106 學年度學生專題報告競賽

超聲波心動圖
預測心臟病患者短期死亡風險

國立台北大學統計學系 陳庭安
國立台北大學統計學系 張庭瑋
國立台北大學統計學系 藍睿豪

國立台北大學統計學系 黃佳慧助理教授 指導

報告日期 2018.06.01

目錄

圖目錄

表目錄

1 研究背景

根據我國衛生福利部(Ministry of Health and Welfare)於 2016 年公告的國人十大死因統計，心臟疾病死亡人數位居第二，僅次於癌症(惡性腫瘤) [1] 。另外，世界衛生組織(World Health Organization)於同年 2016 年的數據顯示，全球年死亡總數約 5600 萬人，其中高達 30%的人死於心血管疾病，為全球死亡人數最高的疾病[2]。

2 研究動機與目的

2.1 研究動機

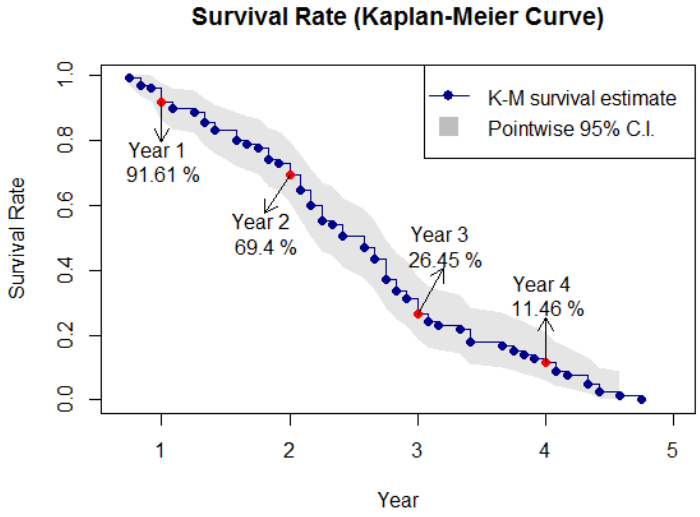
心臟疾病的不可忽視除了在於每年死亡人數居高不下，還有高復發率。罹患心臟病以前預防當然重要，一旦罹患， 防止復發的「二度預防」更是不容輕視，術後追蹤尤為重要。

目前常用於檢測心臟狀況的醫療設備，超聲波心動圖(Echocardiogram)，是一種運用超聲波來顯示心臟結構的檢查方法，具有無侵入性、無放射性與即時提供影像等優點，可用於檢測心臟腔室的大小、肌肉的厚薄、收縮舒張功能的好壞等[3]。

2.2 研究目的

圖一為資料中 132 位患者追蹤期間的存活率，術後一年內情況還算穩定，年底存活率逾 90%。第二年年底存活率不到 70%左右，且與術後 2 年內存活率的跌幅相比，2 年以後有更大的下降幅度，復發情形更為嚴重，因此，若能提前評估患者在邁入第三年前的復發風險，並提早做術後保養，期望能避免心臟再次損傷，以做到二次預防。

故此研究欲透過超聲波心動圖量化患者術後心臟的狀況，進而運用該數值有效預測患者兩年內復發死亡風險，並探究復發風險因子的影響程度，以協助醫師及早做出適當醫療處置，預防疾病的復發。



▲ 圖 1 存活率

3 資料說明

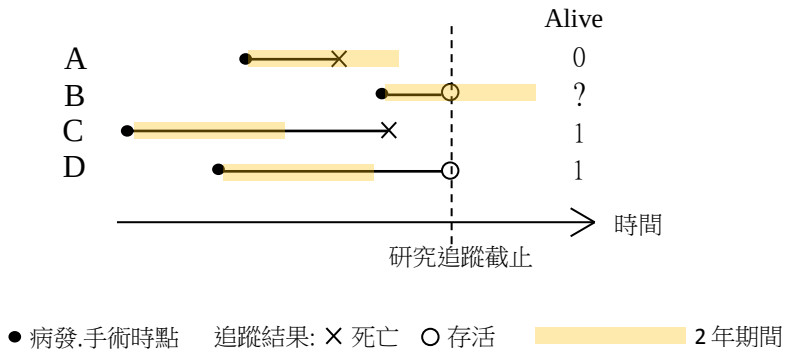
3.1 資料來源

資料取自 UCI 數據平台的超聲波心動圖資料集(Echocardiogram Dataset)，為美國邁阿密的心臟病學專家 Dr. Evlin Kinney，蒐集自 132 位心臟疾病患者，其術後追蹤的結果 [4]。

3.2 變數說明

一、患者資料

1. Age at heart attack: 患者病發年齡，分布年齡 55~65 歲。
2. Survival: 術後追蹤時間。
3. Still Alive: 追蹤結果，死亡或存活。
4. Alive: 術後兩年內死亡或存活，屬新增變數。
   為達成第 2.2 節提及的研究目的，藉由前兩個變數，術後追蹤時間與追蹤結果，將資料中患者分為四類如圖 2。A 類病患不到兩年死亡，Alive 記作 0； C、D 類病患術後兩年仍存活，Alive 記作 1；B 類病患術後追蹤不到兩年存活，滿兩年時是否存活未可知，且追蹤時間長度也不足，故此類病患不在此研究探討範圍內，扣除此類患者，資料樣本數為 90。

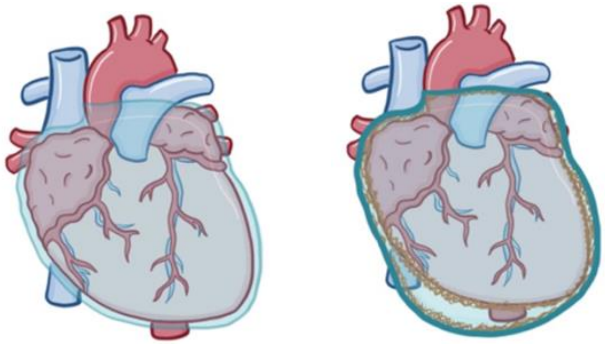


▲ 圖 2 患者術後追蹤結果類型

二、心臟狀況

1. PE(Pericardial effusion）:是否有過多心包積液。

在心臟外面有兩層薄膜覆蓋，醫學上稱之為心包，兩層心包間的間隙含有少量液體便是心包積液。少量積水能保護心臟，過多液體積聚會使心包不易伸展，進而使心臟受壓迫，嚴重恐危及生命[5]。



▲ 圖 3 心包積液示意圖-正常(左)、異常(右)

2. FS（Fractional shortening）: 縮短分率，衡量左心室的收縮程度。
   FS= (LVEDD－LVESD)/LVEDD×100%
   LVEDD(left ventricular end-diastolic dimension): 左心室舒張末期內徑
   LVESD(left ventricular end-systolic dimension): 左心室收縮末期內徑
3. EPSS(E-point septal separation): 舒張早期二尖瓣前葉與室間隔之間的距離，與左心室擴張、收縮力有關。
4. LVDD (left ventricular end-diastolic dimension) : 左心室舒張末期內徑。
5. Wall motion index: 左心室壁運動指標，衡量心跳強弱程度。

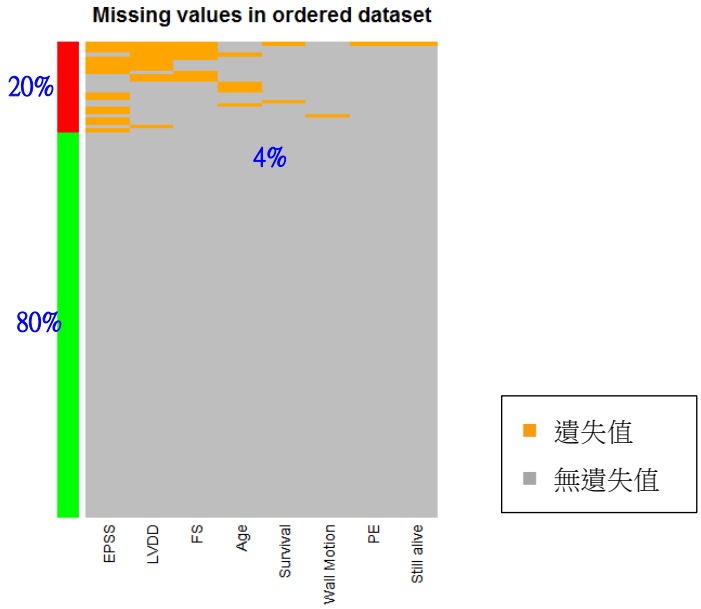
上述衡量心臟狀的指標多與左心室有關，原因是左心室負責體循環，負責把血送到全身，承載的血壓較大，故當心臟病發時，左心室是最容易觀察到異常的部位[6]。

4 資料處理

4.1 遺失值插補

4.1.1 遺失值分布

由圖 4，資料中共有 4%遺失值，有 20%的資料有部分欄位的值有缺失，80%無任何欄位有缺失。



▲ 圖 4 排序後資料中遺失值分布

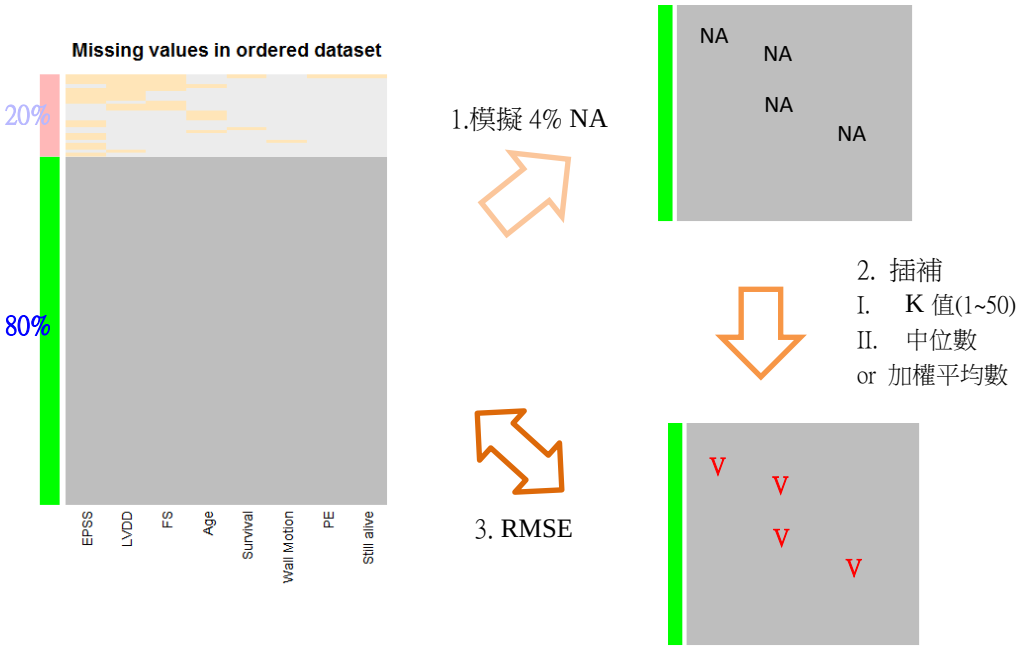
4.1.2 KNN-插補法

此資料採 2003 年提出的 K 近鄰插補法(KNN Imputation)插補遺失值[7]；藉由與遺失資料性質最為相似的 K 筆資料，去推估遺失的資訊。

主要影響插補結果準確性的因素包含參考資料的數目、相似的定義、遺失值的估計方式。若參考資料數目 K 值取過大，參考對象太多，將模糊化資訊的特

徵；若 K 值取太小，則易受到雜訊影響，或因過度依賴樣本資料而失真。資料間相似度的衡量會影響到取到哪些相似的參考對象的數值，進而影響到插補結果。此研究對於資料間相似性的定義為先將資料集標準化之後，再以歐幾里德距離計算其相似程度。

插補流程如圖 5，為了採取適當的 K 值，在此以 80%無遺失值的資料，隨機模擬出具有 4%遺失值的資料，經標準化後計算資料間的歐式距離，對於每個遺失資訊取 K 個與該筆資料最相似的對象，K 值在此研究分別取 1 至 50。若插補對象屬於類別資料，則以 K 個參考對象的數值之眾數補之；若屬連續型資料，則分別考慮以 K 個參考對象的數值之中位數、加權平均數補之。中位數可避免掉離群值所帶來的雜訊影響，而加權平均法在此給予的權重大小算法定義為 exp (-Euclidean distance)，取決於相似度，即標準化後的歐式距離，越是與有遺失值的資料相似的參考對象，給予其值越大的權重。距離與權重的反比關係，使得在此定義的權重公式為一遞減函式。當插補完模擬資料後的結果，與原來完整的資料比較，以 RMSE(Root mean square error)衡量插補的誤差，此研究採用誤差最小的 K 值與估計遺失資訊的算法進行遺失值插補。

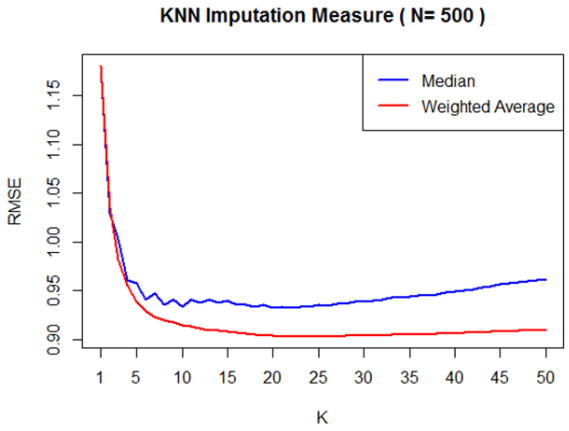


▲ 圖 5 遺失值插補流程

圖 6 為重複 500 次模擬遺失值資料並進行上述插補流程的結果，藍線、紅線分別是以 K 個參考對象值的中位數、加權平均數插補而得的估計誤差。

整體而言加權平均法的誤差明顯較中位數算法來的要小，因此選擇加權平均算法。在 K 值小於 7 時，誤差隨著參考對象個數增加而急速下降，表示在此階

段插補的正確資訊不斷上升；K 值取到 7 之後緩慢下降、趨於平緩，代表能獲得的正確資訊幾乎可以由最接近的 7 個資料點提供，因而在此將取 K 值為 7，且以考慮相似度的加權平均法估計整份資料的遺失值。



▲ 圖 6 模擬遺失值的插補結果

4.2 資料探索與變數轉換

4.2.1 心包積水與短期存活情況的關係

表 1 為有無心包積水(PE)現象患者短期內存活情形。有無心包積水的患者短期內死亡的比例分別為 31%與 27%，此資料在兩獨立母體比例差的檢定下，無足夠證據支持有心包積水患者的死亡比例高於無積水的患者。

▼ 表 1 有無心包積水現象患者存活情形

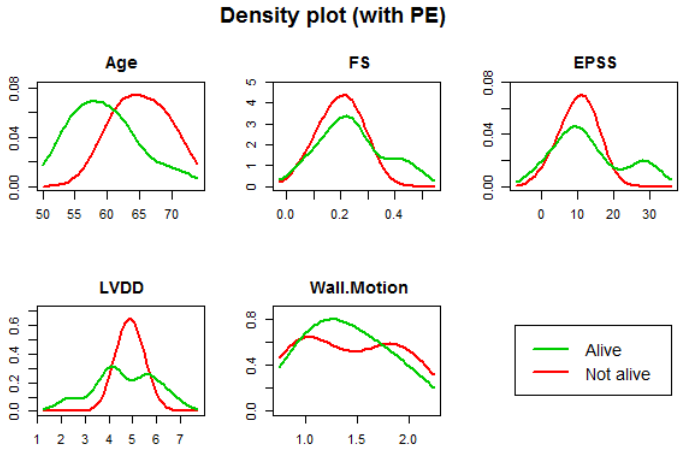
	有心包積水	無心包積水
存活	9 (69%)	56 (73%)
死亡	4 (31%)	21 (27%)

4.2.2 其他變數與短期存活情況的關係

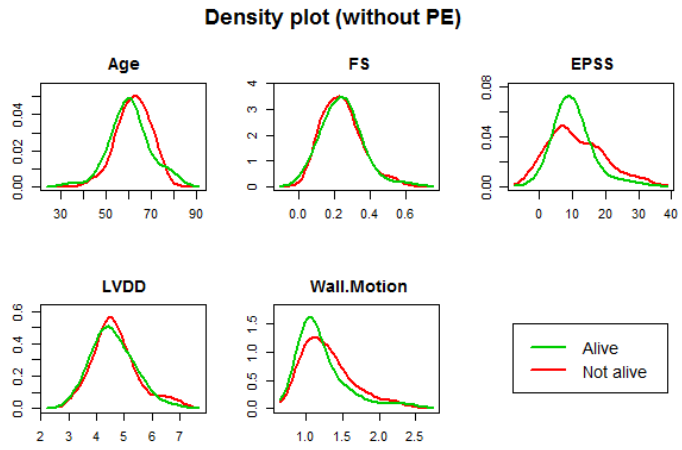
圖 7 及圖 8 分別是考慮患者心包有無積水現象情況下，存活與否的數值分布。心包有積水現象、存活與否的患者，其年齡、心臟狀況差異較無心包積水現象的

患者大，故預期心包有積水的患者判斷短期死亡風險的可能較容易，因而在後續資料探索、預測分類時將是否有心包積水現象的患者分開探討。

由圖 7 知心包有積水的患者，短期死亡的患者群，整體發病時的年齡(Age)、心室壁運動指標(Wall.Motion)較大；左心室收縮程度(FS)較小；舒張早期二尖瓣前葉與室間隔距離(EPSS)與舒張末期左心室直徑(LVDD)較集中於中間值。無心包積水現象的患者，短期死亡群的舒張早期二尖瓣前葉與室間隔距離(EPSS) 較集中於中間值、心室壁運動指標(Wall.Motion)則較大，與心包有積水患者的結果類似，其他數值並無明顯差異(圖 8)。



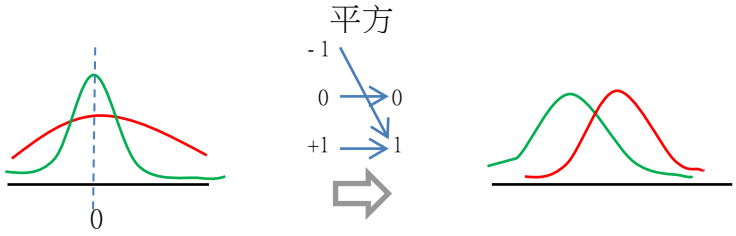
▲ 圖 7 心包積水情況下存活與否的患者各項數值分布圖



▲ 圖 8 無心包積水情況下存活與否的患者各項數值分布圖

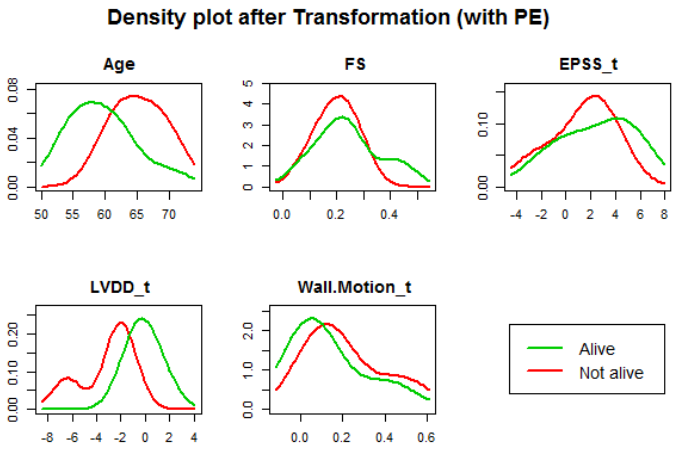
4.2.3 變數轉換

前一小節有提到圖 7 及圖 8 中舒張早期二尖瓣前葉與室間隔距離(EPSS)，與圖 7 中舒張末期左心室直徑(LVDD)值較集中於中間區段，且無明顯左偏或右偏，類似圖 9 左方圖形。為了使之後分類模型(第 5 章)能夠更有效地預測病患死亡的風險，在此考慮做變數轉換，使得兩群數值分布如圖 9 右圖。作法為先將原分布平移至中軸為 0 處，再藉由平方轉換，使得兩端數值映射到大值，中間區段的值則對應至相對小的值。

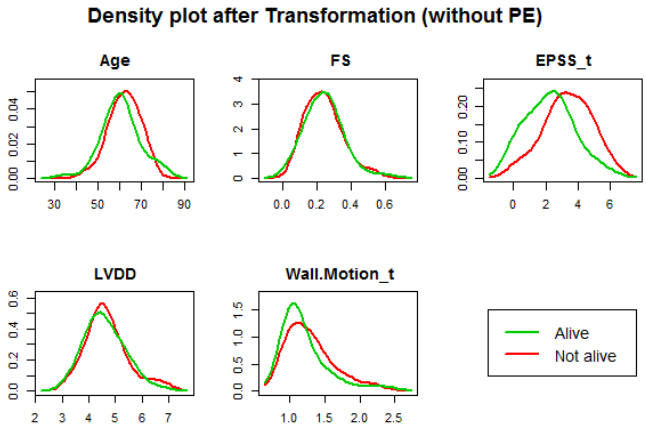


▲ 圖 9 變數轉換概念

圖 7、圖 8 利用平方轉換後，也對部分右偏的變數做對數轉換，轉換後的分布圖如圖 10 及圖 11，經過轉換的變數名稱後方加上「_t」。



▲ 圖 10 轉換後有心包積水情況下存活與否分布圖



▲ 圖 11 轉換後無心包積水情況下存活與否分布圖

4.2.4 解釋變數間的關係與離群值

忽略掉離群值後，轉換後有無心包積水的資料集中，年齡(Age)、左心室收縮程度(FS)、舒張早期二尖瓣前葉與室間隔距離(EPSS)、舒張末期左心室直徑(LVDD)與心室壁運動指標(Wall.Motion)彼此之間無高度相關或完全相關性，因而建模時無須考慮可能出現共線性的問題。

5 資料分析

5.1 變數選擇

分別計算變數轉換前後、有無心包積水的資料集中，每個解釋變數對短期內存活與否的 Wilks' Lambda(式一)，分子為群內變異，分母為群內變異與群間變異的總和，計算結果排序後如表 2。選入具有良好區別不同類別能力的變數，即群間差異越大或群內差異越小、Wilks' Lambda 越小的變數到模型中。

$$\text{Wilks' Lambda} = \text{SSwithin} / (\text{SSwithin} + \text{SSbetween}) \quad \text{(式一)}$$

▼ 表 2 變數排序後 Wilks' Lambda

變數轉換前/心包積水					
Wilks' Lambda	Age	FS	EPSS	LVDD	Wall.Motion
	0.693	0.945	0.956	0.985	0.998
變數轉換前/無心包積水					
Wilks' Lambda	Wall.Motion	Age	LVDD	EPSS	FS
	0.961	0.994	0.998	0.998	1.000
變數轉換後/心包積水					
Wilks' Lambda	LVDD	Age	FS	Wall.Motion	EPSS
	0.567	0.693	0.945	0.945	0.946
變數轉換後/無心包積水					
Wilks' Lambda	EPSS	Wall.Motion	Age	LVDD	FS
	0.831	0.959	0.996	0.996	1.000

5.2 模型選擇

此研究分別採線性判別分析 LDA(Linear Discriminant Analysis)與支持向量機 SVM(Support Vector Machine)分類方法預測短期死亡風險，並以交互驗證 Leave-one-out CV(Cross Validation)方法訓練、測試資料，提高模型可信度。

依前一小節 Wilks' Lambda 小到大排序後的變數，一一加入到分類模型中訓練、測試。考慮變數轉換前後、不同變數選取、分類模型的預測結果準確率如表 3 及表 4。由此二個表格，對於有心包積水的患者最好的預測模型為，經變數轉換後，取 Wilks' Lambda 最小的前三個變數作 LDA，預測準確率近 85%；而對於無心包積水的患者最好的預測模型則是，經變數轉換後，取 Wilks' Lambda 最小的前兩個變數作 LDA，預測準確率達 72%。

▼ 表 3 預測有心包積水現象患者存活準確率

預測準確率(%)				
心包有積水	變數轉換前		變數轉換後	
變數個數	LDA	SVM	LDA	SVM
1	69	62	69	62
2	77	46	77	69
3	69	62	85	62
4	69	62	77	62
5	53	62	69	62

▼ 表 4 預測無心包積水現象患者存活準確率

預測準確率(%)				
心包無積水	變數轉換前		變數轉換後	
變數個數	LDA	SVM	LDA	SVM
1	69	28	70	65
2	69	71	72	64
3	69	28	69	65
4	69	28	68	68
5	71	28	66	69

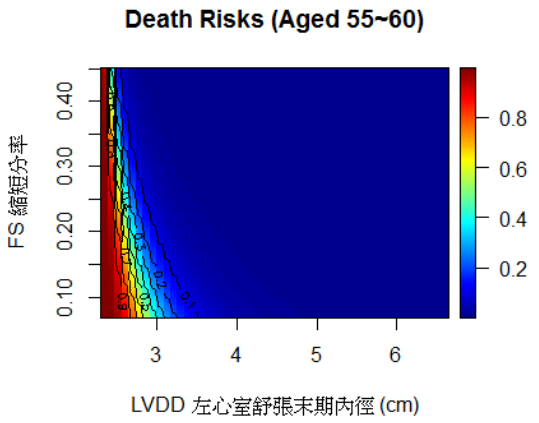
5.3 模型預測結果

圖 12 至 14 為 LDA 藉由 LVDD 左心室舒張末期內徑(cm)、Age 發病年齡、FS 縮短分率，預測不同年齡階段且心包有積水的患者短期死亡的風險的等值線圖。橫軸左心室舒張末期內徑衡量心臟舒張能力，值小表示舒張不完全；縱軸收縮程度顧名思義為衡量心臟的收縮能力，值小則表示收縮不完全。

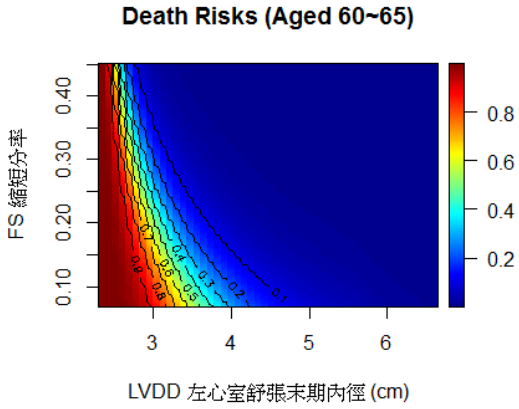
當患者術後心臟舒張、收縮力越不足，無論發病年齡大小，短期復發致死的風險都越高。當考慮患者病發時的年齡，即使同樣的心臟收縮力，越年長的患者越難以再次承受負荷與損傷，短期死亡風險較高。

圖 15 則是 LDA 藉由 EPSS 舒張早期二尖瓣前葉與室間隔距離(mm)、Wall Motion 左心室壁運動指標，預測無心包積水的患者短期死亡風險的等值線圖，

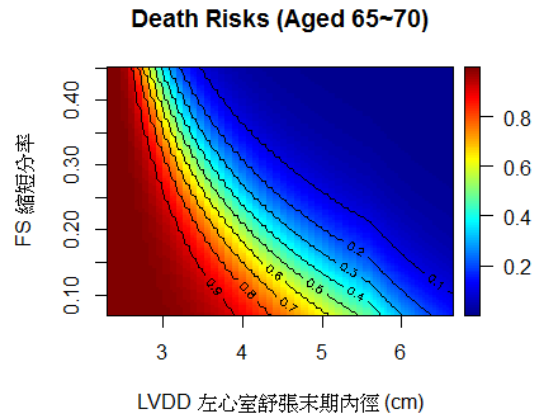
模型所取的變數與心包有積水患者的模型的截然不同。圖中橫軸二尖瓣前葉與室間隔距離過大或過小，隱含左室擴張和收縮力減退；縱軸室壁運動指標衡量心臟跳動強弱，值為 1 左右仍屬正常，值越大，表示跳動程度減弱。當左心室擴張和收縮力減退、心臟跳動程度減弱，患者死亡風險高。



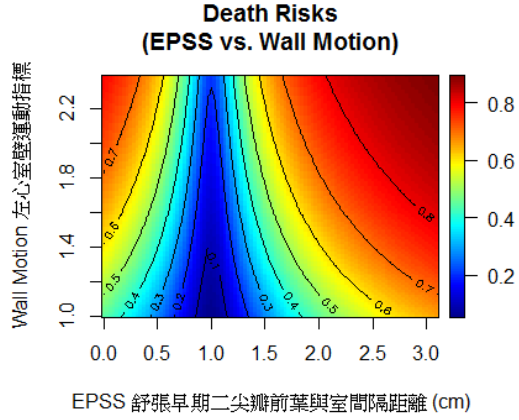
▲ 圖 12 心包有積水患者死亡風險(一)



▲ 圖 13 心包有積水患者死亡風險(二)



▲ 圖 14 心包有積水患者死亡風險預測(三)



▲ 圖 15 心包無積水患者死亡風險預測

6 結論

適合心包有無積水患者的預測指標有所不同，但同樣都考慮了左心室的擴張、收縮能力，另外風險因子還分別包含患者發病年齡與心臟跳動強度。

手術後心臟收縮力弱或心臟跳動強度不夠的患者，心臟較難以承載高壓高速

的血流，若沒有適當保養，復發死亡的可能性很高。另外年長的患者身體狀況較不堪負荷，所以儘管術後心臟收縮力與壯年患者相同，心臟遇到損傷後的結果會比壯年人嚴重。

因此，年長者或心臟功能原本就較弱的族群平時就要特別注意保養，若心臟曾受過損傷或手術，術後的保養更加不得輕忽，以免心臟負荷過重，導致疾病復發。

7 參考資料與文獻

1 Ministry of Health and Welfare of the Republic of China, https://www.mohw.gov.tw/cp-16-33598-1.html

2 World Health Organization, http://www.who.int/cardiovascular_diseases/zh/

3 陳偉光、王泰鴻等(2007)。深知你心: 心臟檢查面面觀。香港。知出版。

4 UCI Machine Learning Repository

5 健康猴: 心包積液的症狀和治療方法, http://www.jiankanghou.com/jibing/33368.html

6 中亞健康網: 心臟血管, https://www.ca2-health.com/cvd

7 Batista , G. E. & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Journal of Applied Artificial Intelligence*, 17:519-533

8 分工

- 陳庭安—研究動機與目的，資料處理，分析，結論，參考資料與文獻蒐集
- 張庭瑋—研究背景，研究動機與目的，資料說明，參考資料與文獻蒐集
- 藍睿豪—研究動機與目的，分析方法，參考資料與文獻蒐集