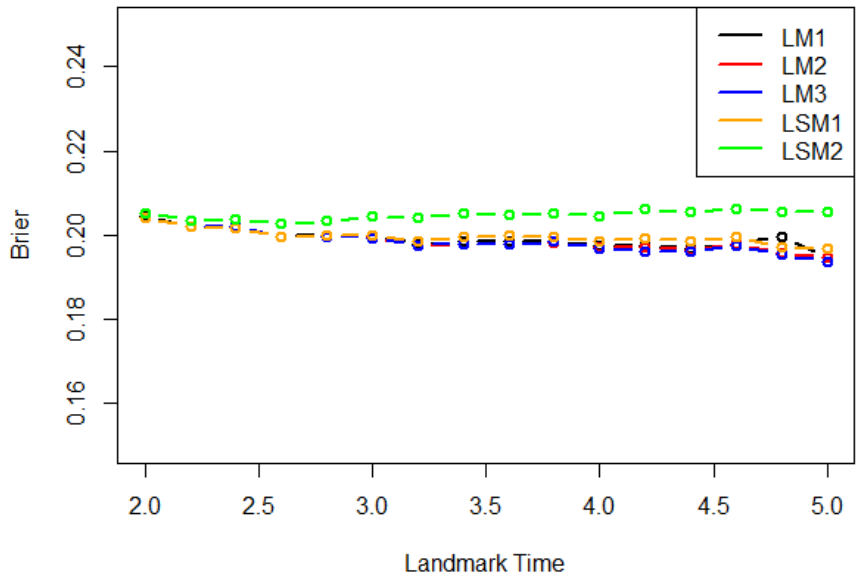


圖 4-13：特徵時間點u為 2 至 5 年下，五種模式預測 5 年後存活機率 Brier 分數(Clayton Copula)

doi:10.6342/NTU201903431

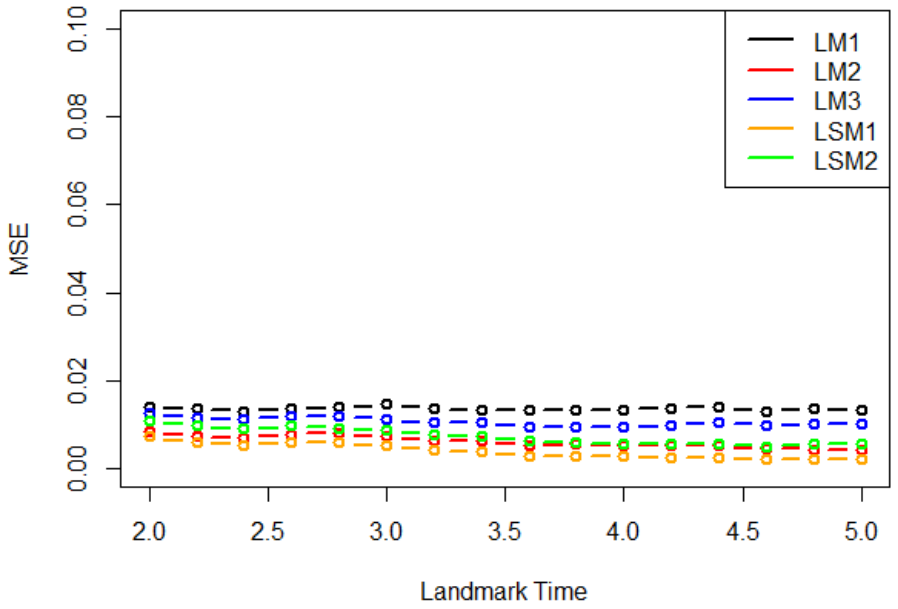


圖 4-14：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(Clayton Copula)

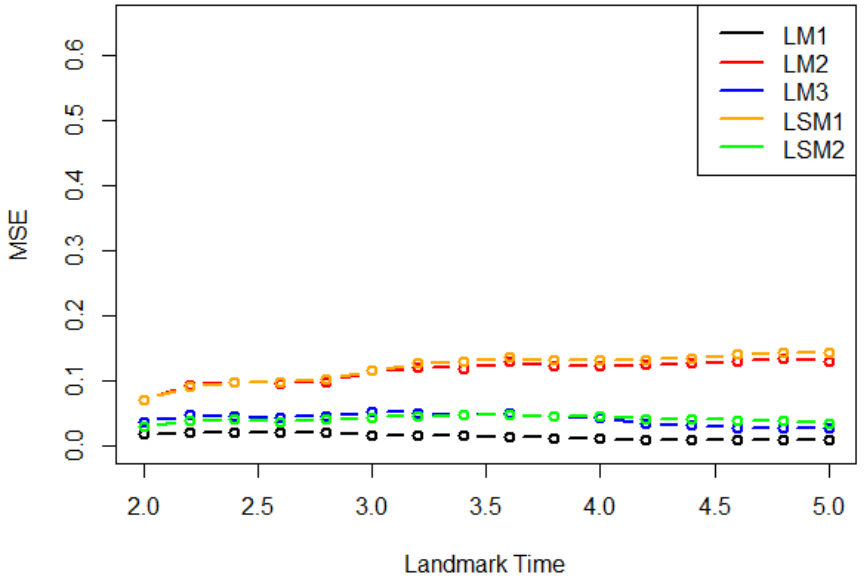


圖 4-15：在特徵時間點u為 2 至 5 年下，未發生標誌事件個案五種模式預測 5 年後存活機率均方差(Clayton Copula)

doi:10.6342/NTU201903431

第五章 實例分析

第一節 大腸直腸癌研究

大腸直腸癌在目前社會中是各項癌症中常見的一種，不論男性或女性皆有可能受到影響，雖然多數病患在得病初期接受手術治療後即有可能好轉，但仍有大部分病患最終死於此疾病。為了確保進入研究之病患皆符合資格，病患將由 NCCTG (North Central Cancer Treatment Group)病理學委員會進行資格審查，在此研究中，進入研究之病患必須符合幾項條件：

1. 參與研究的病患都必須接受過位於大腸或直腸上之腫瘤切除手術。
2. 切除之腫瘤須已侵入鄰近器官或轉移至附近淋巴結。
3. 病患在進入研究前未曾接受過任何位於腰椎或骨盆的放射治療及化療藥品(5-FU)。
4. 5 年內並未罹患其他惡性疾病，除鱗狀細胞癌、基底細胞癌以及子宮頸原位癌。
5. 病患未有其他局部轉移之腫瘤無法與原發性病灶一起切除。
6. 符合條件者需於手術結束後五週內進入研究。

而大腸直腸癌的預後多與臨床與病理因素息息相關，例如：年齡、腫瘤大小、組織類型分級…等，而另外一個重要的預後因子則是癌症復發的有無，此預後因子對於病患未來存活情形會有極大影響。因此若能提前預測出已發生復發病患未來存活機率，則能提供醫師在診斷上良好的參考依據，並對病患做出最適當的治療且也可以對於醫療資源進行更妥善的分配。此筆資料可在 R 當中 **survidm (Luis. et al., 2019)**套件中獲得。

doi:10.6342/NTU201903431

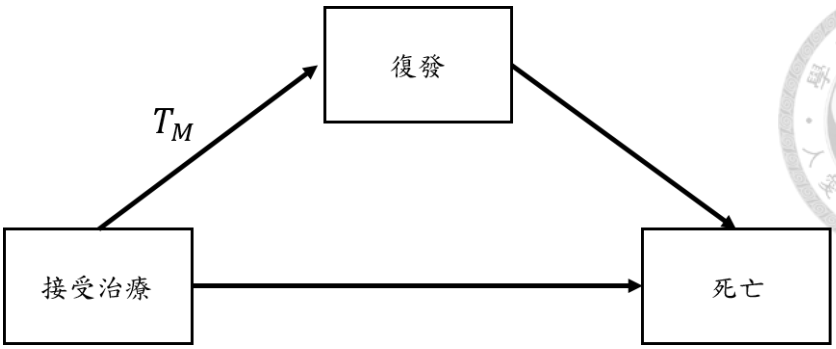


圖 5-1：大腸直腸癌資料結構圖

本文所使用到變數介紹如下表：

變數名稱	定義
time1	進入研究後至發生復發/死亡/設限的時間
event1	發生復發或設限指標函數(0：未發生復發，1：復發)
Stime	進入研究後發生復發/時間
event	死亡或設限指標函數(0：設限，1：死亡)

表 5-1：大腸直腸癌資料變數介紹

此筆資料共紀錄了 929 位病患的資料，其中 506 位發生復發，而 506 位中則有 452 人在研究終止前死亡。對於發生復發的個案，復發時間T_M之 25 百分位數約為 0.6 年、中位數約為 1.13 年、75 百分位數約為 2.1 年，而平均數為 1.64 年。死亡時間T_D之 25 百分位數約為 2.21 年、中位數約為 5.41 年、75 百分位數約為 6.48 年，而平均數為 4.57 年。資料中復發時間與死亡時間之間 Kendall's tau 相關係數為 0.8。將其中 2/3 資料作為訓練資料集，其餘為測試資料集，並進行 3-fold 的交叉驗證。本文探討特徵時間點在 2、3、4、5 年下，以上述所建構之五種特徵點模式進行預測，並比較其預測之存活機率與實際存活機率之一致性，真值計算部分則以測試資料集繪製 Kaplan-Meier 存活曲線圖做為代表。下圖為在各特徵時間點下預測未來存活機率之存活曲線圖以及 AUC、Brier 分數、均方差等評估模式的準則。

doi:10.6342/NTU201903431

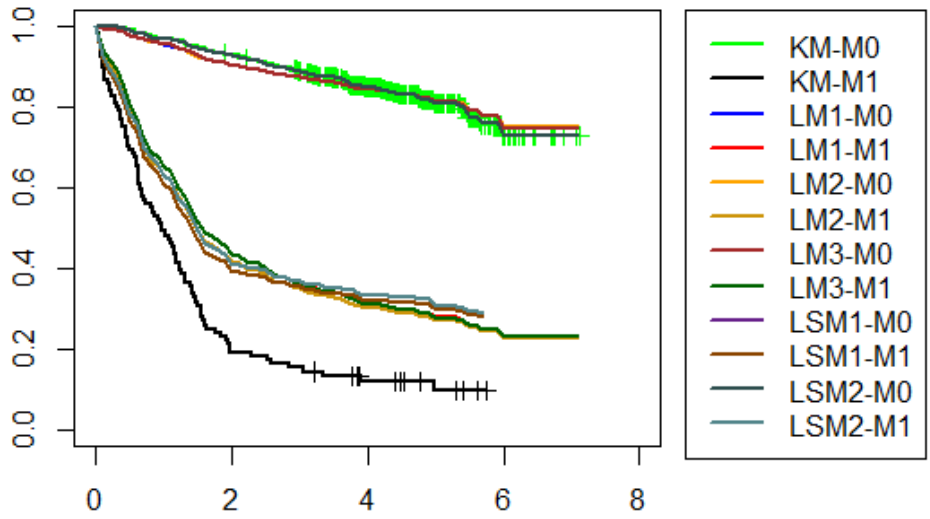


圖 5-2：特徵時間點$u = 2$年下，以五種特徵點模式預測個案未來 7 年存活機率的存活曲線圖

(e.g. KM-M0 表示在特徵時間點 2 年內尚未發生標誌事件個案之 Kaplan-Meier 存活曲線，LM1-M0 表示在特徵時間點 2 年內尚未發生標誌事件個案以表 3-1 中 LM1 模式進行預測的存活曲線圖。)

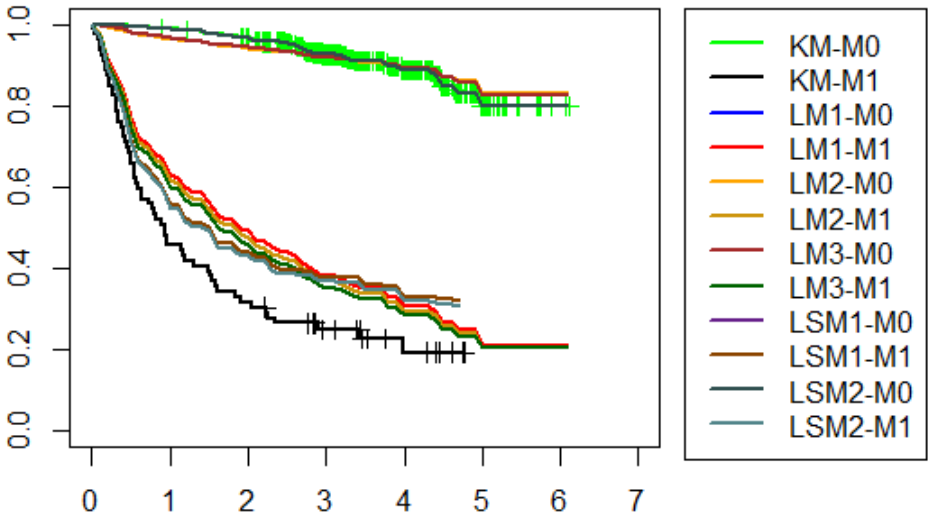


圖 5-3：特徵時間點$u = 3$年下，以五種特徵點模式預測個案未來 6 年存活機率的存活曲線圖

doi:10.6342/NTU201903431

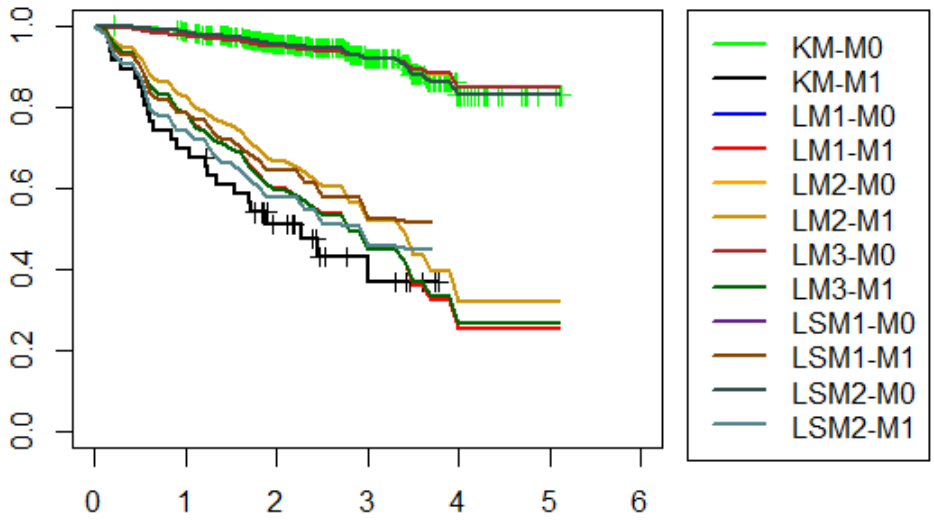


圖 5-4：特徵時間點$u = 4$年下，以五種特徵點模式預測個案未來 5 年存活機率的存活曲線圖

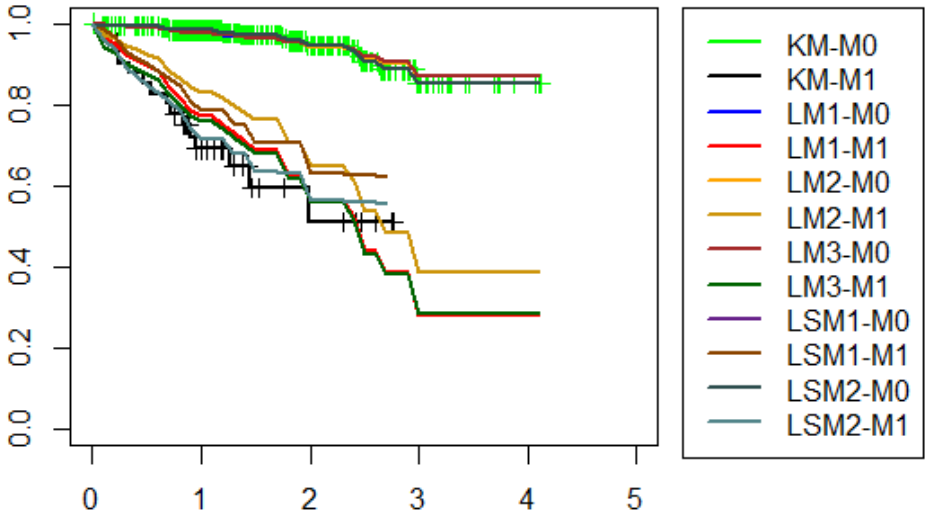


圖 5-5：特徵時間點$u = 5$年下，以五種特徵點模式預測個案未來 4 年存活機率的存活曲線圖

doi:10.6342/NTU201903431

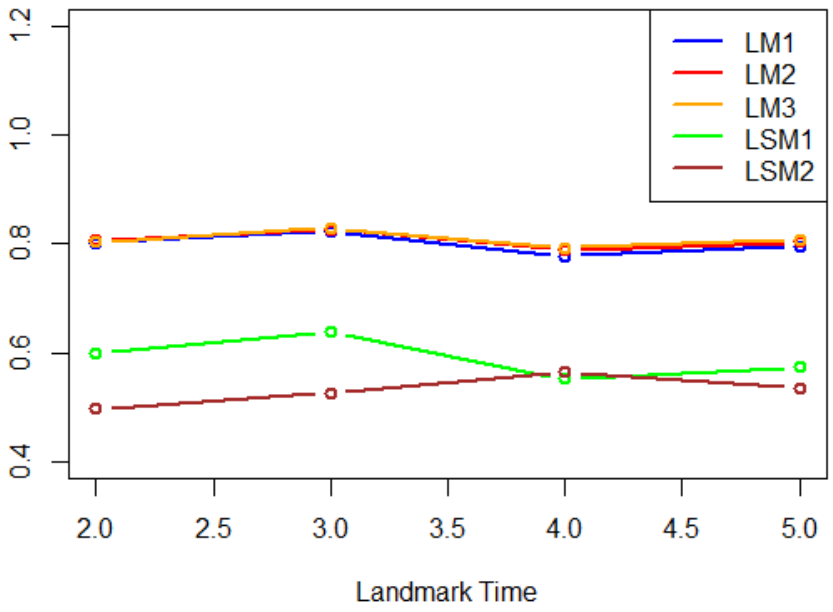


圖 5-6：特徵時間點$u = 2$、3、4、5下，以五種特徵點模式預測未來 2 年存活機率 AUC 表現

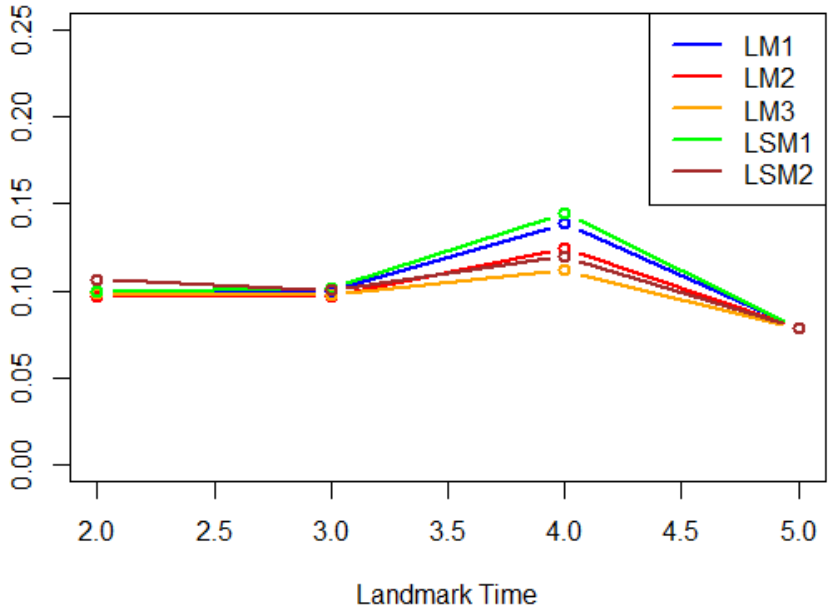


圖 5-7：特徵時間點$u = 2$、3、4、5下，以五種特徵點模式預測未來 2 年存活機率 Brier 分數表現

doi:10.6342/NTU201903431

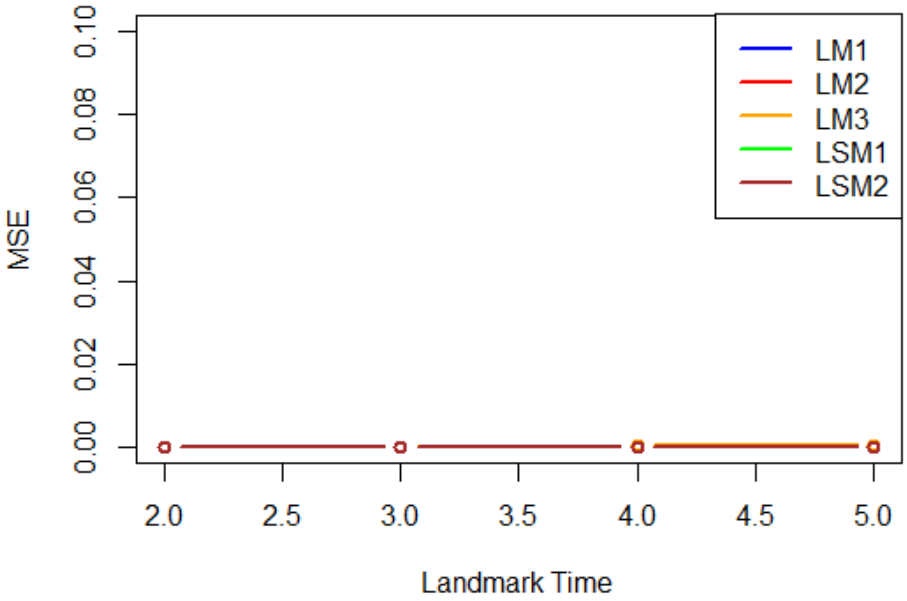


圖 5-8：特徵時間點$u = 2$、3、4、5下，未復發個案，五種特徵點模式預測未來 4 年存活機率均方差表現

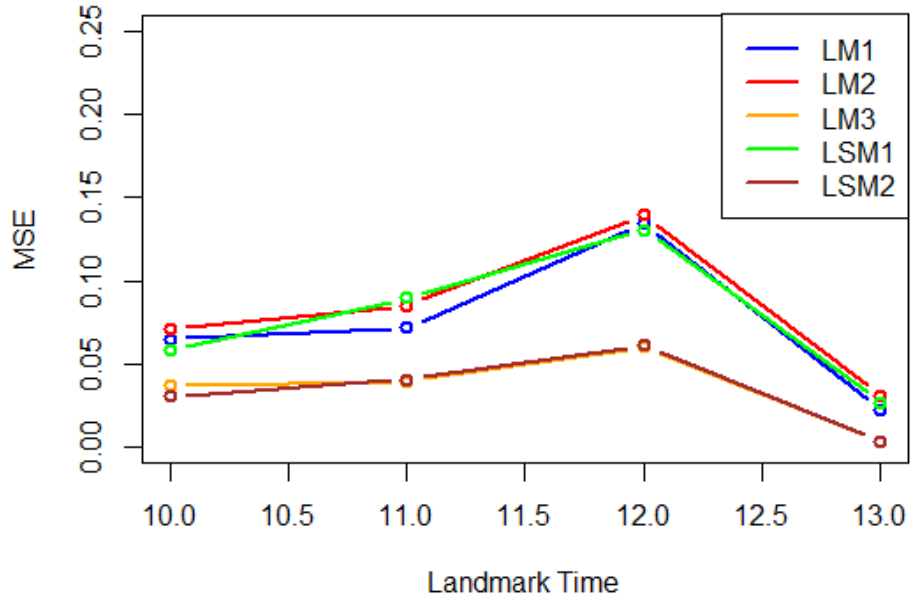


圖 5-9：特徵時間點$u = 2$、3、4、5下，已復發個案，五種特徵點模式預測未來 4 年存活機率均方差表現

doi:10.6342/NTU201903431

由圖 5-2 至圖 5-5 結果可以發現，LM3 以及 LSM2 在預測存活曲線圖上，由於透過分段估計以及函數調整，比起其餘模式預測存活機率更接近真值。在評估準則表現上，AUC(圖 5-6)表現上，LSM1 以及 LSM2 表現明顯較其餘特徵點模式差，可能原因為標誌事件在預測模式中為重要預測因子，若將他視為分層依據，則缺少此重要變數，因此 AUC 表現較差，Brier 分數(圖 5-7)表現，LM3 相較於其他模式表現更佳，若透過均方差(圖 5-8、圖 5-9)觀察 LSM2 以及 LM3 相較於其他模式表現明顯更好，因此認為此筆資料中以 LM3 以及 LSM2 兩模式能得到較準確的預測結果。

第二節 PAQUID 研究

阿茲海默症是一種發病進程緩慢且會隨著時間不斷惡化的神經退化性疾病，一般發病年齡在 65 歲以上，在已開發國家中，阿茲海默症是需要最多社會資源的疾病，一旦發病後平均存活時間大約是 4.5 年，因此我們希望經由模型的預測，能更準確的提供醫師患者在未來的情況，讓醫師做出更妥善的判斷以及提供最完善的治療。

本文以 PAQUID (Personnes Ag´ees QUID)研究中之資料進行實際資料分析，PAQUID 為一大型的流行病學研究，此研究主要針對法國西南部吉倫特(Gironde)以及多爾多涅(Dordogne)地區等 75 各社區，藉由心理學家進行家訪或電訪，在首次拜訪後，在接下來 1、3、5、8、10、13、15、17、20 年後皆會再度拜訪，總共 10 次拜訪在 20 年間。研究最後總共蒐納了 3,777 位居住於家中 65 歲以上的老年人口，對於阿茲海默症判斷標準上共分為兩階段，第一階段篩選使用美國精神醫學學會出版之精神疾病與診斷手冊第三版(DSM-IIIR)，第二階段則使用阿茲海默症臨床診斷標準(NINCDS-ADRDA)。此研究主要目的有二，一是希望藉由此研究了解阿茲海默症之盛行率、發生率等臨床表現，另

doi:10.6342/NTU201903431

一個則是透過機構相關研究以及死亡率來試圖找出最主要的危險因子。此筆資料收錄了這當中的 1,000 位民眾，可在 R 當中 **SmoothHazard** (Touraine et al., 2014)套件獲得。

這 1000 筆資料中，判定為阿茲海默症民眾共有 186 位，而研究結束前死亡的個案共有 724 位。當中，在進入研究前已被診斷出阿茲海默症個案有 23 位，研究結束前死亡 21 位，其餘 163 位則是在進入研究後才被診斷為阿茲海默症，研究結束前死亡 106 位。本研究以半競爭風險資料結構進行分析，資料結構如下：

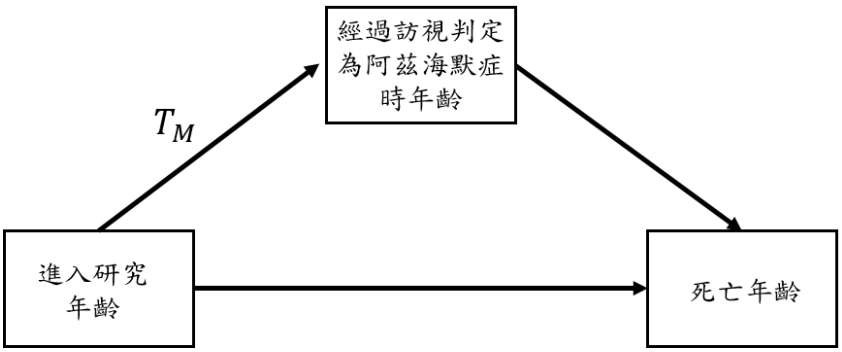


圖 5-10：PAQUID 研究資料路徑圖

本文所使用到變數介紹如下表：

變數名稱	定義
Dementia	至研究結束前是否被判定為阿茲海默症(0：否，1：是)
Death	至研究結束前是否已死亡(0：仍未死亡，1：死亡)
e	在進入研究時的年齡
l	對於進入研究前已曾被診斷為阿茲海默症個案，紀錄其接受該次診斷時年齡；未被診斷出阿茲海默症之個案則紀錄其進入研究前最後一次接受診斷時年齡。

doi:10.6342/NTU201903431

r	對於進入研究後才被診斷為阿茲海默症個案，紀錄接受訪視當時的年齡，進入研究前已被診斷出阿茲海默症個案則紀錄進入研究後第一次訪視年齡；未被判定為阿茲海默症個案則紀錄最後一次訪視的年齡資訊。
t	對於死亡個案：紀錄死亡時年齡，未死亡個案則紀錄最近一次獲得的年齡資訊。

表 5-2：PAQUID 資料變數介紹

這筆資料中， T_M在定義上可解釋為「個案進入研究年齡至第一次被檢測出為阿茲海默症年齡所經過的時間」定義為表 5-1 中變數 r-e，25 百分位數為 1.15 年、50 百分位數為 5.42 年、75 百分位數為 12.51 年，平均值約為 7.31 年；而T_D表示「個案進入研究年齡到紀錄死亡年齡所經過的時間」定義為表 5-1 中變數 t-e，25 百分位數為 6.17 年、50 百分位數為 11.45 年、75 百分位數為 17.01 年，平均值約為 10.98 年。T_M與T_D之 Kendall's tau 相關係數為 0.58。

本文將探討在特徵時間點 u 為 10、11、12、13 下，以五種特徵點模式預測未來存活機率的表現。在$u = 10$、11、12、13下，整筆資料中已被診斷為阿茲海默症個案分別有 89、106、106、140 筆。將其中 2/3 資料作為訓練資料集，其餘為測試資料集，並進行 3-fold 的交叉驗證 (cross validation)。而模式調整上，本文將T_M時間分成四段，分別為$(0,4]$、$(4,6]$、$(6,10]$以及$(10,u]$，並藉由$log(T_M)^x$做轉換，真值部分，將測試資料集繪製 Kaplan-Meier 存活曲線圖做為代表。下圖為各特徵時間點下，預測未來存活機率情形及預測的評估準則：

doi:10.6342/NTU201903431

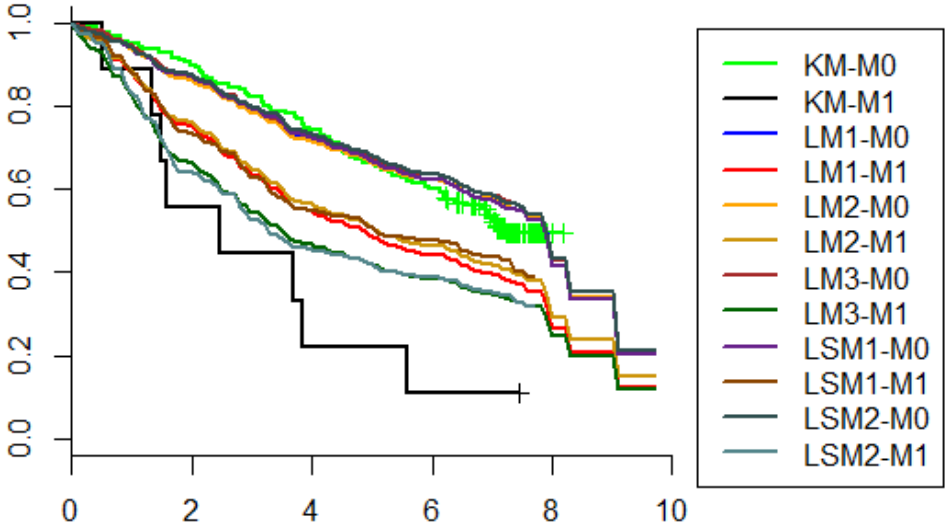


圖 5-11：特徵時間點$u = 10$年下，以五種特徵點模式預測個案未來 10 年存活機率的存活曲線圖

(e.g. KM-M0 表示在特徵時間點 10 年內尚未發生標誌事件個案之 Kaplan-Meier 存活曲線，LM1-M0 表示在特徵時間點 10 年內尚未發生標誌事件個案以表 3-1 中 LM1 模式進行預測的存活曲線圖。)

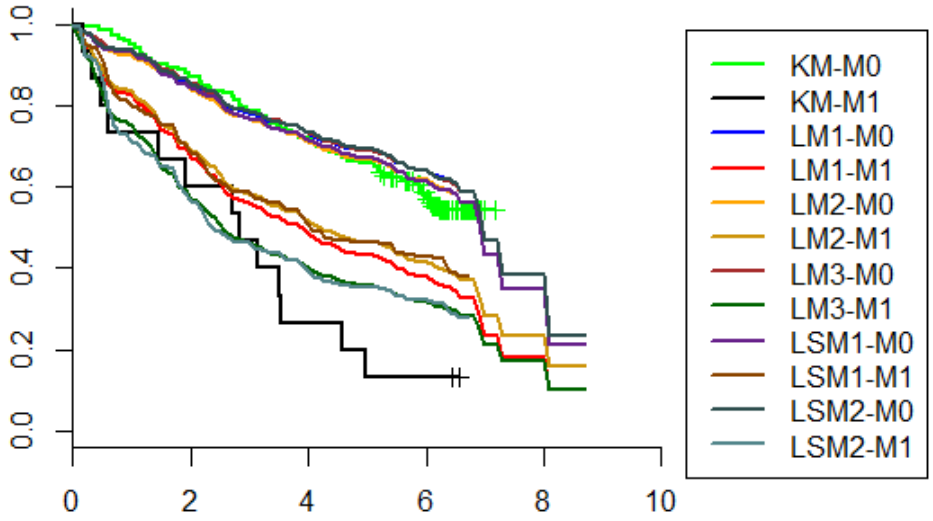


圖 5-12：特徵時間點$u = 11$年下，以五種特徵點模式預測個案未來 9 年存活機率的存活曲線圖

doi:10.6342/NTU201903431

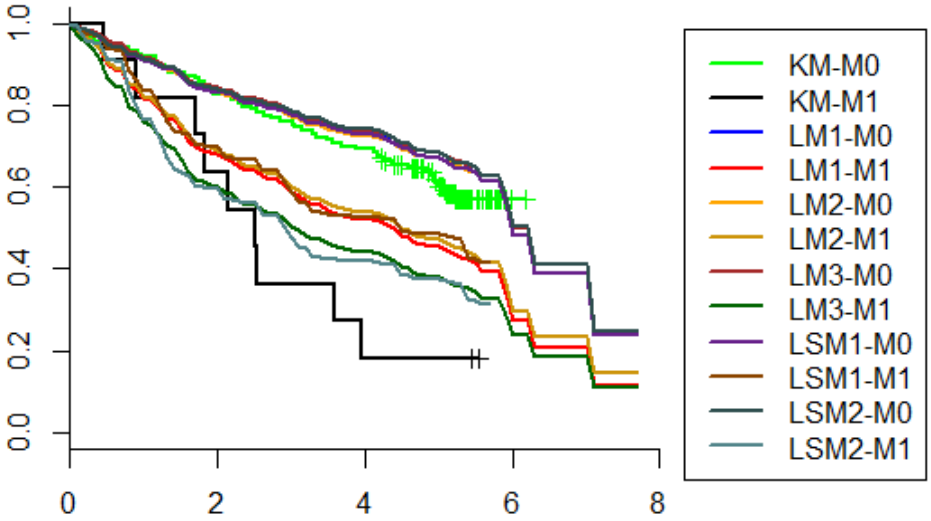


圖 5-13：特徵時間點$u = 12$年下，以五種特徵點模式預測個案未來 8 年存活機率的存活曲線圖

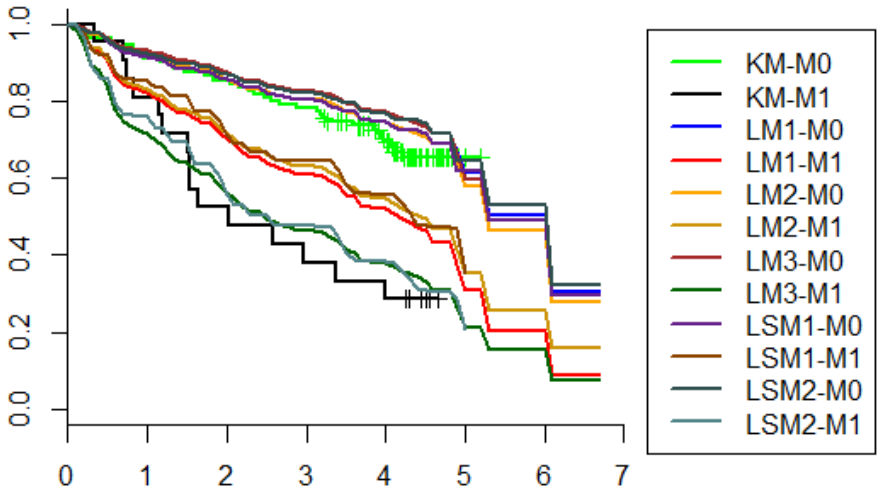


圖 5-14：特徵時間點$u = 13$年下，以五種特徵點模式預測個案未來 7 年存活機率的存活曲線圖

doi:10.6342/NTU201903431

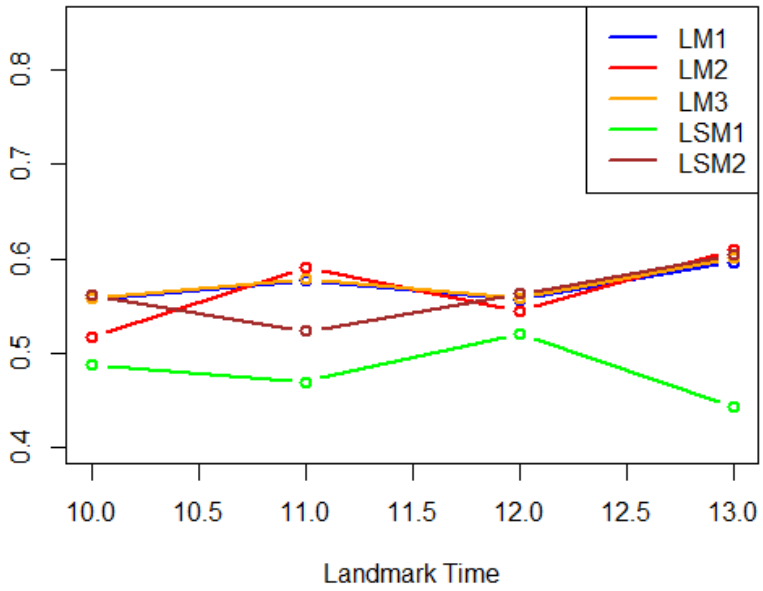


圖 5-15：特徵時間點$u = 10$、11、12、13下，以五種特徵點模式預測未來 4 年存活機率 AUC 表現

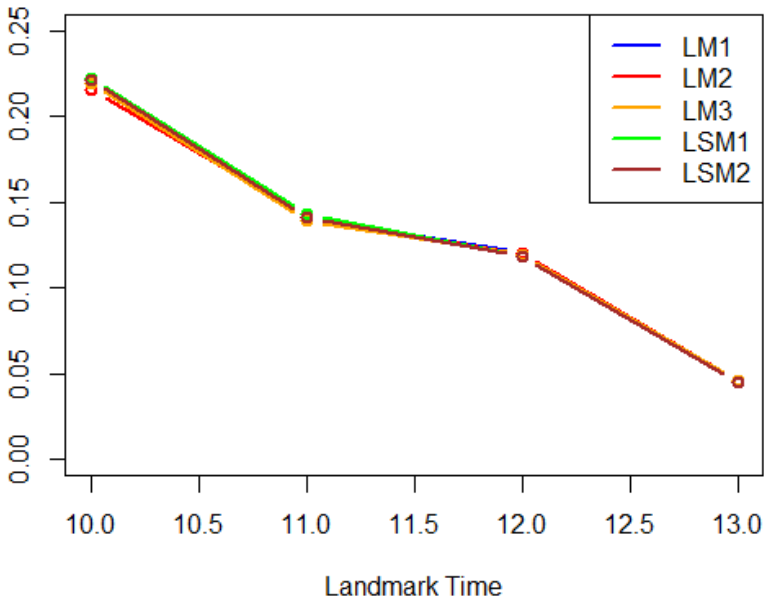


圖 5-16：特徵時間點$u = 10$、11、12、13下，以五種特徵點模式預測未來 4 年存活機率 Brier 分數表現

doi:10.6342/NTU201903431

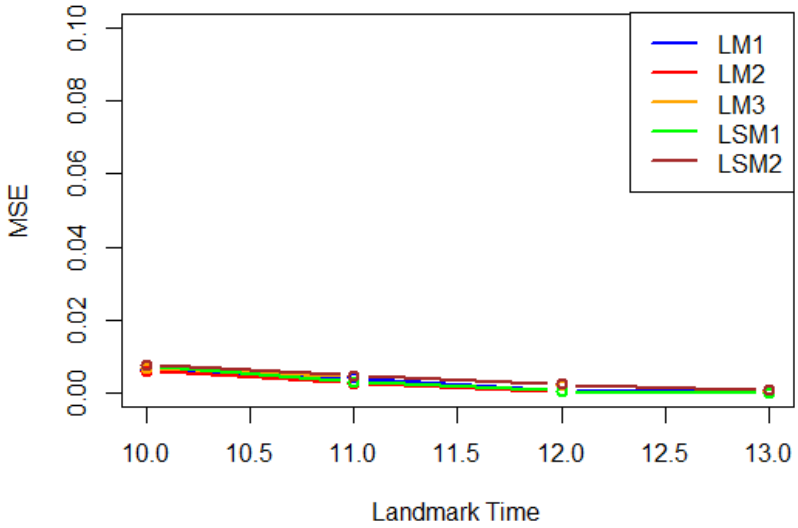


圖 5-17：特徵時間點$u = 10$、11、12、13下，未被判定為阿茲海默症個案，五種特徵點模式預測未來 4 年存活機率均方差表現

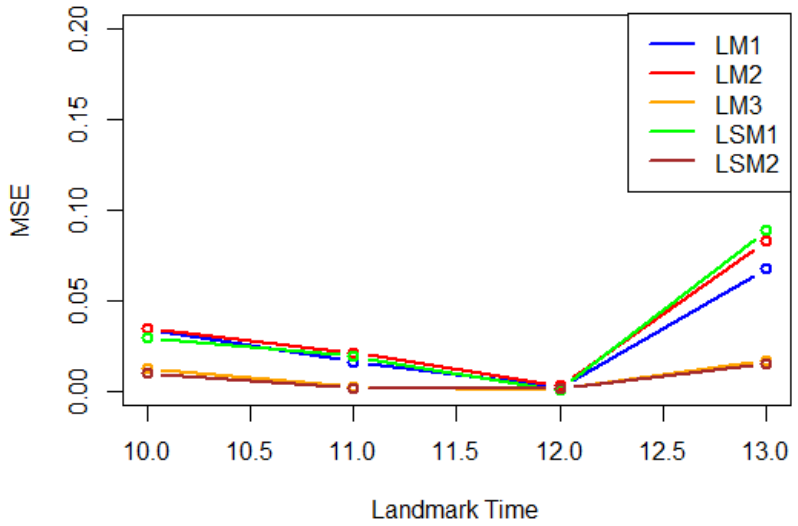


圖 5-18：特徵時間點$u = 10$、11、12、13下，被判定為阿茲海默症個案，五種特徵點模式預測未來 4 年存活機率均方差表現

由圖 5-11 至圖 5-14 結果可以發現，LM3 以及 LSM2 在預測存活曲線圖上，由於透過分段估計以及函數調整，比起其餘模式預測存活機率更接近真值。在評估準則表現上，AUC(圖 5-15)表現上，LSM1 以及 LSM2 表現較差，可能原因為，標誌事件在預測模式中為重要預測因子，若將他視為分層依據，模式中則缺少此重要變數，因此 AUC 表現相較其餘模式則較差，Brier 分數(圖 5-16)表現，各特徵模式則差異不大，若透過均方差(圖 5-17、圖 5-18)觀察

doi:10.6342/NTU201903431

LSM2 以及 LM3 相較於其他模式表現更佳，因此認為此筆資料中以 LM3 以及 LSM2 兩模式能得到較準確的預測結果。

上述初步分析模式仍有缺失之處：首先，分析此資料可先排除在進入研究前已被診斷出阿茲海默症個案有 23 位，再者，並未考慮年齡與其他重要預後因素等重要因素，最後，進入研究之年齡必須納入模式。

第六章 討論與總結

本文在 LM3 以及 LSM2 中進行分段估計，於實際資料分析上是直接以百分位數作為分段依據，並未特別探討分段點如何決定，對於$g(x)$函數決定則是直接以對數轉換，因此分段點的決定依據及$g(x)$函數如何做選擇可以做更進一步研究。評估準則部分，本文雖採用 AUC、Brier 分數以及均方差探討特徵點模式預測表現，但 AUC 以及 Brier 分數為判斷模式整體表現，若只想針對某特定情況個案(如：已發生標誌事件個案)觀察其預測表現則無法達成，因此文中透過均方差做判斷，但若只以均方差做為唯一判斷依據略顯不足，故是否能綜合多項評估準則後，提出新的評估準則或是在評估準則的選擇也可做更進一步研究。

本研究主要是在假設標誌歷程只與標誌事件的發生以及將標誌時間(T_M)長短有關下，以兩變數作為預測因子，並以 Cox 模式將存活時間與預測因子做連結預測未來存活機率。目的除了是以 AUC、Brier 分數以及均方差等評估準則比較不同特徵點模式在預測能力外，也將特徵點模式再作延伸，把標誌時間作分段估計的處理，設計出更具彈性的模式後再與其他特徵點模式進行比較。在模擬分析上，考慮(T_M,T_D)三種不同聯合模式，目的為在比較各特徵點模式後，希望能試著設計出較彈性的模式幫助往後在進行預測時，能夠藉著分段估計方

doi:10.6342/NTU201903431

式達到更佳的預測力外，希望找出對於前述$g(x)$函數調整的方式，能夠提供未來研究者在建立模式時的參考依據。在參數調整過程中，我們發現對於存活機率高估的狀況能夠將標誌時間以$log(T_M)^x$函數做轉換後，高估情況越明顯則給予參數x較高的次方，若預測情形為低估則以$(T_M)^x$函數作為轉換型式，同理低估情況越明顯則給予參數x越高的次方。除此之外，在調整過程中發現，若要針對已發生標誌事件個案做調整有時會降低對於未發生標誌事件個案預測的準確度，故本文後續多考慮了以是否標誌事件做為分層依據，建立分層特徵點模式，但以分層特徵點模式預測時會受到資料量影響，可從第四章模擬中可看出，在資料量過少情況下，預測較不穩健，因此雖然能些微改善未發生標誌事件個案預測的準確度，但在建構模式時需要特別注意到這兩者間的權衡。

在第五章實例分析中，本文以兩筆資料進行分析，兩筆資料最大差異為標誌時間平均發生的長短，因此可以比較本文建構之特徵點模式在此兩種情況下預測的結果，實務上，年齡或其他預後因素也須納入，以建構更好的預測模式。本論文設計更具彈性的特徵點模式估計未來時間點下之存活機率之動態預測，提供在實際臨床運用上參考依據。

doi:10.6342/NTU201903431

第七章 參考文獻

Cortese, G., Gerds, T. A., and Anderson, P. K. (2013). Comparing predictions among competing risks models with time-dependent covariates. *Statistics in Medicine, 32, 3089-3101.*

Dancourt, V., Quantin, C., Abrahamowicz, M., Binquet, C., Alioum, A. and Faivre, J. (2004). Modeling recurrence in colorectal cancer. *Journal of Clinical Epidemiology, 57, 243-251.*

Dartigues, D-F., Gagnon, M., Barberger-Gateau, P., Letenneur, L., Commenges, D., Sauvel, C., Michel, P. and Salamon, R. (1992). The paquid epidemiological program on brain ageing. *Neuroepidemiology, 14-18.*

Etzioni, R., Pepe, M., Longton, G., Hu, C. and Goodman, G., (1999) Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Medical Decision Making, 19, 242-251.*

Graf, E., Schmoor, C., Sauerbrei, W. and Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine, 18, 2529-2545.*

Hanley, J. A. and McNeil, B., (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143, 29-36.*

Heagerty, P. J., and Zheng, Y. (2005) Survival Model Predictive Accuracy and ROC Curves. *Biometrics, 61, 92-105.*

doi:10.6342/NTU201903431

Hilden, J., Habbema, J. D. F. and Bjerregard, B. (1978). The measurement of performance in probabilistic diagnosis III: methods based on continuous functions of the diagnostic probabilities. *Methods of Information in Medicine, 17, 238-246.*

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

Laurie, JA., Moertel, CG., Fleming, TR., Wieand, HS., Leigh, JE., Rubin, J., McCormack, GW., Gerstner, JB., Krook, JE. and Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: An evaluation of levamisole and the combination of levamisole and fluorouracil: The North Central Cancer Treatment Group and the Mayo Clinic. *Journal of Clinical Oncology, 7:1447-1456.*

Linnet, K. (1989). Assessing diagnostic tests by a strictly proper scoring rule. *Statistics in Medicine, 8, 609-618.*

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., and Sestelo, M. (2019). Estimation in the progressive illness-death model: A nonexhaustive review. *Biometrical Journal, 61, 245-263.*

Neri, A., Marrelli, D., Rossi, S., De Stefano, A., Mariani, F., De Marco, G., Caruso, S., Corso, G., Cioppa, T., Pinto, E. and Roviello, F. (2007). Breast cancer local recurrence: Risk factors and prognostic relevance of early time to recurrence. *World Journal of Surgery, 31, 36-45.*

doi:10.6342/NTU201903431

Peng, W-C. (2010). Estimation of Conditional Survival probabilities using marker time as predictors. *Division of Biostatistics, Graduate Institute of Epidemiology College of Public Health, National Taiwan University Master Thesis*.

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.*

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67, 819-829.*

Rizopoulos, D., Murawska, M., Andrinopoulou, E-R., Molenberghs, G., Takkenberg, J. J. and Lesaffre, E. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal, 59, 1261-1276*.

Suresh, K., Jeremy, M.G.T., Daniel, E.S., Daignault, S. and Tsodikov, A. (2017). Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical Journal, 59, 1277-1300.*

Slate, E. H. and Turnbull, B. W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistical in Medicine, 19, 617-637.*

Spiegelhalter, D. J. (1986). Probabilistic prediction in patient management and clinical trials. *Statistics in Medicine, 5, 421-433.*

Su, D-H. (2016). Statistical analysis for dynamic survival prediction involving marker processes. *Division of Biostatistics, Graduate Institute of Epidemiology College of Public Health, National Taiwan University Doctoral Dissertation.*

Taylor, J. M. G., Yu, M. and Sander, H. (2005). Individualized Predictions of Disease Progression Following Radiation Therapy for Prostate Cancer. *Journal of Clinical Oncology, 23, 816-825.*

doi:10.6342/NTU201903431

Taylor, J. M. G., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T. and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics, 69, 206-213.*

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.*

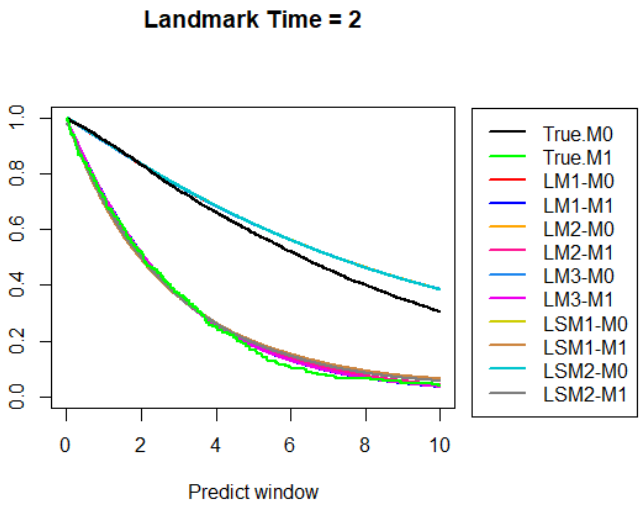
Yang, C-Y. (2009). Predicting survival with disease progression as a time-dependent covariate in proportional hazards model. *Division of Biostatistics, Graduate Institute of Epidemiology College of Public Health, National Taiwan University Master Thesis.*

doi:10.6342/NTU201903431

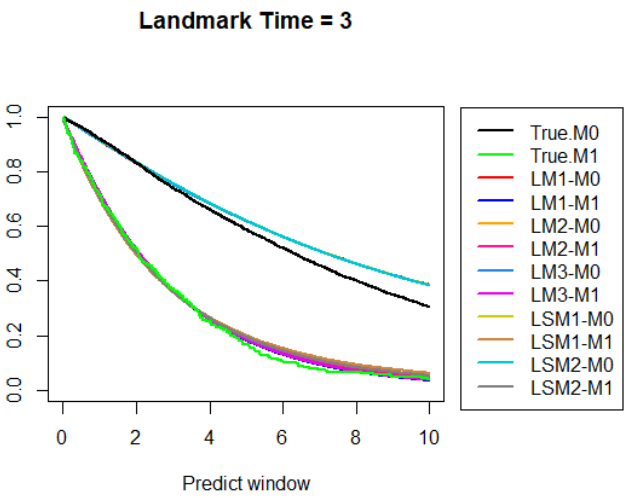
附錄

在 Kendall's tau 為 0.6，特徵時間點$u = 2$、3、5下，預測未來 10 年存活曲線圖以及各項評估準則。

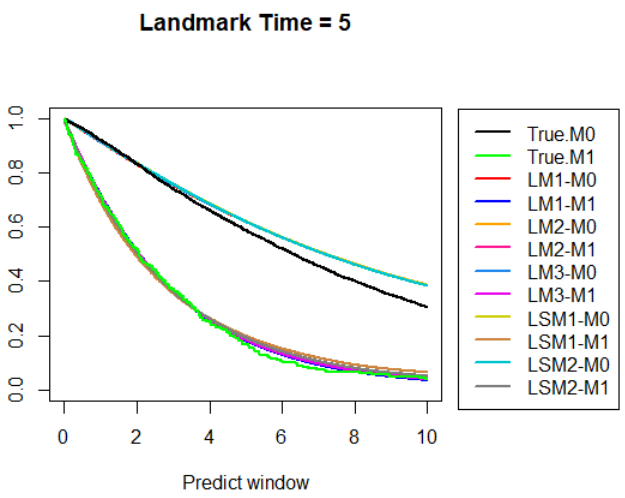
- **時間依賴性 Cox 模式**



圖附-1：標誌時間點 2 年下預測未來 10 年存活曲線圖

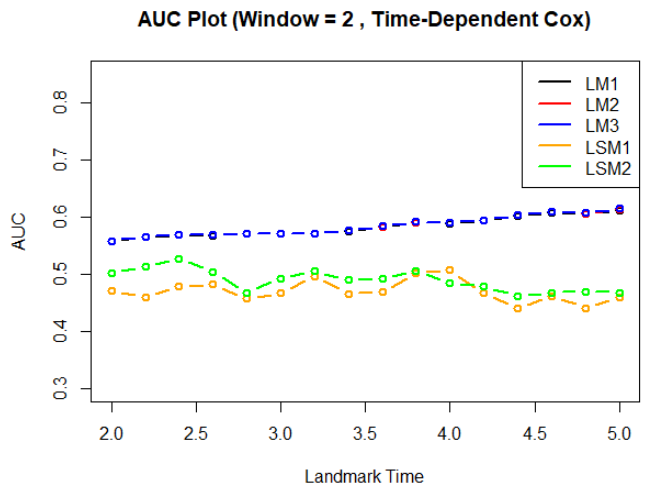


圖附-2：標誌時間點 3 年下預測未來 10 年存活曲線圖

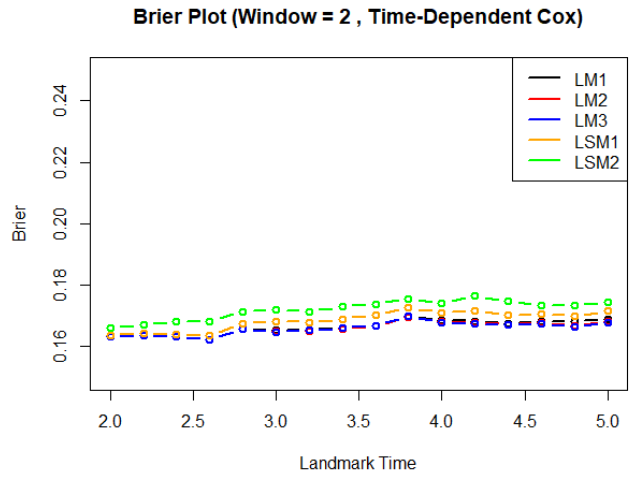


圖附-3：標誌時間點 5 年下預測未來 10 年存活曲線圖

doi:10.6342/NTU201903431

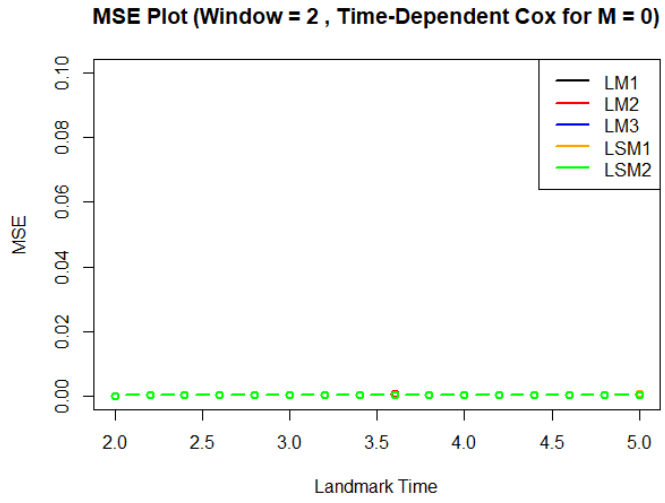


圖附-4：各特徵點模式預測 2 年後 AUC 表現

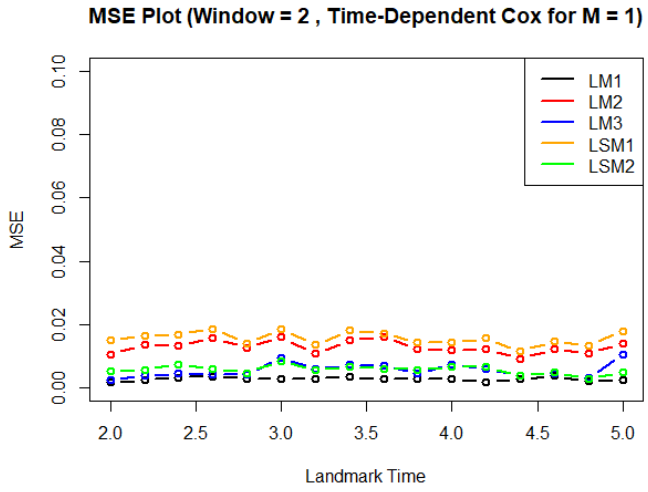


圖附-5：各特徵點模式預測 2 年後 Brier 分數表現

圖型標題解釋(Window:預測區間，資料生成方式)

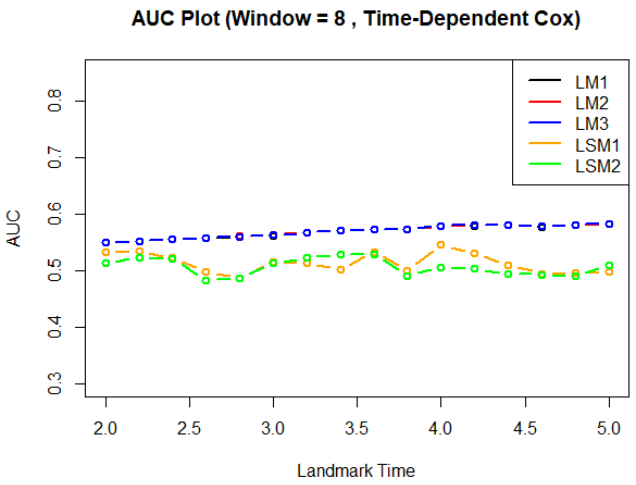


圖附-6：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差

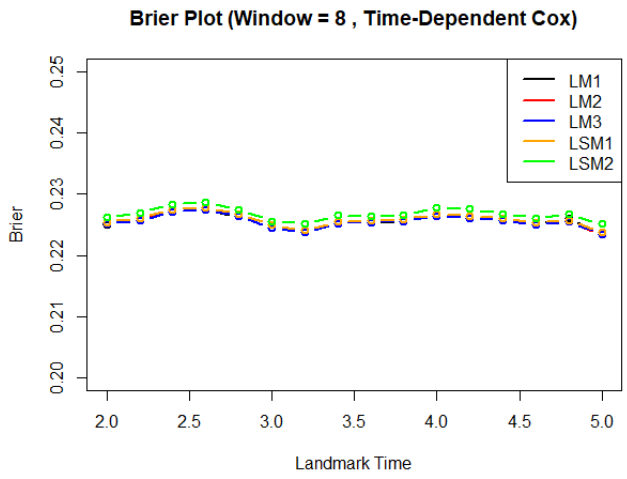


圖附-7：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

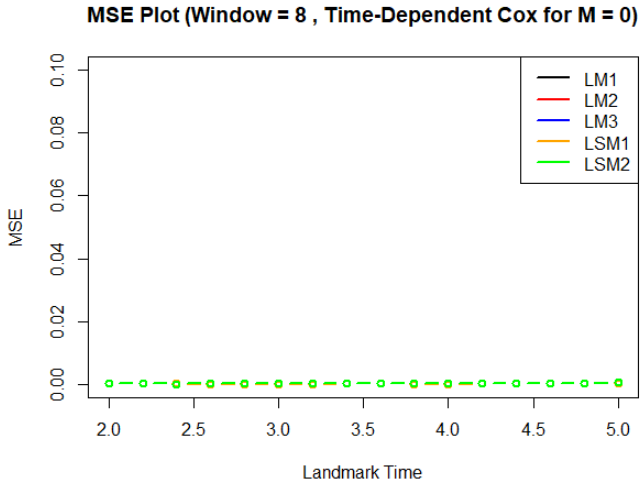
doi:10.6342/NTU201903431



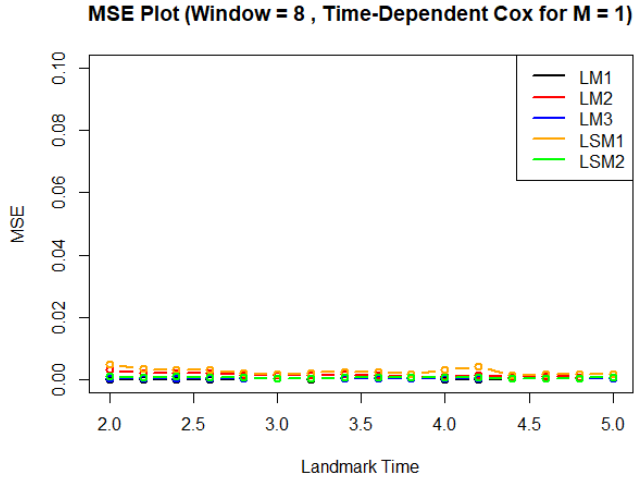
圖附-8：各特徵點模式預測 8 年後 AUC 表現



圖附-9：各特徵點模式預測 8 年後 Brier 分數表現

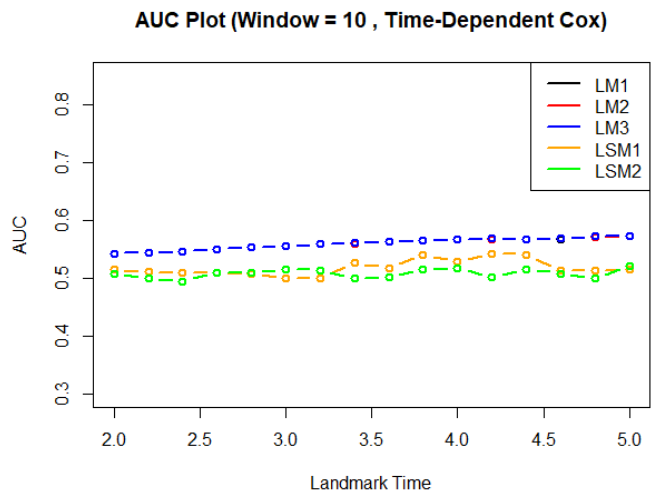


圖附-10：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差

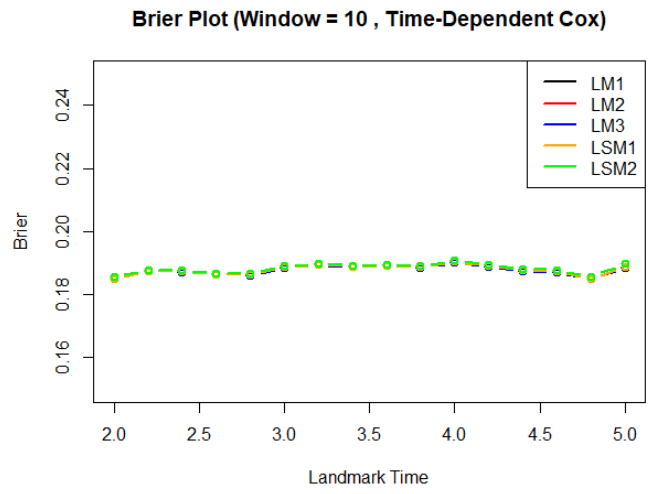


圖附-11：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

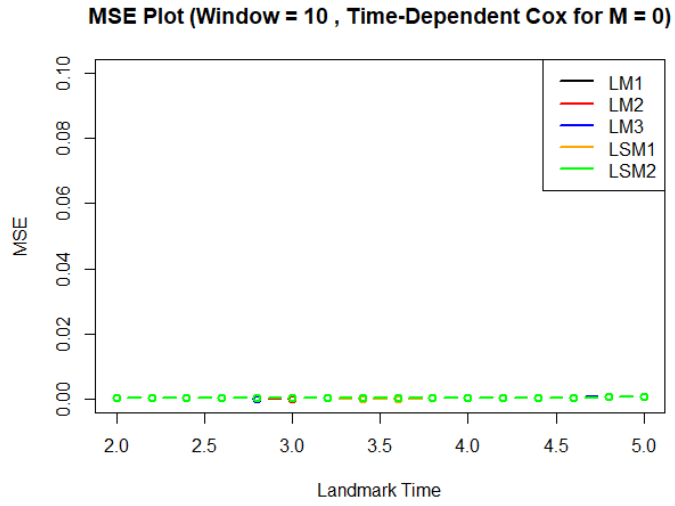
doi:10.6342/NTU201903431



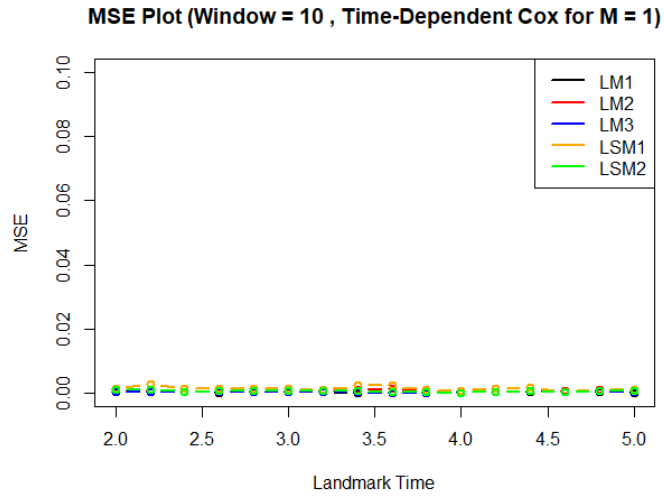
圖附-12：各特徵點模式預測 10 年後 AUC 表現



圖附-13：各特徵點模式預測 10 年後 Brier 分數表現



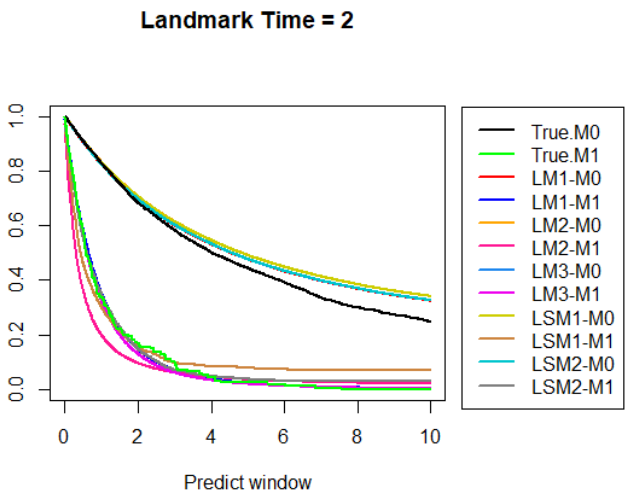
圖附-14：各特徵點模式預測未發生標誌事件個案 10 年後存活機率均方差



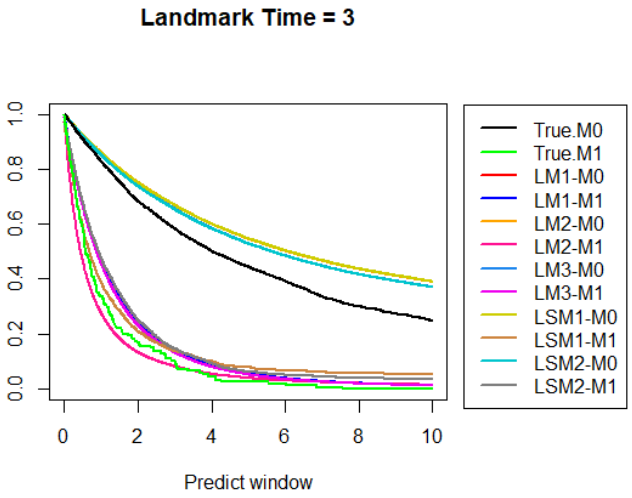
圖附-15：各特徵點模式預測已發生標誌事件個案 10 年後存活機率均方差

doi:10.6342/NTU201903431

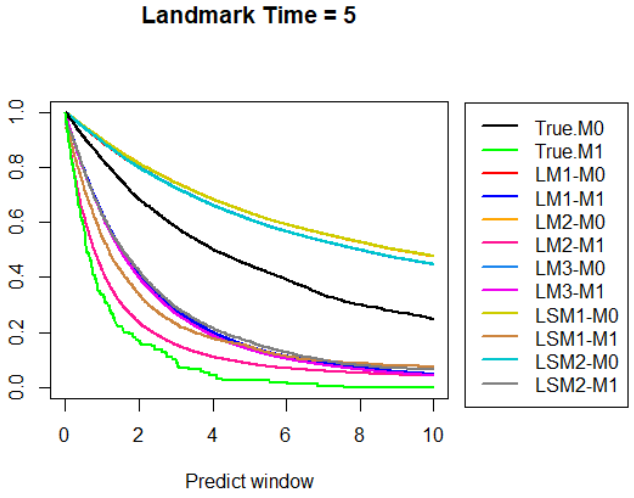
- **對數常態(Log Normal)**



圖附-16：標誌時間點 2 年下預測未來 10 年存活曲線圖

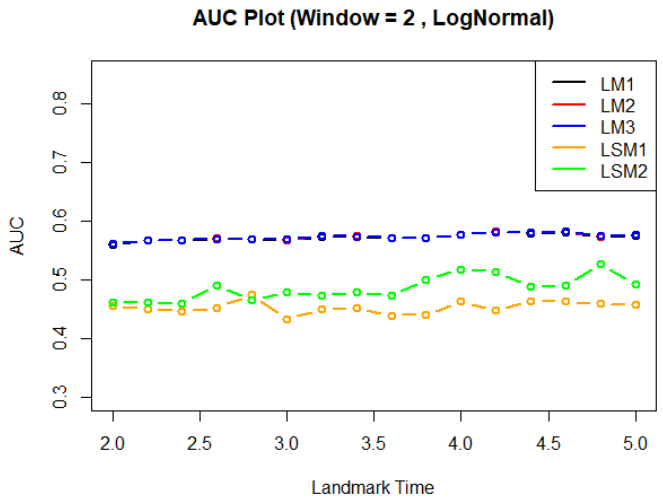


圖附-17：標誌時間點 3 年下預測未來 10 年存活曲線圖

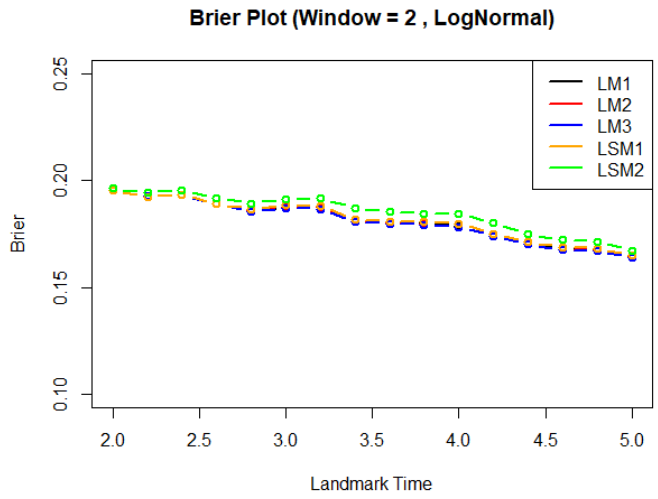


圖附-18：標誌時間點 5 年下預測未來 10 年存活曲線圖

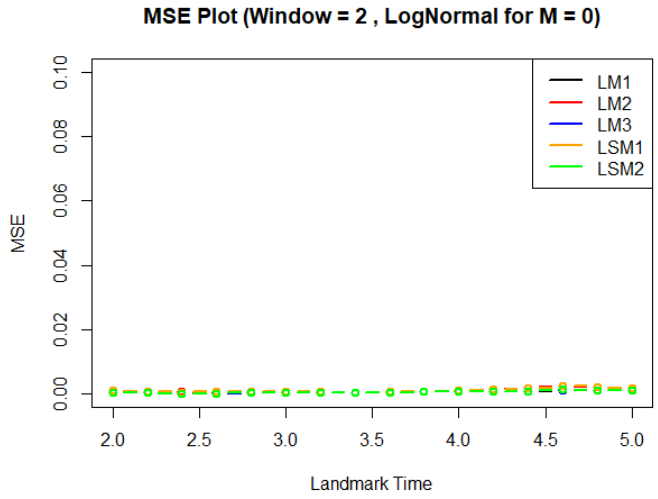
doi:10.6342/NTU201903431



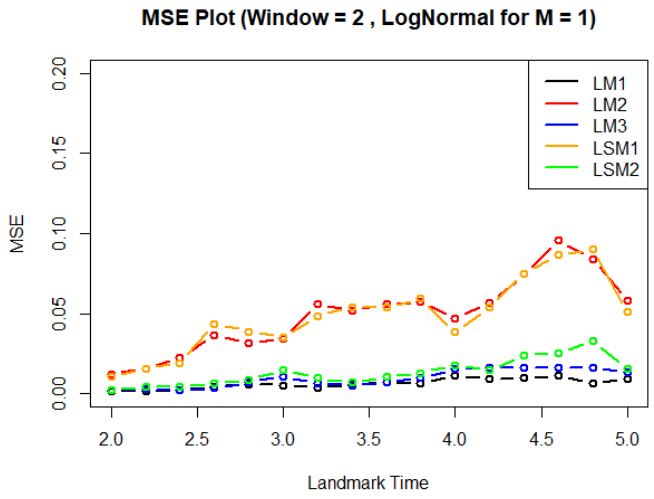
圖附-19：各特徵點模式預測 2 年後 AUC 表現



圖附-20：各特徵點模式預測 2 年後 Brier 分數表現

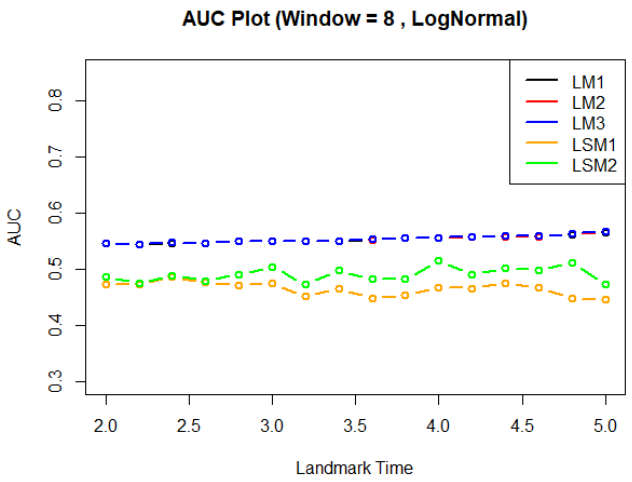


圖附-21：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差



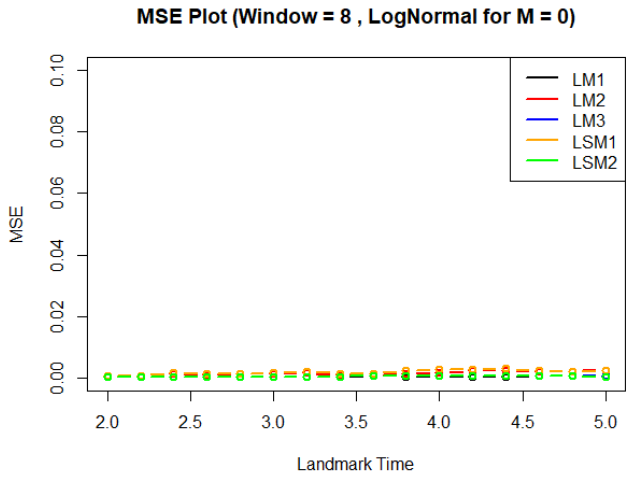
圖附-22：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

doi:10.6342/NTU201903431

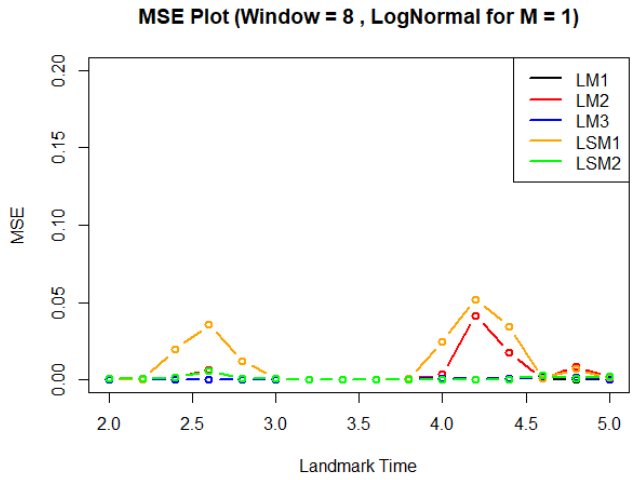


圖附-23：各特徵點模式預測 8 年後 AUC 表現

圖附-24：各特徵點模式預測 8 年後 Brier 分數表現

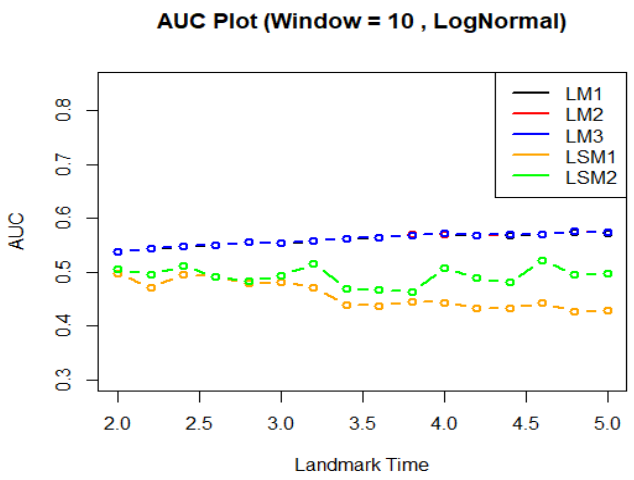


圖附-25：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差

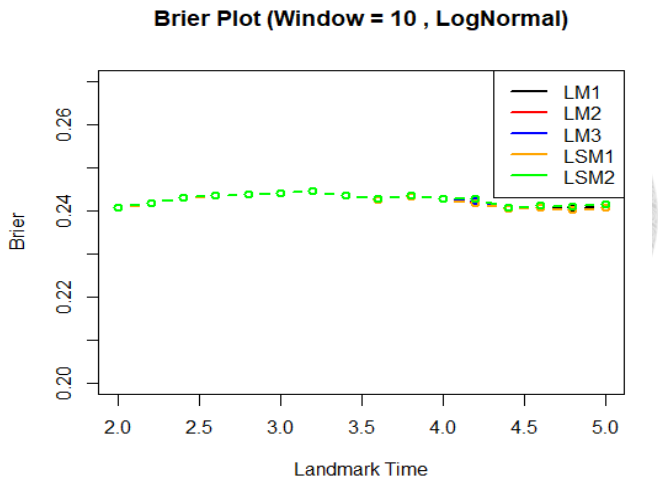


圖附-26：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

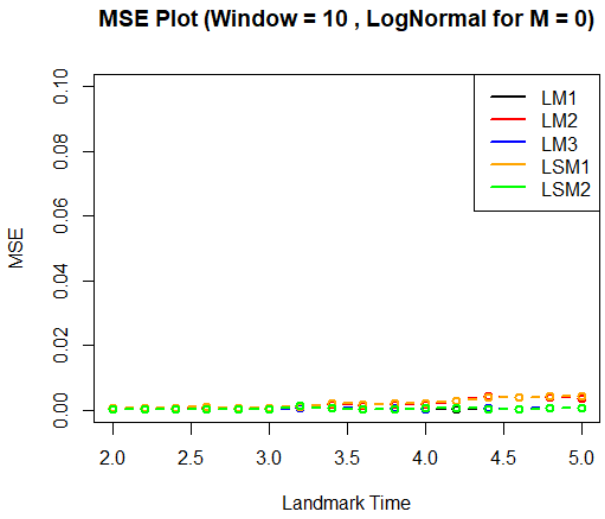
doi:10.6342/NTU201903431



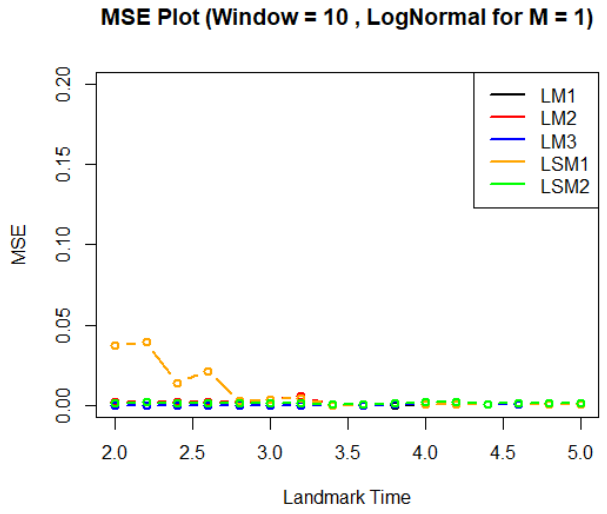
圖附-27：各特徵點模式預測 10 年後 AUC 表現



圖附-28 ：各特徵點模式預測 10 年後 Brier 分數表現



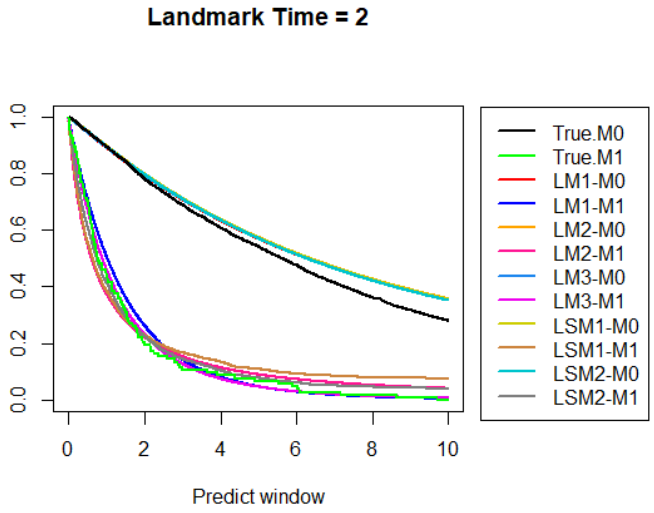
圖附-29：各特徵點模式預測未發生標誌事件個案 10 年後存活機率均方差



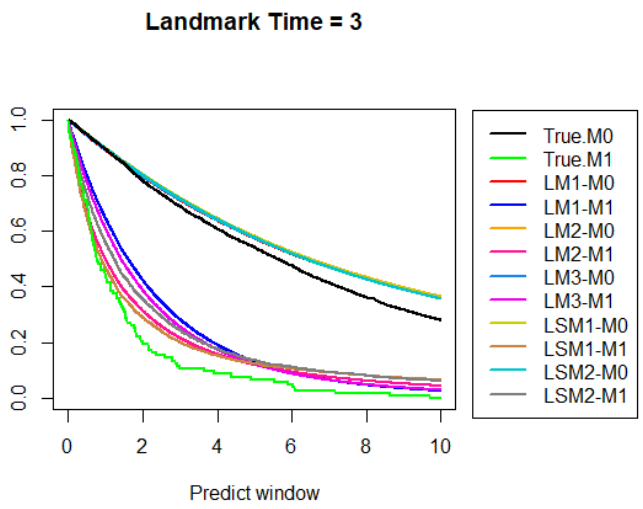
圖附-30：各特徵點模式預測已發生標誌事件個案 10 年後存活機率均方差

doi:10.6342/NTU201903431

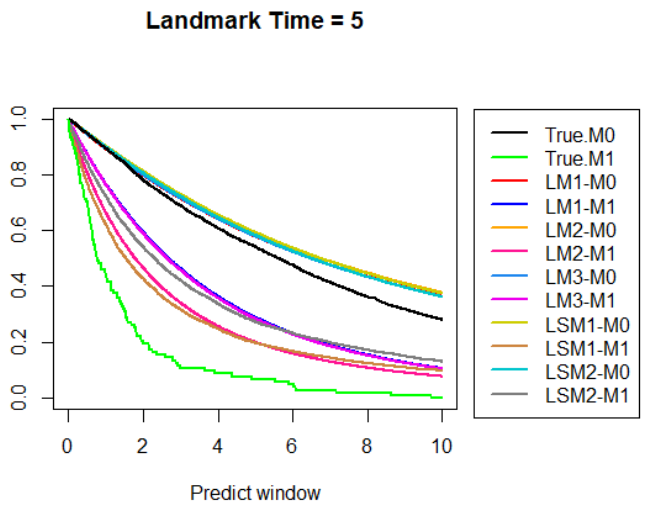
- **Clayton Copula**



圖附-31：標誌時間點 2 年下預測未來 10 年存活曲線圖

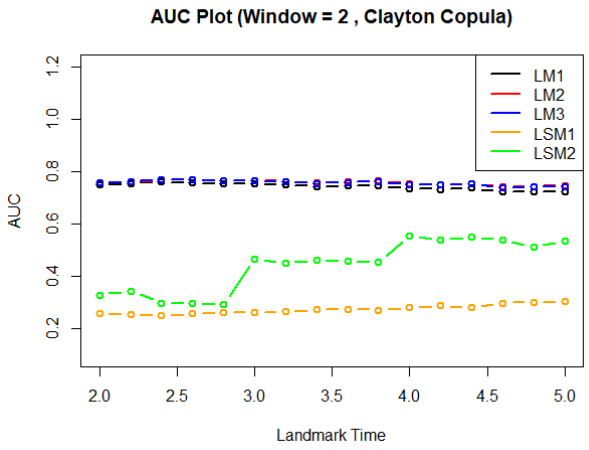


圖附-32：標誌時間點 3 年下預測未來 10 年存活曲線圖

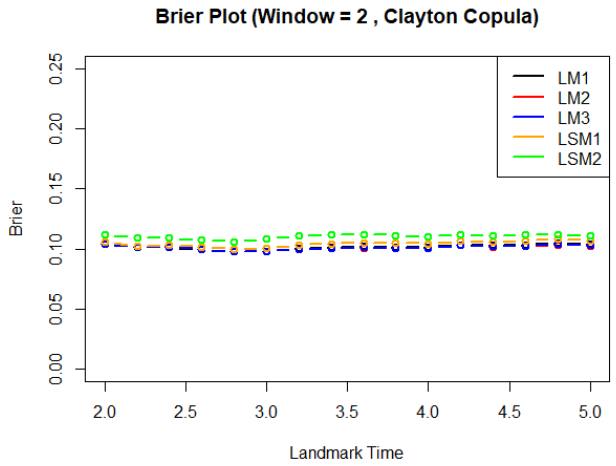


圖附-33：標誌時間點 5 年下預測未來 10 年存活曲線圖

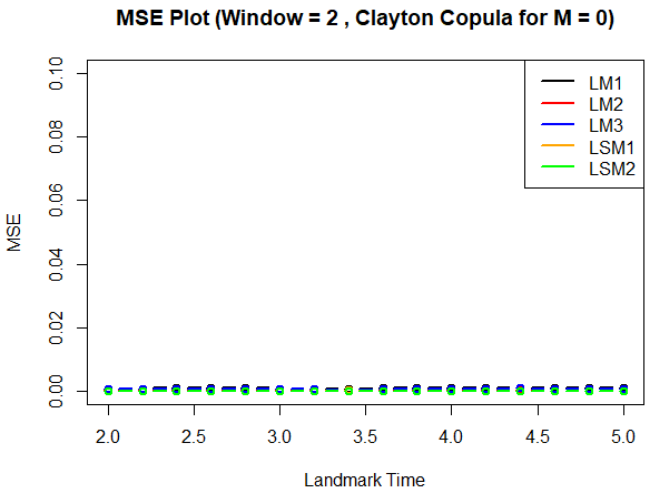
doi:10.6342/NTU201903431



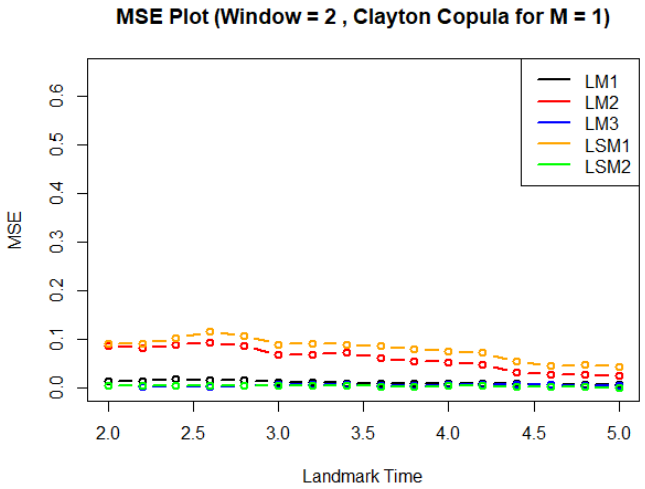
圖附-34：各特徵點模式預測 2 年後 AUC 表現



圖附-35 ：各特徵點模式預測 2 年後 Brier 分數表現

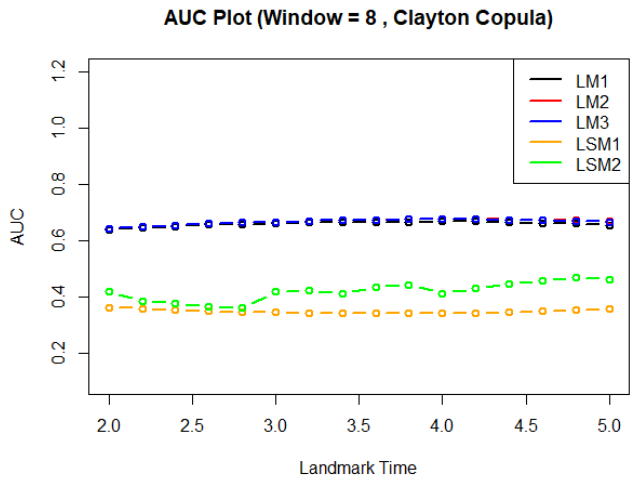


圖附-36：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差

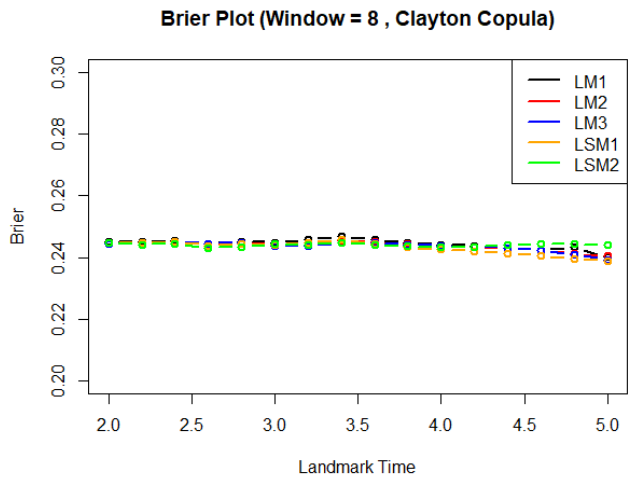


圖附-37：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

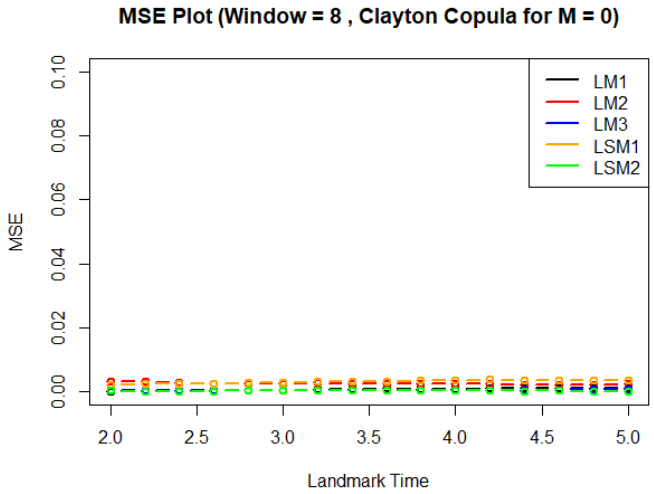
doi:10.6342/NTU201903431



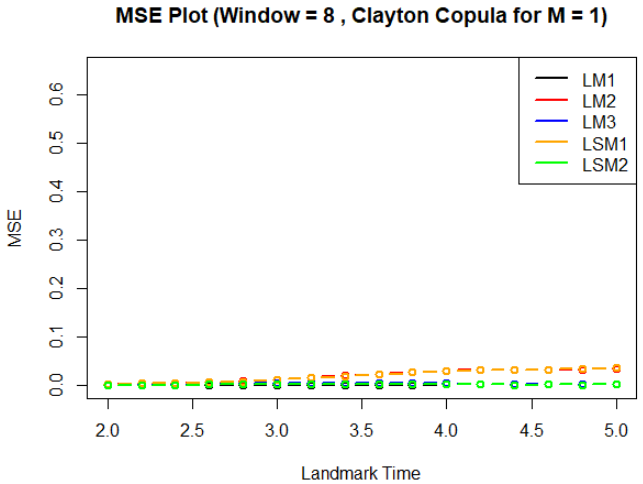
圖附-38：各特徵點模式預測 8 年後 AUC 表現



圖附-39 ：各特徵點模式預測 8 年後 Brier 分數表現

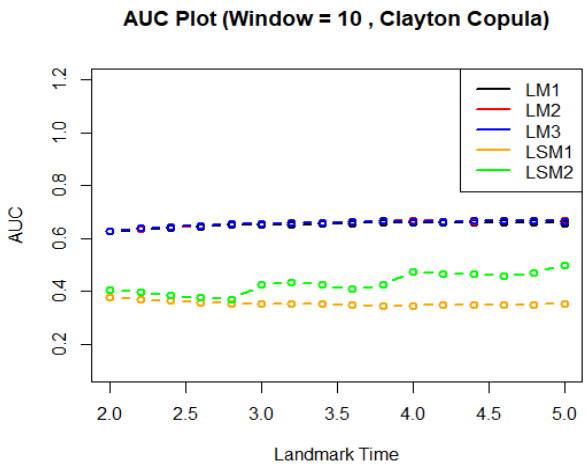


圖附-40：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差

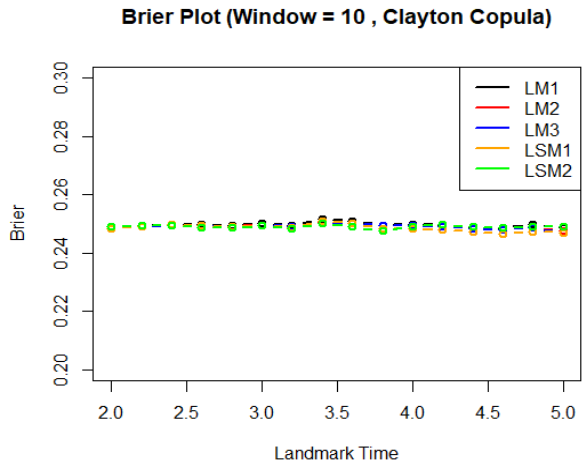


圖附-41：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

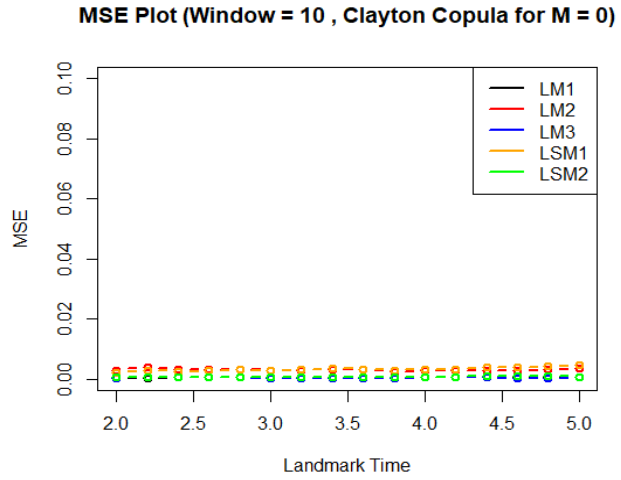
doi:10.6342/NTU201903431



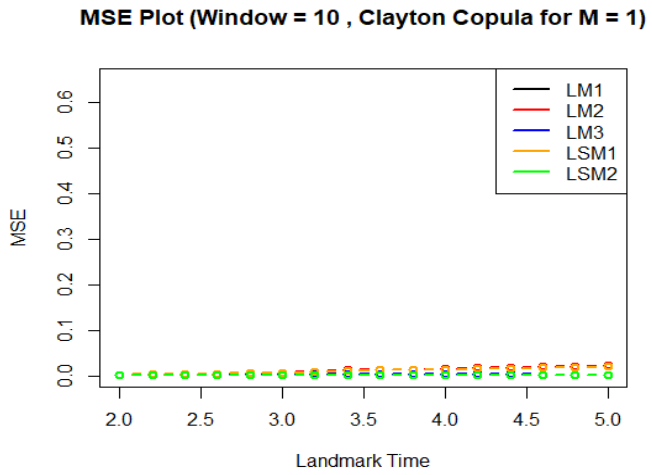
圖附-42：各特徵點模式預測 10 年後 AUC 表現



圖附-43 ：各特徵點模式預測 10 年後 Brier 分數表現



圖附-44：各特徵點模式預測未發生標誌事件個案 10 年後存活機率均方差

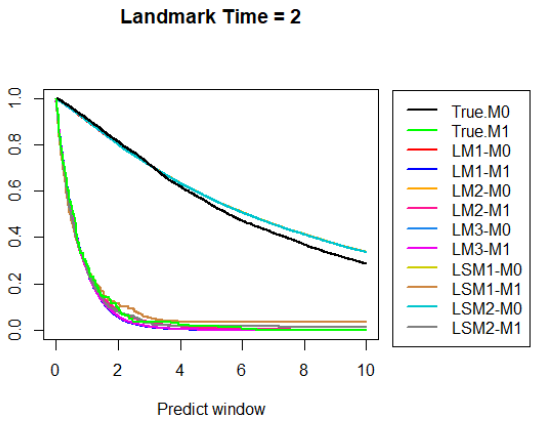


圖附-45：各特徵點模式預測已發生標誌事件個案 10 年後存活機率均方差

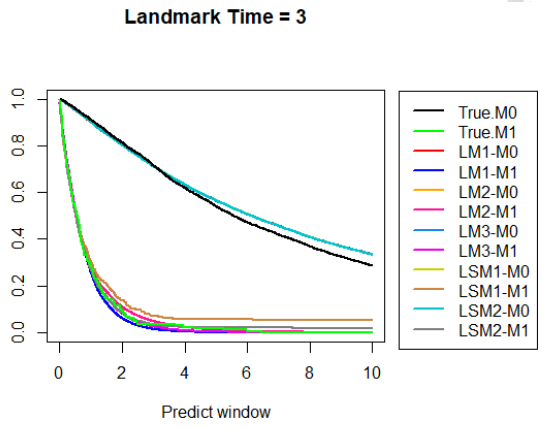
doi:10.6342/NTU201903431

在 Kendall's tau 為 0.8，特徵時間點$u = 2$、3、4、5下，預測未來 10 年存活曲線圖以及各項評估準則。

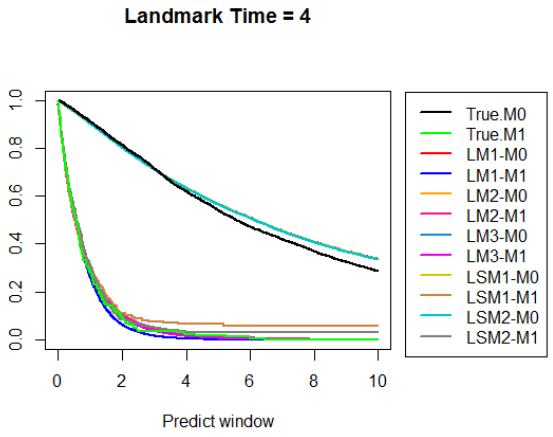
- **時間依賴性 Cox 模式**



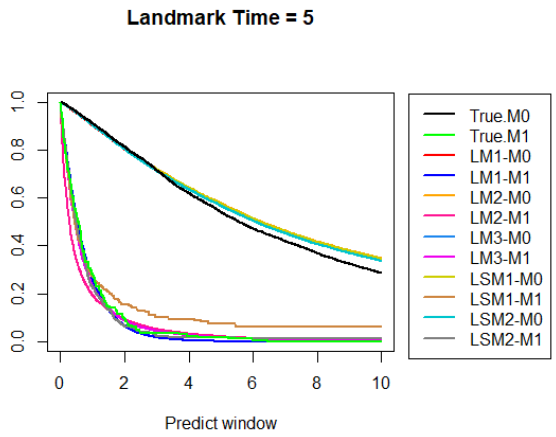
圖附-46：標誌時間點 2 年下預測未來 10 年存活曲線圖



圖附-47：標誌時間點 3 年下預測未來 10 年存活曲線圖

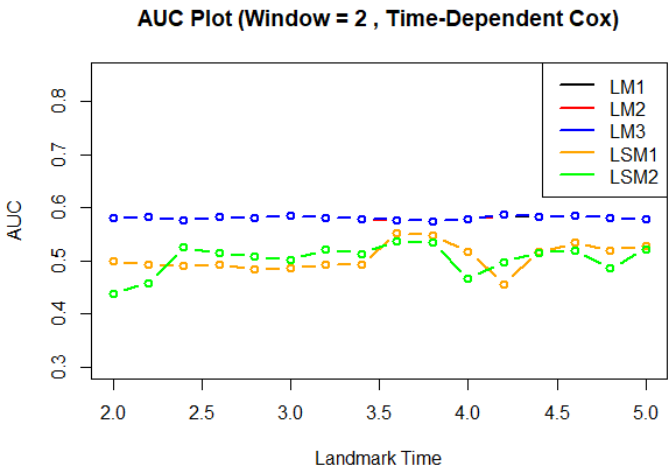


圖附-48：標誌時間點 4 年下預測未來 10 年存活曲線圖

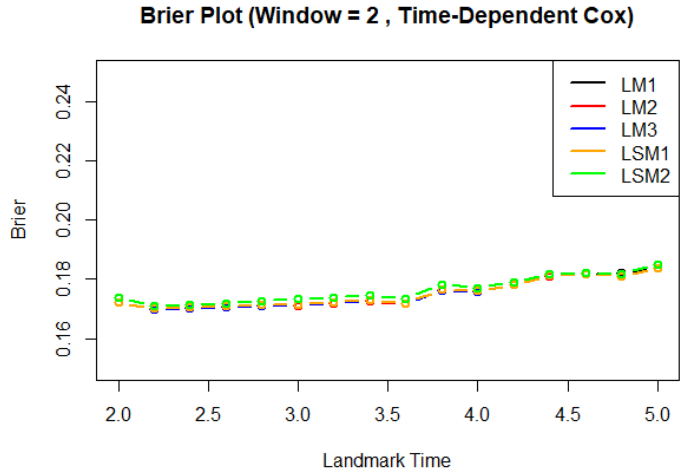


圖附-49：標誌時間點 5 年下預測未來 10 年存活曲線圖

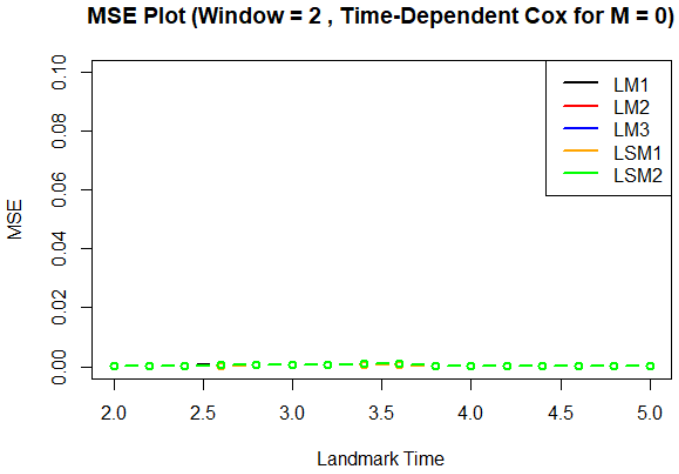
doi:10.6342/NTU201903431



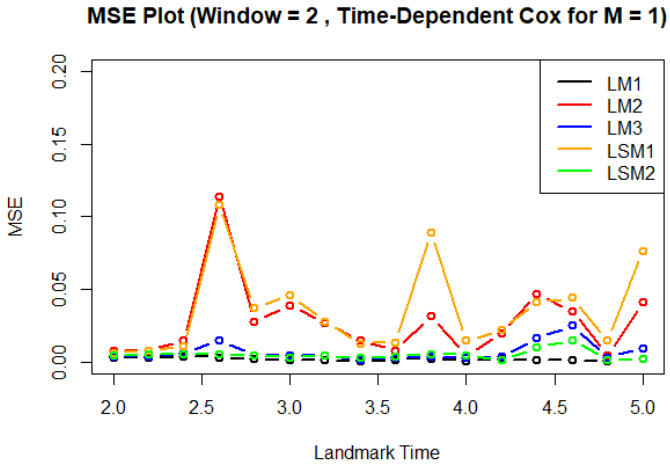
圖附-50：各特徵點模式預測 2 年後 AUC 表現



圖附-51 ：各特徵點模式預測 2 年後 Brier 分數表現

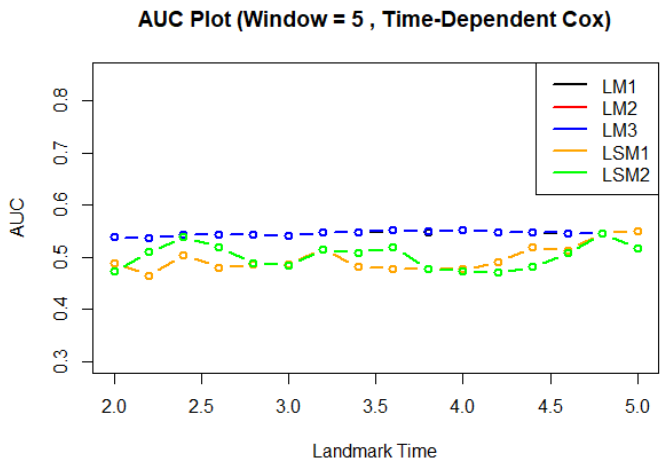


圖附-52：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差



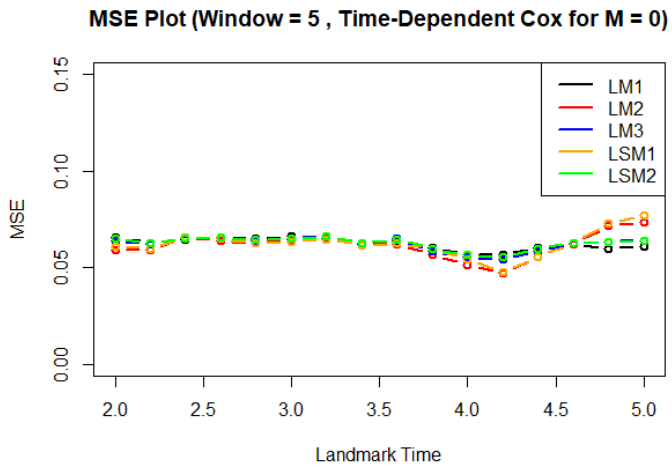
圖附-53：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

doi:10.6342/NTU201903431

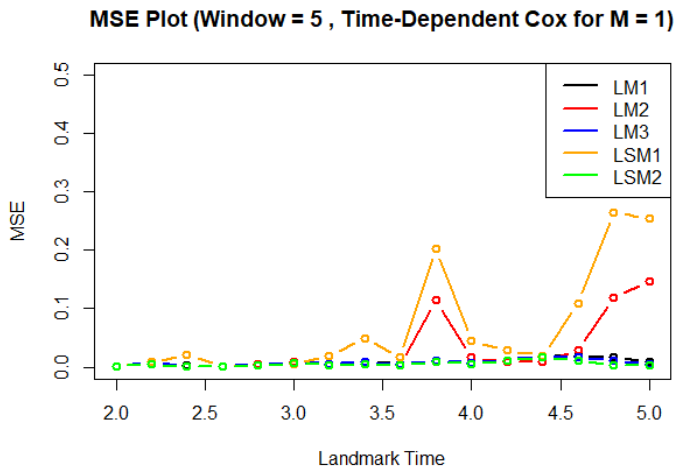


圖附-54：各特徵點模式預測 5 年後 AUC 表現

圖附-55：各特徵點模式預測 5 年後 Brier 分數表現

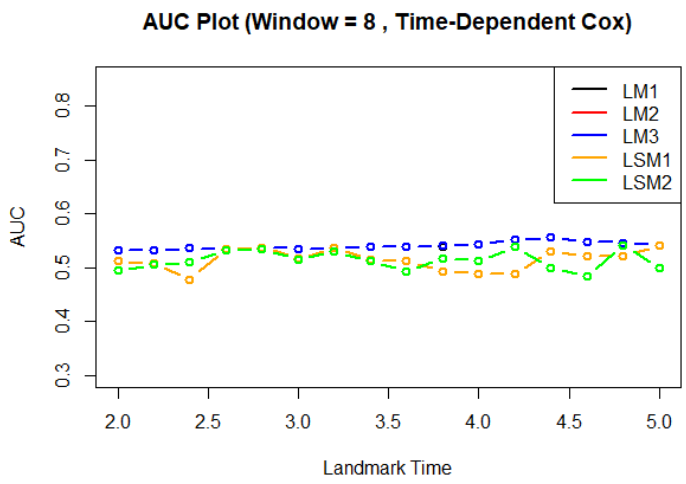


圖附-56：各特徵點模式預測未發生標誌事件個案 5 年後存活機率均方差

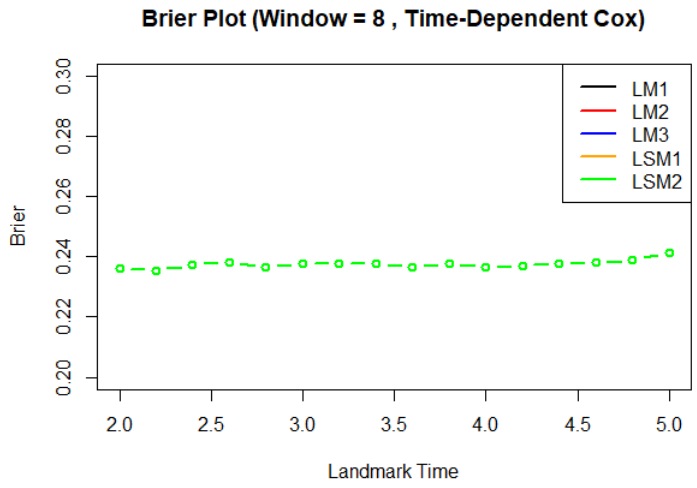


圖附-57：各特徵點模式預測已發生標誌事件個案 5 年後存活機率均方差

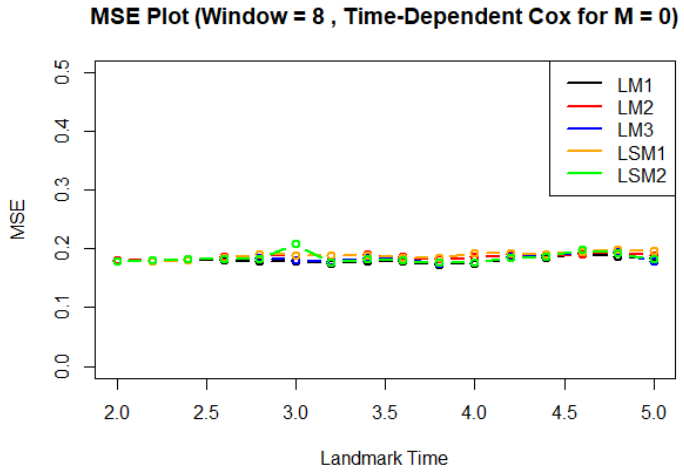
doi:10.6342/NTU201903431



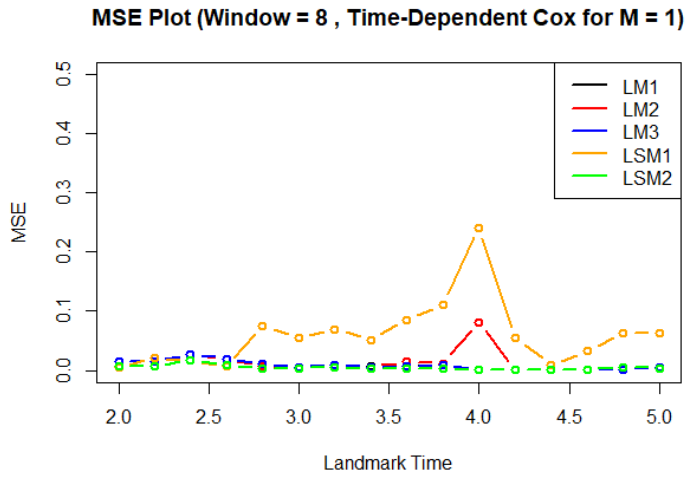
圖附-58：各特徵點模式預測 8 年後 AUC 表現



圖附-59：各特徵點模式預測 8 年後 Brier 分數表現

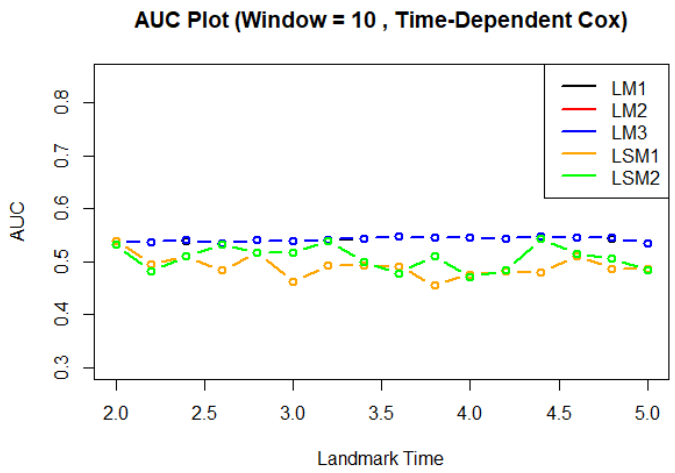


圖附-60：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差



圖附-61：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

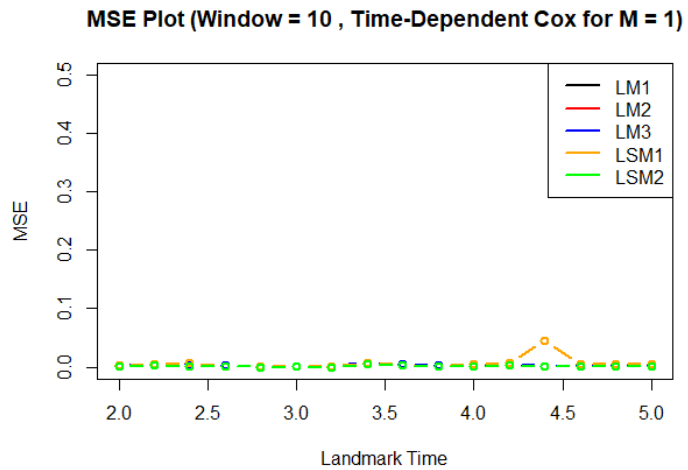
doi:10.6342/NTU201903431



圖附-62：各特徵點模式預測 10 年後 AUC 表現

圖附-63：各特徵點模式預測 10 年後 Brier 分數表現

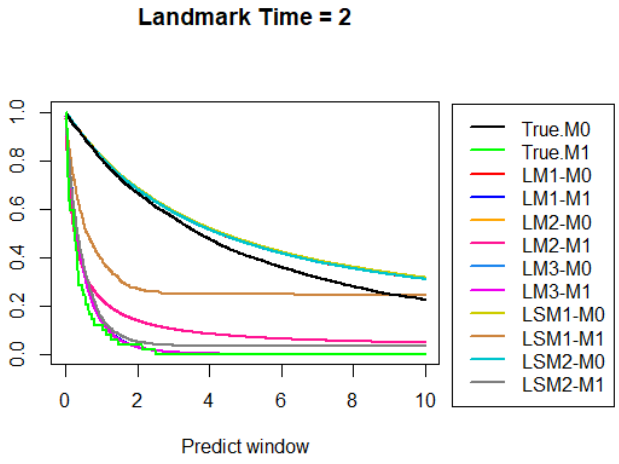
圖附-64：各特徵點模式預測未發生標誌事件個案 10 年後存活機率均方差



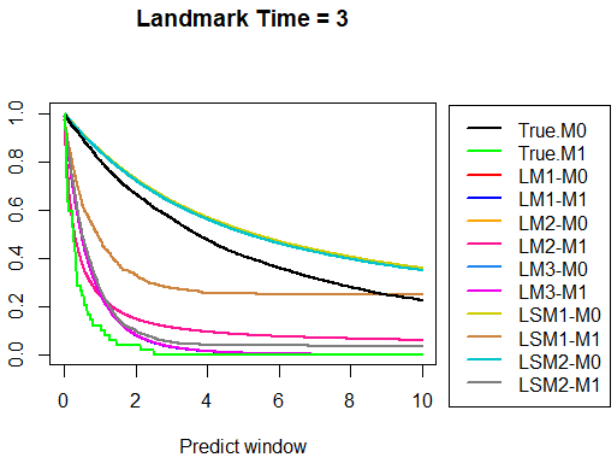
圖附-65：各特徵點模式預測已發生標誌事件個案 10 年後存活機率均方差

doi:10.6342/NTU201903431

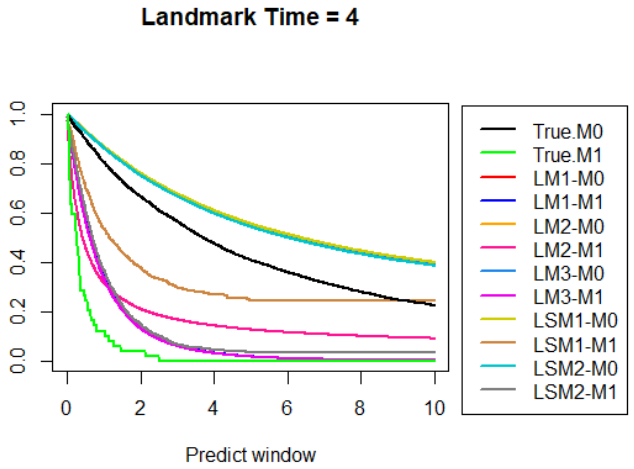
- **對數常態(Log Normal)**



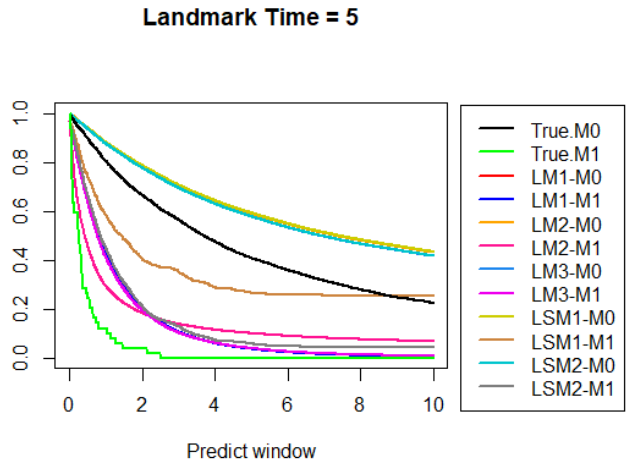
圖附-66：標誌時間點 2 年下預測未來 10 年存活曲線圖



圖附-67：標誌時間點 3 年下預測未來 10 年存活曲線圖

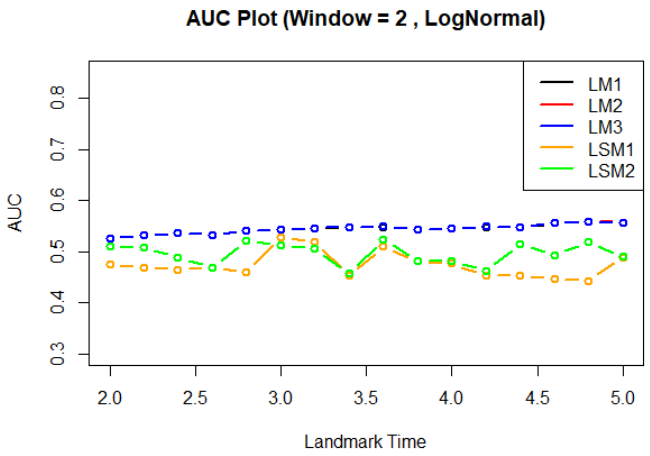


圖附-68：標誌時間點 4 年下預測未來 10 年存活曲線圖



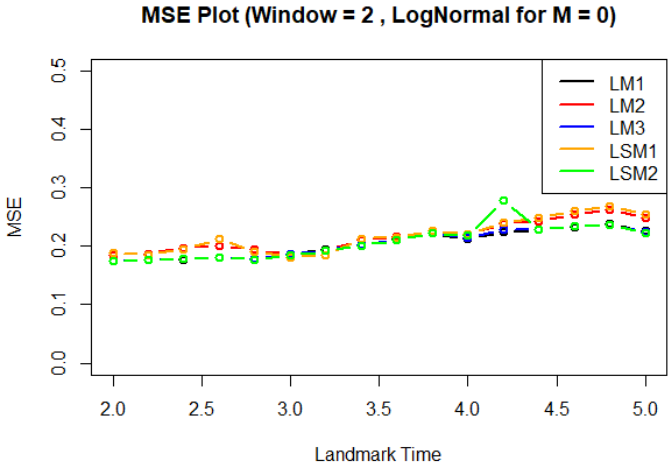
圖附-69：標誌時間點 5 年下預測未來 10 年存活曲線圖

doi:10.6342/NTU201903431

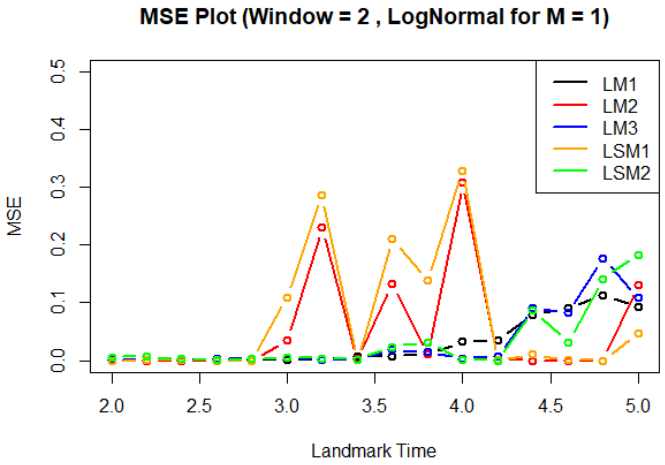


圖附-70：各特徵點模式預測 2 年後 AUC 表現

圖附-71 ：各特徵點模式預測 2 年後 Brier 分數表現

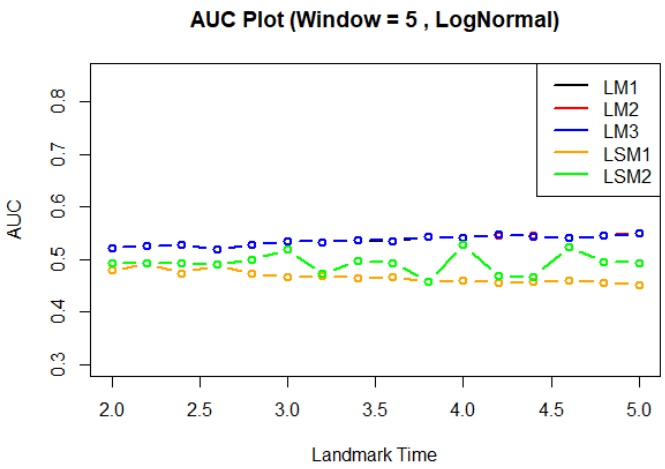


圖附-72：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差

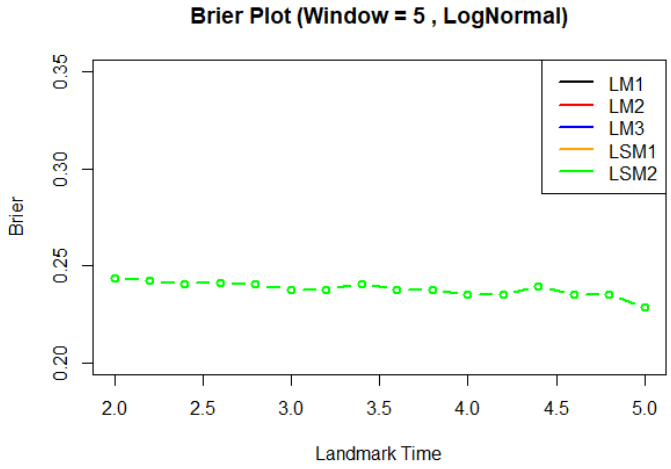


圖附-73：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

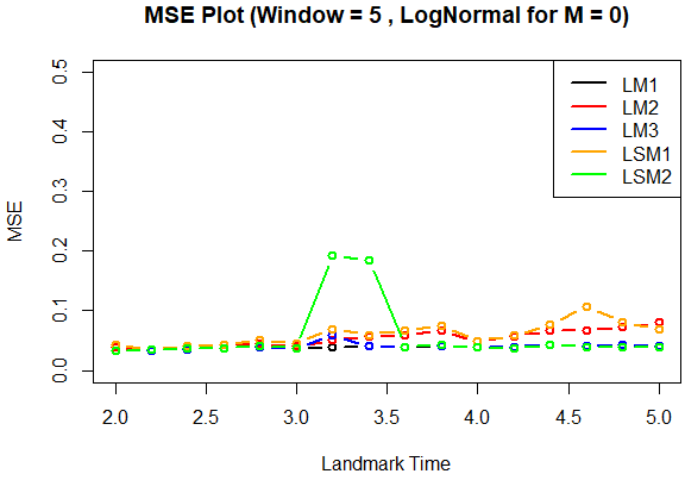
doi:10.6342/NTU201903431



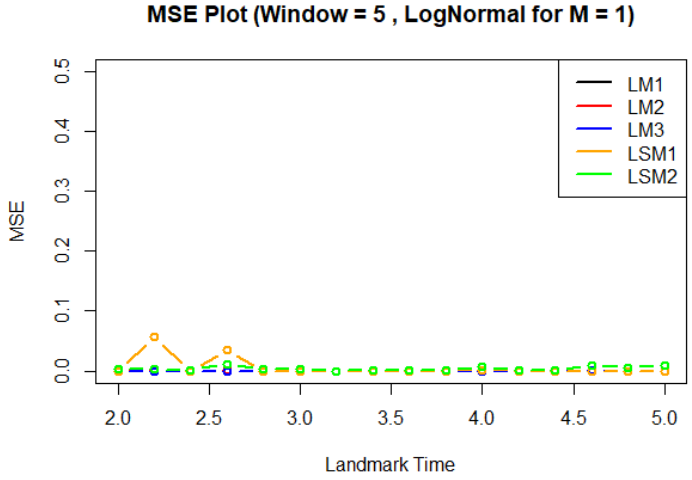
圖附-74：各特徵點模式預測 5 年後 AUC 表現



圖附-75：各特徵點模式預測 5 年後 Brier 分數表現

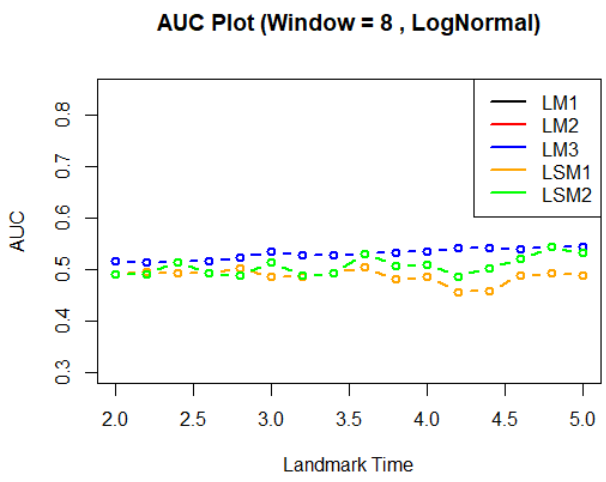


圖附-76：各特徵點模式預測未發生標誌事件個案 5 年後存活機率均方差

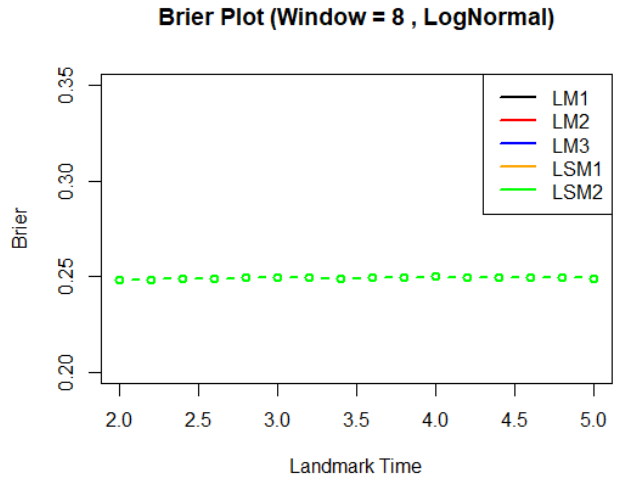


圖附-77：各特徵點模式預測已發生標誌事件個案 5 年後存活機率均方差

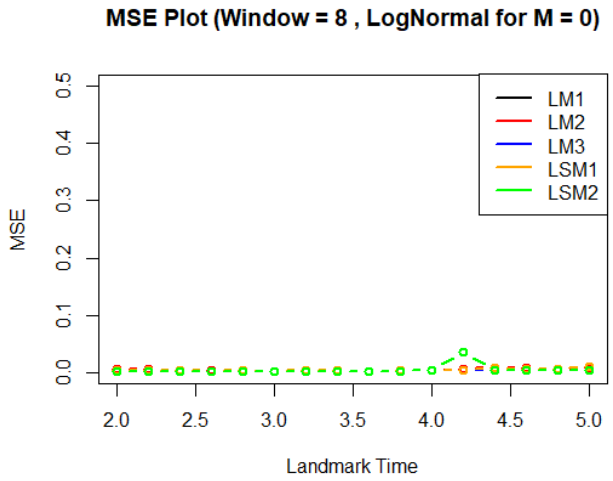
doi:10.6342/NTU201903431



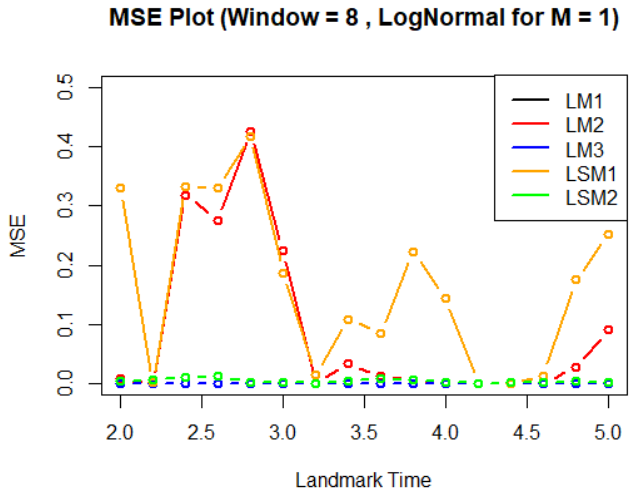
圖附-78：各特徵點模式預測 8 年後 AUC 表現



圖附-79：各特徵點模式預測 8 年後 Brier 分數表現

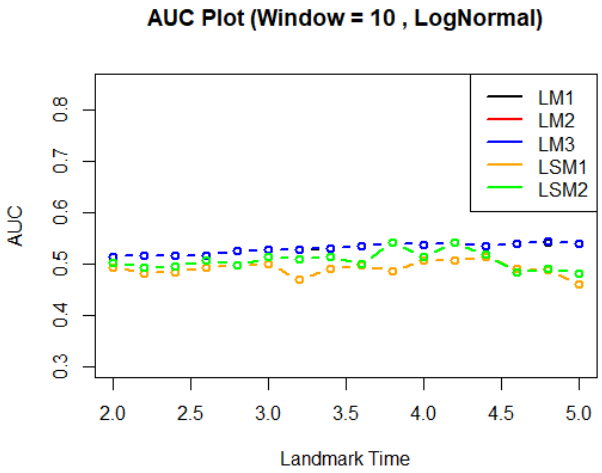


圖附-80：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差

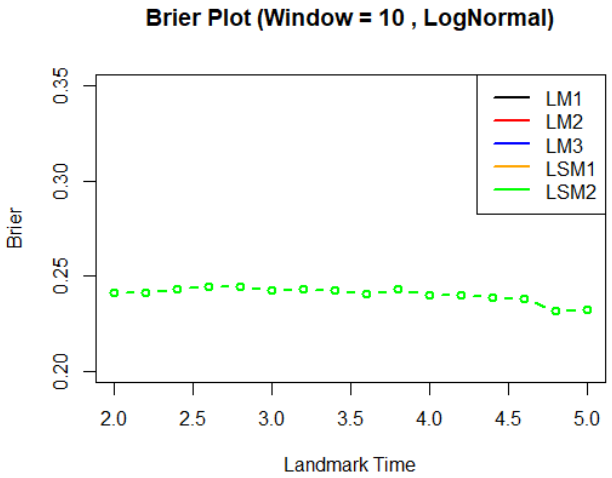


圖附-81：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

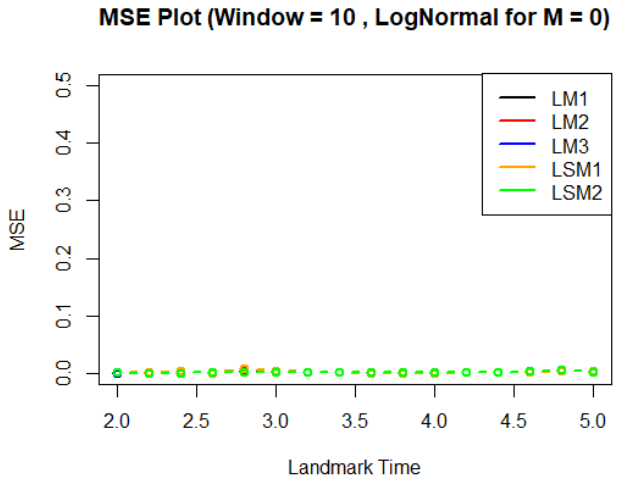
doi:10.6342/NTU201903431



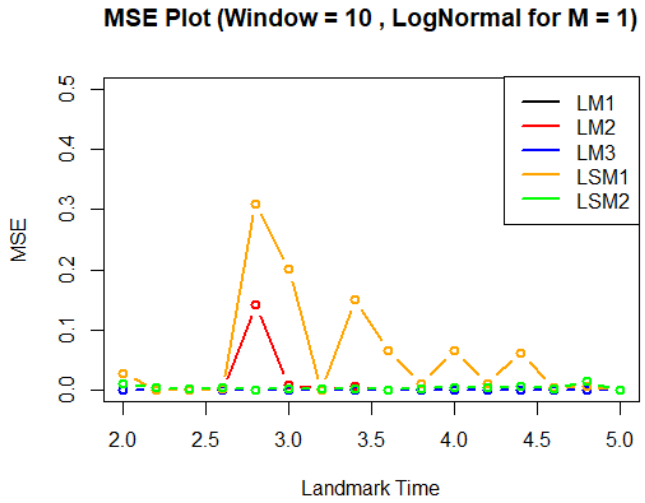
圖附-82：各特徵點模式預測 10 年後 AUC 表現



圖附-83：各特徵點模式預測 10 年後 Brier 分數表現



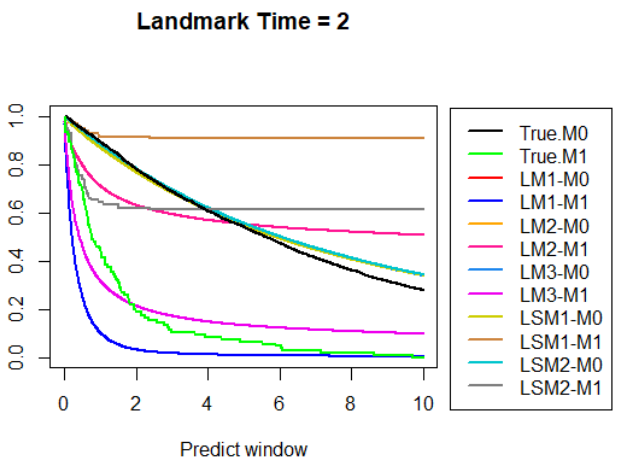
圖附-84：各特徵點模式預測未發生標誌事件個案 10 年後存活機率均方差



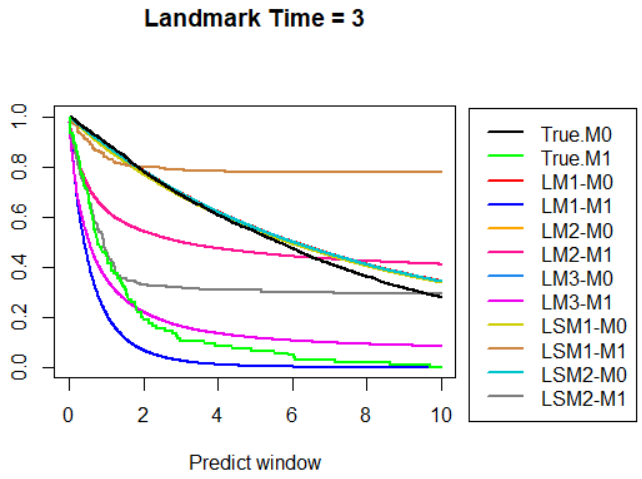
圖附-85：各特徵點模式預測已發生標誌事件個案 10 年後存活機率均方差

doi:10.6342/NTU201903431

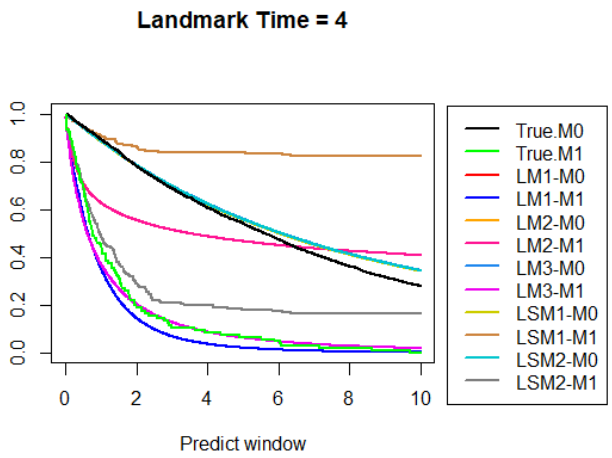
- **Clayton Copula**



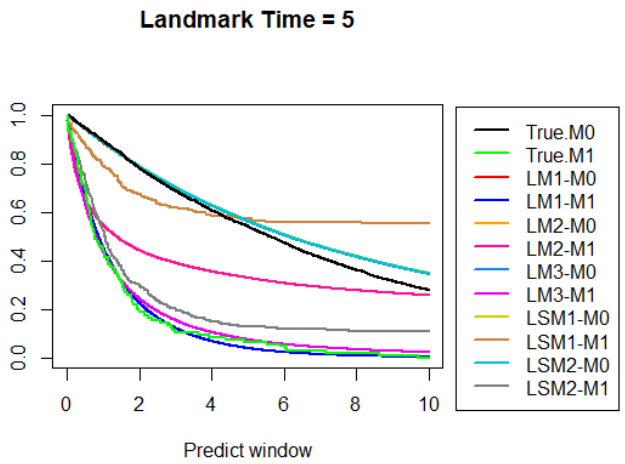
圖附-86：標誌時間點 2 年下預測未來 10 年存活曲線圖



圖附-87：標誌時間點 3 年下預測未來 10 年存活曲線圖

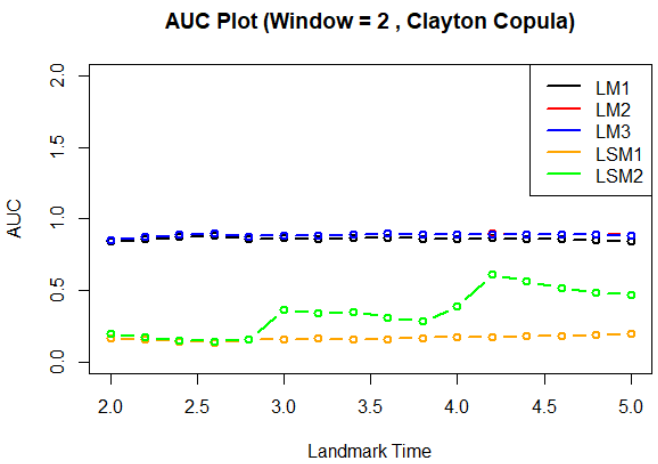


圖附-88：標誌時間點 4 年下預測未來 10 年存活曲線圖

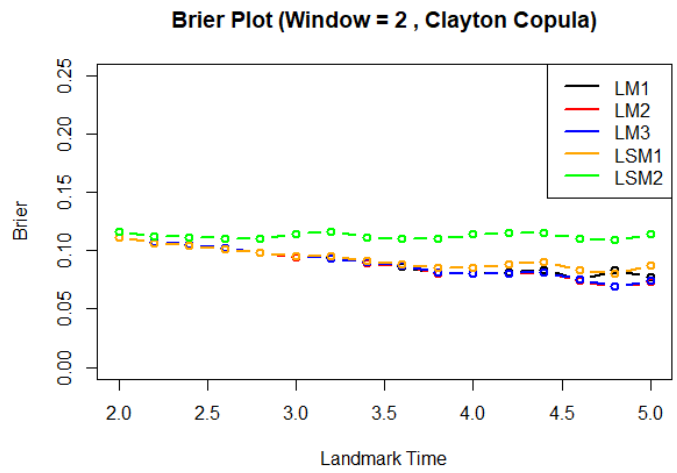


圖附-89：標誌時間點 5 年下預測未來 10 年存活曲線圖

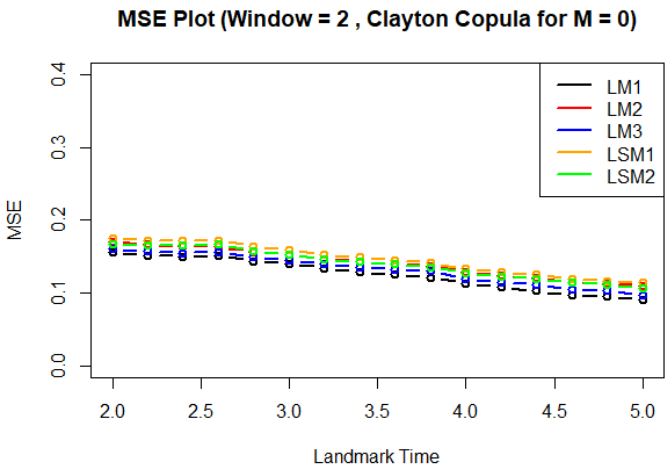
doi:10.6342/NTU201903431



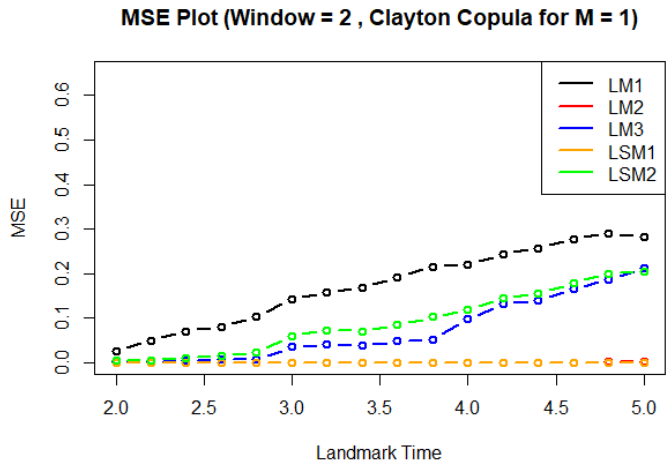
圖附-90：各特徵點模式預測 2 年後 AUC 表現



圖附-91：各特徵點模式預測 2 年後 Brier 分數表現

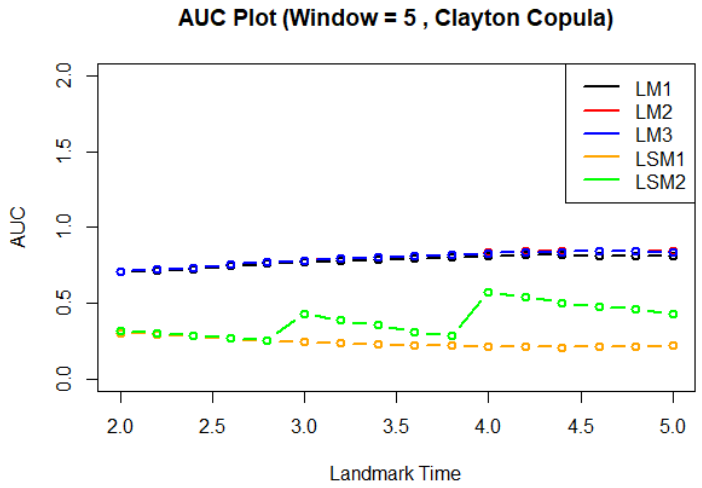


圖附-92：各特徵點模式預測未發生標誌事件個案 2 年後存活機率均方差

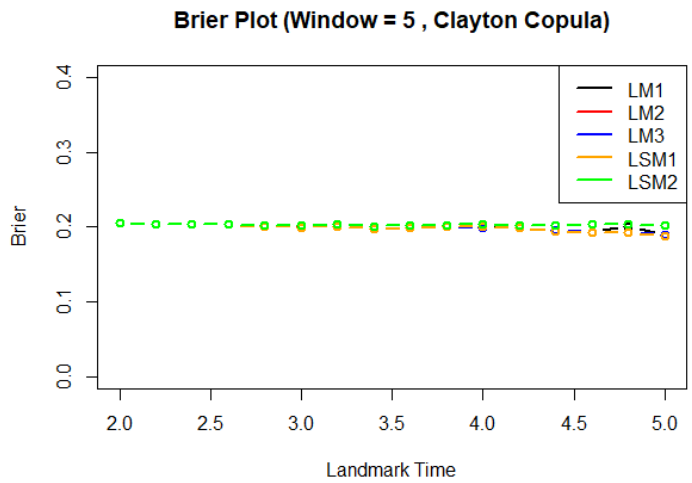


圖附-93：各特徵點模式預測已發生標誌事件個案 2 年後存活機率均方差

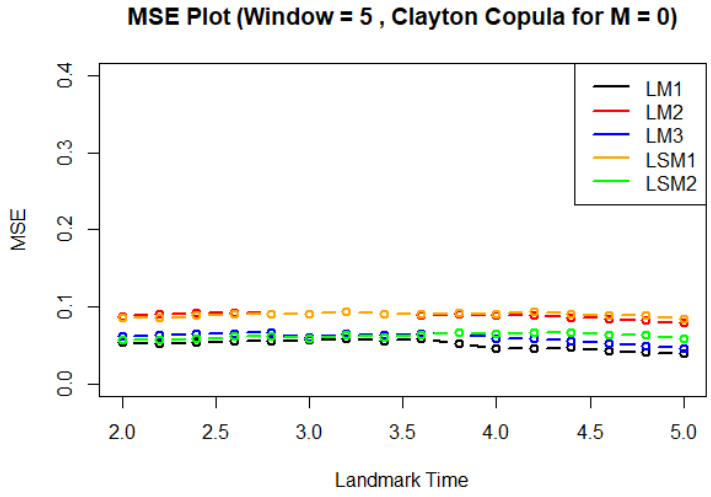
doi:10.6342/NTU201903431



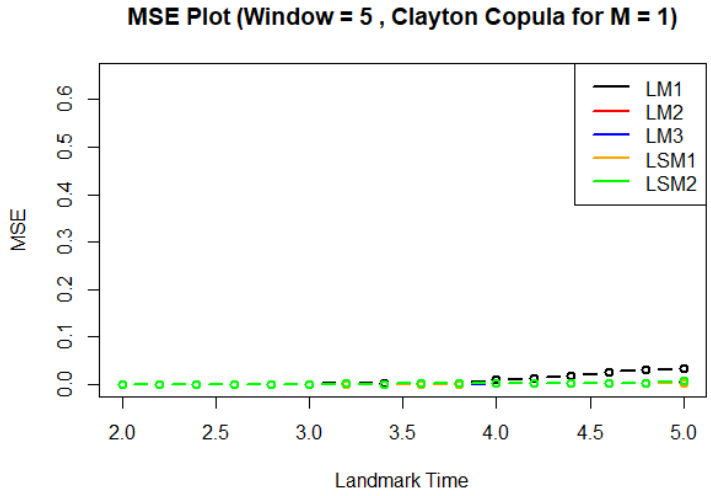
圖附-94：各特徵點模式預測 5 年後 AUC 表現



圖附-95：各特徵點模式預測 5 年後 Brier 分數表現

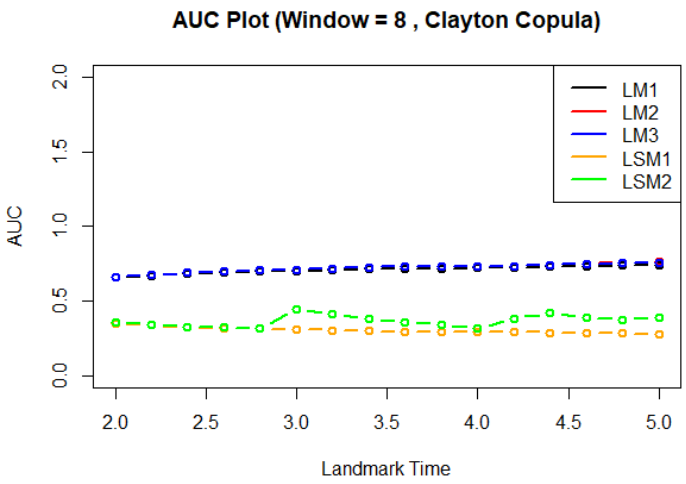


圖附-96：各特徵點模式預測未發生標誌事件個案 5 年後存活機率均方差

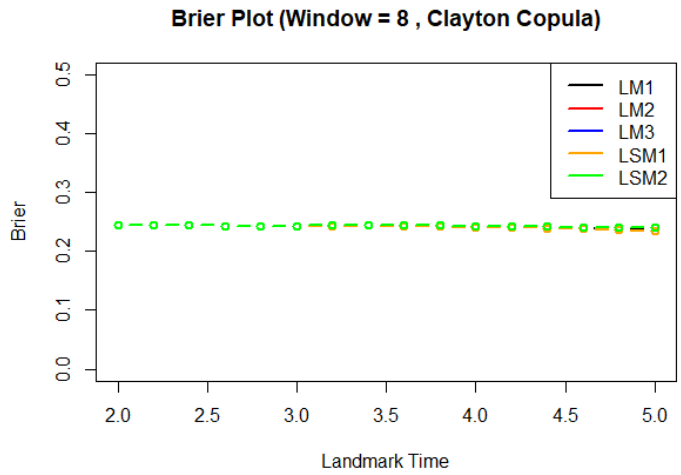


圖附-97：各特徵點模式預測已發生標誌事件個案 5 年後存活機率均方差

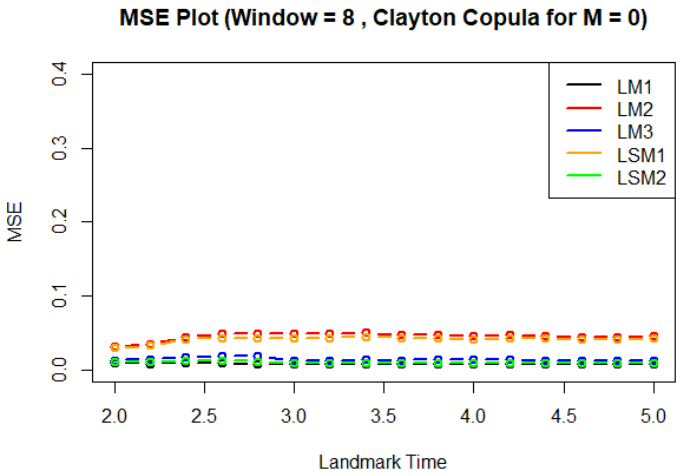
doi:10.6342/NTU201903431



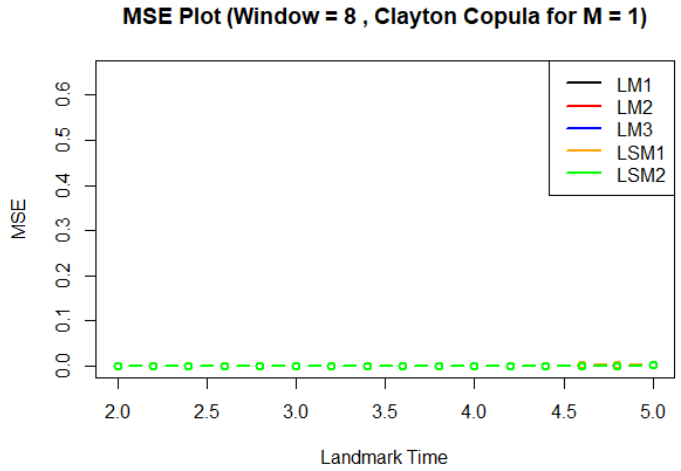
圖附-98：各特徵點模式預測 8 年後 AUC 表現



圖附-99：各特徵點模式預測 8 年後 Brier 分數表現

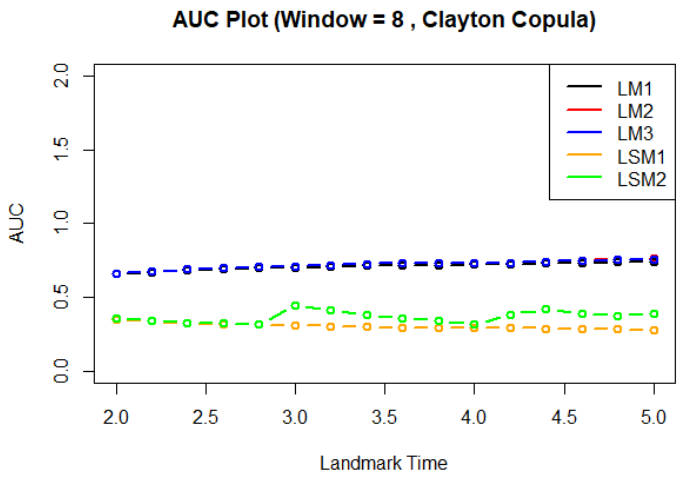


圖附-100：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差

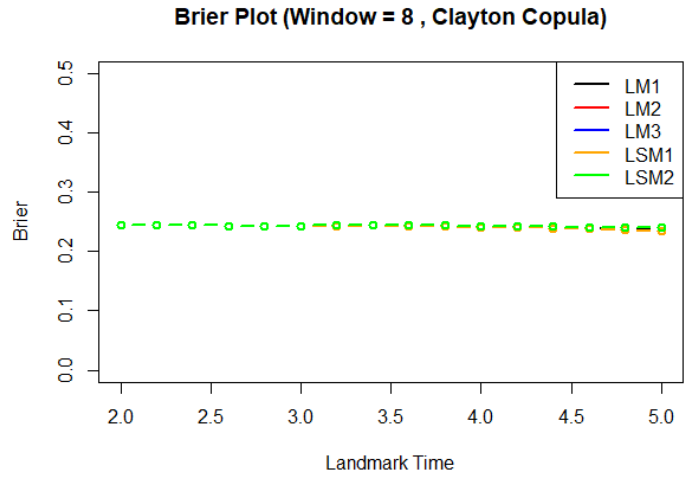


圖附-101：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

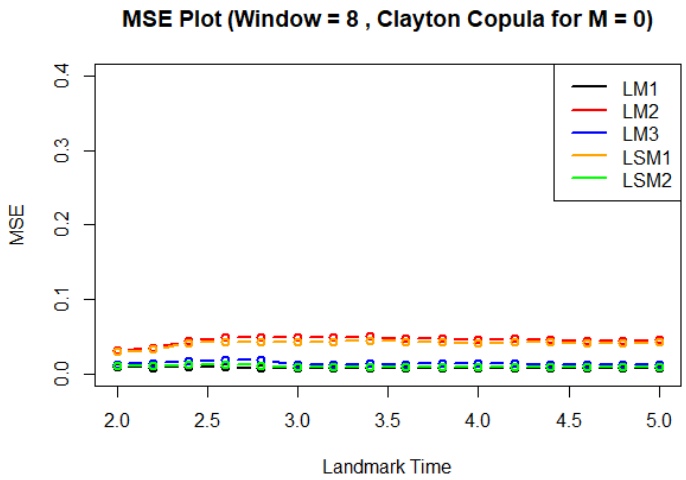
doi:10.6342/NTU201903431



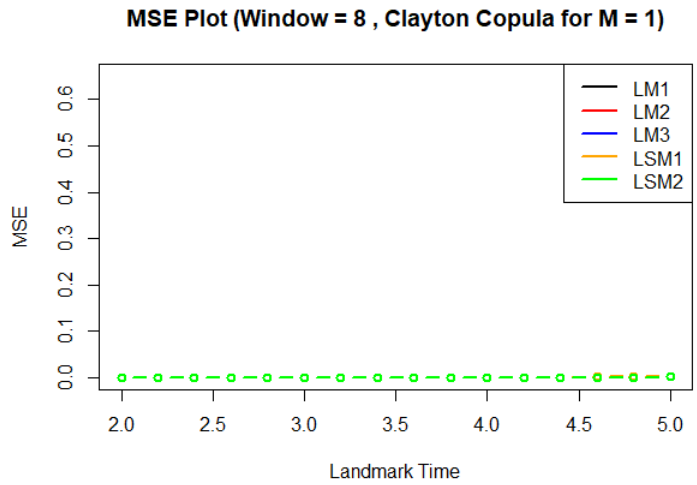
圖附-102：各特徵點模式預測 8 年後 AUC 表現



圖附-103：各特徵點模式預測 8 年後 Brier 分數表現



圖附-104：各特徵點模式預測未發生標誌事件個案 8 年後存活機率均方差



圖附-105：各特徵點模式預測已發生標誌事件個案 8 年後存活機率均方差

doi:10.6342/NTU201903431