

東海大學統計研究所

碩士論文

雙樣本右設限存活資料之
數據驅動機動檢定

研究生：陳志維
指導教授：張玉媚博士

中華民國一百年七月

東海大學碩士班研究生

論文口試委員審定書

統計學系碩士班陳志維君所提之論文

雙樣本右設限存活資料之數據驅動機動檢定

經本委員會審議，認為符合碩士資格標準。

論文口試委員召集人 陳春樹 (簽章)

委員 張玄媚

俞一君

中華民國 100 年 7 月 14 日

致謝

終於，我的學習過程也來到這最後的階段了。想當初，懵懵懂懂的我，對學習意興闌珊，對於未來，一點想法也沒有。所幸，遇見了張玉媚博士，輕鬆愉快的上課氣氛及深入淺出的教學，使我對鑽研統計的興趣越來越濃烈，於是決定攻讀碩士學位。

就讀研究所期間，感謝啟蒙我的張老師願意擔任我的指導教授，並給了我研究方向。在課業上有疑惑時，老師總是不厭其煩地為我指點迷津，讓我能釐清問題、一步步完成研究。在研究之餘，老師的言行給了我很多啟發，如果我能有正確的學習態度及待人接物之道，那一定是老師給予的影響，志維萬分感激老師。

論文口試期間，承蒙陳春樹博士與俞一唐博士在百忙之中，給予詳細地審閱與批示，並提供許多寶貴的意見，使論文的內容更為豐富完整，在此致上最誠摯的謝意。

在沒有扎實的基礎之下，論文的完成，是沒辦法這麼順利的，所以我得感謝所有系上所有的老師。謝謝老師們為了傳授統計相關的專業知識給我們而努力、詳細地教學，謝謝您們。

在研究所修業的兩年，時間過得很緊湊，生活上許多大大小小的事情無法處理得面面俱到。感謝系上的佳玲、雅文、淑姿、如君與小津學姊幫忙協助我們課業上與行政上的許多事務，有您們在，我們少了一點雜務的紛擾，也就能多出一點寶貴的研究時間。

即將拿到畢業證書，腦子裡浮現的是研一的時候，大家擠在悶熱的研究室裡討論功課、趕作業，還有大家聚在一起聊著似乎無關緊要的事，很多很多，都是以後值得回味的一頁。感謝炳然在程式撰寫以及統計觀念方面教了我許多；感謝兩青、莉芹、咨瑋與怡穎在乏味的研究生活，總能摩擦出一些歡樂的火花；感謝淑玲與維珍，為我帶來了一些思維的刺激；感謝柏尚在研究上的一起努力與討論；感謝所有的同學，我們在研究所經歷的點點滴滴，一定能讓未來的我們更有解決困難的能力。

啟志學長，感謝你對我的鼓勵，我會對自己更有自信；宋仔、阿尼奇與阿如，感謝你們往往在我忙於研究、廢寢忘食之際，拉我一把，到外頭呼吸一下大自然的空氣，讓我的腦子能有空間整理整理思緒；小兔，謝謝你給家裡帶來的歡笑，我一定會更加疼你的。

最後，最要感謝的是我的奶奶、爸爸與媽媽，有您們無條件地付出、疼愛與栽培，才能有今天的我，希望我可以是您們的榮耀以及驕傲。謹以這兩年來的成果，獻給我最愛的奶奶、爸爸與媽媽。

陳志維 謹致於

東海大學統計學系研究所

中華民國一百年七月

摘要

欲檢定兩組存活分配是否相同，加權對數秩 (Weighted Log-Rank) 檢定與加權 KM (Weighted Kaplan-Meier) 檢定是被使用得最廣泛的兩種方法。因為這兩種檢定在各種的狀況下，各有不同的優點，而我們很難得知真實狀況為何，所以事先決定選用的檢定之檢定力可能不會最大。因此，為了將這兩種檢定的優點結合，本文著重於同時使用這兩種檢定的機動檢定。

我們延續 Chi and Tsai (2001) 的想法，考慮使用加權對數秩統計量及加權 KM 統計量的線性組合之機動檢定，選取使其達到最大的權重，再利用排序技巧造其抽樣分配，並與 Chi and Tsai 所提給定相同權重線性組合的檢定比較。透過模擬研究，可以說明我們所提出的方法具有優越性。最後，我們將此檢定方法應用於實際資料。

關鍵字：數據驅動；線性組合的檢定；排序技巧；右設限資料；機動檢定；加權

KM 檢定；加權對數秩檢定

Abstract

The weighted log-rank (WLR) tests and the weighted Kaplan-Meier (WKM) test are the most common used tests for testing the equality of two survival distribution. Since each test has different advantages against various alternatives and it's hard to know the real alternative, the pre-selected test can't gain more power. Therefore, in order to combine the advantages of these two classes of tests, a versatile test based on WLR tests and WKM test is then proposed. We select the appropriate weights by maximizing the versatile test statistic, and use the permutation method to construct the sampling distribution. Some numerical experiments are performed for illustrating the superiority of the proposed method and then the proposed testing procedure is applied to two real data sets.

Key Words: Data-driven; Linear combination test; Permutation method; Right-censored data; Versatile test; Weighted Kaplan-Meier test ; Weighted Log-Rank tests

目錄

第一章、緒論	1
第二章、加權對數秩檢定與加權 KM 檢定	4
第一節、加權對數秩檢定	4
第二節、加權 KM 檢定	5
第三章、數據驅動機動檢定	7
第四章、研究模擬	9
第五章、實例分析	11
第六章、結論與未來研究方向	13
參考文獻	14
附錄	16

圖表目錄

表一、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 30$ 及 $\alpha = 0.05$ 下，誤差率及 檢定力之估計·····	16
表二、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 50$ 及 $\alpha = 0.05$ 下，誤差率及 檢定力之估計·····	17
表三、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 80$ 及 $\alpha = 0.05$ 下，誤差率及 檢定力之估計·····	18
表四、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 30$ 及 $\alpha = 0.05$ 下，數據驅動 機動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差·····	19
表五、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 50$ 及 $\alpha = 0.05$ 下，數據驅動 機動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差·····	19
表六、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 80$ 及 $\alpha = 0.05$ 下，數據驅動 機動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差·····	19
表七、實例分析，其各種檢定之檢定統計量及單邊 p 值·····	20
圖一、各個設置下的風險函數圖·····	21
圖二、腎臟手術資料的存活函數估計圖及累積風險估計圖·····	22
圖三、燒燙傷消毒資料的存活函數估計圖及累積風險估計圖·····	22

第一章、緒論

在臨床試驗中，主要目標是比較各個組別的存活時間，藉以評斷藥劑的效用。因此，檢定兩組存活分配是否相同是一重要問題。

最常被用來檢定兩組存活分配是否相同的檢定統計量為對數秩 (Log-Rank, 記作 LR) (Mantel, 1966) 與 Peto-Prentice-Wilcoxon (記作 PPW) (Gehan, 1965; Peto and Peto, 1972; Prentice; 1978)，其中對數秩檢定對檢定風險是否成比例具有局部最大檢定力，而 PPW 檢定較能檢定出早期風險差異。此外，加權對數秩檢定 (Weighted Log-Rank, 記作 WLR) 依據兩組風險估計值的差異，給予不同權重加總而得到，因此常被使用來檢定相關的問題。但不幸地，當兩組的風險呈現交叉時，WLR 統計量對於檢定存活函數差異不夠靈敏，基於此缺失，Pepe and Fleming (1989) 提出了加權 KM (Weighted Kaplan-Meier, 記作 WKM) 統計量，此檢定統計量是依據兩組 Kaplan-Meier (1958) 存活函數估計值的差異，給予不同權重加總而得，用以檢定兩存活函數的差異。結果顯示，分別在風險成比例、早期風險差異時，WKM 檢定是可以和 LR 檢定與 PPW 檢定抗衡的，而當兩組的風險呈現交叉時，WKM 檢定的表現比 WLR 檢定優異。

此外，權重給得不同，WLR 檢定有各種不同優點，藉以應付不同的對立假設。不過，實際上，真實的對立假設很難得知，權重的選擇就難以決定。為了在大多數狀況下，都能保持更好的檢定力，Lee (1996) 提出了一些折衷的策略，其檢定統計量是依據 WLR 檢定族的線性組合或者選取 WLR 檢定族中的最大值。Shen and Cai (2001) 研究了依據選取 WKM 檢定族中的最大值的機動檢定。最

近，Yang and Prentice (2010) 提出一個以對資料配適 Yang and Prentice (2005) 所提出的模型，藉此決定 WLR 檢定族之權重的想法。

因為 WLR 檢定與 WKM 檢定在各種狀況下，各有不同的優點，因此 Chi and Tsai (2001) 提出一機動檢定，其檢定統計量是依據了兩組獨立右設限資料的 WLR 統計量與 WKM 統計量的線性組合或 WLR 統計量與 WKM 統計量的最大值。Chi and Tsai (2001) 指出，在不同的對立假設下，這些機動檢定比 Lee (1996) 提出依據線性組合之檢定穩健得多。然而，Chi and Tsai (2001) 將 WLR 統計量與 WKM 統計量以相同的權重合併在一起，即各給定 0.5 與 0.5，結果在不同的對立假設下，此方法無法經常得到最高的檢定力。在本文，我們將延續著 Chi and Tsai 的想法，並提出一數據驅動機動檢定，由資料自行選取最佳的權重，使得 WLR 統計量與 WKM 統計量的線性組合達到最大，藉此將檢定力達到最大，但其漸進分配已偏離標準常態分配，所以需藉由排序 (Permutation) 技巧造其抽樣分配。

欲進行臨床試驗，需先行擬定完善的計畫書，因此在進行臨床試驗前就已決定使用何種檢定方法。我們很難得知兩個組別真實的存活分配為何，單單使用 WLR 統計量與 WKM 統計量，無法在不同的對立假設下都有好的表現，而我們所提出的數據驅動機動檢定以線性組合的方式將 WLR 統計量與 WKM 統計量合併，其結合了 WLR 統計量與 WKM 統計量各自的優點。不論真實的對立假設為何，我們所提出的方法提供了檢定出兩組存活分配是否有差異的優越準則。

本文於第二章回顧用於檢定兩組獨立右設限資料存活分配是否相同的 WLR

檢定與 WKM 檢定。在第三章介紹我們提出的數據驅動機動檢定，再藉由排序 (Permutation) 技巧造其抽樣分配。第四章則透過模擬方法研究其誤差率與檢定力的表現。在第五章利用實際資料說明所提方法的應用。最後，第六章為結論與未來研究方向。

第二章、加權對數秩檢定與加權 KM 檢定

令 T_{ij} 及 C_{ij} 分別為第 i 組的第 j 個患者的失敗時間及設限時間， $i = 1, 2$ ， $j = 1, \dots, n_i$ 。就獨立受測的資料而言，在右設限之下，觀察到的時間是 X_{ij} ， $X_{ij} = \min\{T_{ij}, C_{ij}\}$ 及 $\delta_{ij} = I\{T_{ij} \leq C_{ij}\}$ ，其中 $I\{E\}$ 是指標函數，表示當 E 事件發生時， $I\{E\}$ 為 1，否則為 0。令 $S_i(t) = P(T_{ij} > t)$ 及 $G_i(t) = P(C_{ij} > t)$ 分別為 T_i 及 C_i 的存活函數。令計數過程 $N_i(t) = \sum_{j=1}^{n_i} I\{X_{ij} \leq t, \delta_{ij} = 1\}$ 為第 i 組在時間點 t 已觀察到的失敗個數及風險過程 $Y_i(t) = \sum_{j=1}^{n_i} I\{X_{ij} \geq t\}$ 為第 i 組在時間點 t 仍涉險的人數。

第一節、加權對數秩檢定

為了檢定兩組存活分配是否相同，虛無假設會設定 $H_0: S_1(t) = S_2(t)$ 在所有的時間點 t ，對立假設則設定為 $H_1: S_1(t) \neq S_2(t)$ 在某些時間點 t ，或者設定為 $H_1: S_1(t) < S_2(t)$ 在某些時間點 t 。最常使用的檢定統計量為 Fleming and Harrington (1991) 所提出的 WLR 檢定統計量，其表示成

$$K_1 = \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \int_0^T \hat{W}(t) \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} \left[\frac{dN_1(t)}{Y_1(t)} - \frac{dN_2(t)}{Y_2(t)} \right],$$

其中 $T = \sup\{t: Y_1(t)Y_2(t) > 0\}$ ， $\hat{W}(t) = \{\hat{S}(t-)\}^p \{1 - \hat{S}(t-)\}^q$ ， $p \geq 0$ 且 $q \geq 0$ ， $\hat{S}(t)$ 是合併兩組樣本的 Kaplan-Meier (1958) 存活函數估計值，而 $\hat{S}(t-)$ 是合併兩組樣本在時間點 t 的前一時間點之 Kaplan-Meier (1958) 存活函數估計值。Fleming and Harrington (1991) 稱 $\hat{W}(t)$ 為 $G^{p,q}$ 族， $G^{0,0}$ 與 $G^{1,0}$ 即為 LR 檢定統計量與 PPW 檢定統計量，其分別在風險成比例和早期風險差異之下，表現得很優異。在虛無假設

之下， K_1 的漸進分配為平均數 0、變異數 σ_{11}^2 的常態分配， σ_{11}^2 的一致性估計值可表示成

$$\hat{\sigma}_{11}^2 = \frac{n_1 + n_2}{n_1 n_2} \int_0^T \hat{W}^2(t) \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)} \left(1 - \frac{\Delta N_1(t) + \Delta N_2(t) - 1}{Y_1(t) + Y_2(t) - 1} \right) \frac{d(N_1(t) + N_2(t))}{Y_1(t) + Y_2(t)}, \quad (1)$$

其中 $\Delta N_i(t) = N_i(t) - N_i(t-)$ ， $i = 1, 2$ 。因此，在顯著水準 α 之下，當 $K_1^* = K_1 / \sqrt{\hat{\sigma}_{11}^2} \geq z_\alpha$ ，我們有足夠的證據拒絕虛無假設，支持第二組的存活率較第一組高，其中 z_α 是標準常態分配右尾面積為 α 所對應的值。

第二節、加權 KM 檢定

因為 WLR 檢定是依據秩而來的，其可能在特定的對立假設之下，對檢定存活時間的差異不夠靈敏。因此，Pepe and Fleming (1989) 提出了 WKM 檢定統計量，其依據兩組的 Kaplan-Meier (1958) 存活函數估計值的差異，給予不同權重加總而得，其可表示成

$$K_2 = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^{T_c} \hat{w}(t) \{ \hat{S}_2(t) - \hat{S}_1(t) \} dt,$$

其中 $T_c = \sup \{ t : \min(\hat{S}_1(t), \hat{S}_2(t), \hat{G}_1(t), \hat{G}_2(t)) > 0 \}$ ， $\hat{S}_i(t)$ 及 $\hat{G}_i(t)$ 分別為第 i 組的失敗時間及設限時間的 Kaplan-Meier (1958) 存活函數估計值， $i = 1, 2$ ，且 $\hat{w}(t)$ 為隨機權重函數，若設限比例過大，其將降低試驗後期兩組的 Kaplan-Meier (1958) 存活函數估計值差異的權重。若此權重函數選取適宜，得到的檢定統計量才會穩

定，Pepe and Fleming (1989) 建議使用 $\hat{w}(t) = \frac{\hat{G}_1(t-)\hat{G}_2(t-)}{p_1\hat{G}_1(t-) + p_2\hat{G}_2(t-)}$ ，其中

$p_i = n_i / (n_1 + n_2)$, $i = 1, 2$ 。與 WLR 檢定相仿，在虛無假設之下， K_2 的漸進分配為平均數 0、變異數 σ_{22}^2 的常態分配， σ_{22}^2 的一致性估計值可表示成

$$\hat{\sigma}_{22}^2 = - \int_0^{T_c} \left\{ \int_t^{T_c} \hat{w}(u) \hat{S}(u) du \right\}^2 \frac{p_1 \hat{G}_1(t-) + p_2 \hat{G}_2(t-)}{\hat{G}_1(t-) \hat{G}_2(t-)} \frac{d\hat{S}(t)}{\hat{S}(t) \hat{S}(t-)}, \quad (2)$$

其中， $\hat{S}(t)$ 是合併兩組樣本的 Kaplan-Meier (1958) 存活函數估計值，而 $\hat{S}(t-)$ 是合併兩組樣本在時間點 t 的前一時間點之 Kaplan-Meier (1958) 存活函數估計值，且 $d\hat{S}(t) = \hat{S}(t) - \hat{S}(t-)$ 。同樣地，在顯著水準 α 之下，當 $K_2^* = K_2 / \sqrt{\hat{\sigma}_{22}^2} \geq z_\alpha$ ，我們有足夠的證據拒絕虛無假設，支持第二組的存活率較第一組高，其中 z_α 是標準常態分配右尾面積為 α 所對應的值。Pepe and Fleming (1989) 同時證明了在風險成比例時，此檢定統計量是能與 LR 檢定統計量抗衡的，而在早期風險差異及風險呈現交叉時，其表現得較優異。不過，若設限比例過大，試驗後期兩組的 Kaplan-Meier (1958) 存活函數估計值差異的權重將會給得較低，使得此檢定統計量對於檢定後期風險差異不夠靈敏。

雖然 WLR 檢定與 WKM 檢定對於各種對立假設有著不同的優點，但在許多狀況下，往往不能每種檢定都有令人滿意的表現。因此，我們提出一數據驅動機動檢定，將於下個章節闡明。

第三章、數據驅動之機動檢定

由於 WLR 檢定與 WKM 檢定對於不同的狀況有各自的優點，因此將它們合併在一起，希望能得到更傑出的檢定統計量是很自然的想法。在本章，我們考慮使用 WLR 統計量與 WKM 統計量的線性組合之機動檢定檢定兩組右設限資料存活分配是否相同。如 Chi and Tsai (2001) 所建議，它們使用相同的權重，也就是各自給定 0.5 與 0.5，以合併後的檢定統計量來進行檢定。然而，在不同的對立假設下，此方法無法總是得到最高的檢定力且不夠順應資料。為了減輕其在各種對立假設下的弱點，我們嘗試提出一種新的檢定方法，由資料自行選取權重。因此，以下將要介紹如何由資料自行選取 WLR 統計量與 WKM 統計量各自的權重及如何利用排序(Permutation) 技巧造其抽樣分配。

我們的想法是依據 Chi and Tsai (2001) 提出的檢定統計量，其表示成

$$\{K(\beta) = \beta K_1^* + (1-\beta)K_2^* : 0 \leq \beta \leq 1\},$$

在虛無假設 $H_0 : S_1(t) = S_2(t)$ 之下，給定 β 的 $K(\beta)$ 之漸進分配為平均數 0、變異數 σ_β^2 的常態分配， σ_β^2 的一致性估計值可表示成

$$\hat{\sigma}_\beta^2 = \beta^2 + (1-\beta)^2 + 2\beta(1-\beta)\hat{\rho},$$

其中 $\hat{\rho} = \hat{\sigma}_{12}^2 / \sqrt{\hat{\sigma}_{11}^2 \hat{\sigma}_{22}^2}$ ， $\hat{\sigma}_{12}^2 = -\int_0^{T_c} \hat{W}^2(t) \left\{ \int_t^{T_c} \hat{w}(u) \hat{S}(u) du \right\} \frac{d\hat{S}(t)}{\hat{S}(t)}$ ，而 $\hat{\sigma}_{11}^2$ 及 $\hat{\sigma}_{22}^2$ 已

事先分別定義在(1)及(2)。因此，在顯著水準 α 之下，當 $K^*(\beta) = K(\beta) / \sqrt{\hat{\sigma}_\beta^2} \geq z_\alpha$ 時，就可以斷定第二組的存活率較第一組高。

欲使用 WLR 統計量與 WKM 統計量的線性組合之機動檢定，首先由資料自

行選取適當的權重 $\hat{\beta} \in [0,1]$ 來得到檢定統計量

$$K^*(\hat{\beta}) = \{\hat{\beta}K_1^* + (1-\hat{\beta})K_2^*\} / \sqrt{\hat{\sigma}_{\hat{\beta}}^2}, \quad (3)$$

在本文， β 的估計值 $\hat{\beta} \equiv \arg \max_{\beta \in [0,1]} \{K^*(\beta)\}$ 。但經此選取權重準則得到的檢定統計量 $K^*(\hat{\beta})$ ，其漸進分配已偏離標準常態分配，因此我們藉由排序技巧造其抽樣分配，才能忠實地檢定兩組存活分配是否相同。

使用排序技巧造 $K^*(\hat{\beta})$ 之抽樣分配的步驟如下：

步驟一、依原始資料，求出 β 的估計值 $\hat{\beta} \equiv \arg \max_{\beta \in [0,1]} \{K^*(\beta)\}$ 及對應的檢定統計量

$$(3)，即 $\hat{\beta}$ 及 $K^*(\hat{\beta})$ 。$$

步驟二、把原始資料的兩個組別合併在一起，並以抽出不放回的方式，隨機將每個患者編號 1 至 $n_1 + n_2$ 。

步驟三、將編號 1 至 n_1 的患者視為第一組、編號 $n_1 + 1$ 至 n_2 的患者視為第二組，

依此新資料，求出在 β 的估計值 $\hat{\beta} \equiv \arg \max_{\beta \in [0,1]} \{K^*(\beta)\}$ 時的檢定統計量

$$(3)，稱它為 $K_l^*(\hat{\beta})$ 。$$

步驟四、重複執行步驟二至三 N 次，將得到 $K_l^*(\hat{\beta})$ ， $l = 1, \dots, N$ ，即造出了 $K^*(\hat{\beta})$

$$\text{之抽樣分配，p 值} = \sum_{l=1}^N I\{K_l^*(\hat{\beta}) > K^*(\hat{\beta})\} / N。$$

因此，在顯著水準 α 之下，p 值 $\leq \alpha$ 時，就可以斷定第二組的存活率較第一組高。

第四章、模擬研究

本章利用模擬研究了解我們所提的數據驅動機動檢定的誤差率及檢定力表現。此模擬研究包含了一個兩組存活分配相同以及四個兩組存活分配不相同的設置。在誤差率的研究，兩組的存活分配皆是從尺度參數 1 的指數分配生成的。在檢定力的研究，兩組的存活分配差異型式包含了韋伯比例風險模型、早期風險差異、晚期風險差異及風險交叉差異。圖一繪出了這幾種情況的風險函數圖，其中，設置 II 至 IV 是利用有著不同風險函數的分段指數分配來生成的。每一個設置，皆在適當的樣本數 ($n_1 = n_2 = 30, 50$ 及 80)，且設限分配皆為均勻分配(0, 2) 下進行模擬。在誤差率研究的設置之下，兩組的設限比率皆為 0.43。在設置 I 至 IV，兩組的設限比率各有所不同，其範圍為 0.24 至 0.73。這些檢定統計量的誤差率和檢定力是在顯著水準 0.05 之下，根據 3000 次的模擬所估計而得。因此，誤差率估計值的標準差為 0.004 ($\approx \sqrt{(0.05 \times 0.95)/3000}$)。

表一至表三列出了各種檢定統計量在 $n_1 = n_2 = 30, 50$ 及 80 下，其誤差率及檢定力的估計值。這些檢定統計量包含了 Pepe and Fleming (1989) 提出的 WKM 檢定、四個屬於 Fleming and Harrington (1991) 提出的 WLR 檢定族 ($G^{p,q}$)，以及四個我們所提出數據驅動機動檢定統計量 $K_{p,q}^*(\hat{\beta})$ ，其中 $K_{p,q}^*(\hat{\beta})$ 為 WKM 和 $G^{p,q}$ 的線性組合。我們同時也列出與 $K_{p,q}^*(\hat{\beta})$ 對應比較的 $K_{p,q}^*(0.5)$ ， $K_{p,q}^*(0.5)$ 為 $\hat{\beta} \equiv 0.5$ 時的 $K_{p,q}^*(\hat{\beta})$ ，即 Chi and Tsai (2001) 所提出。表四至表六則分別列出在 $n_1 = n_2 = 30, 50$ 及 80 下，模擬 3000 次所得到權重估計值 $\hat{\beta}$ 的平均數和標準差。

在表一，我們可以得知，在 $n_1 = n_2 = 30$ 之下，大多數的檢定統計量皆大過預

先設定好的誤差率 0.05；若將樣本數加大到 50，見表二，大多數的檢定統計量的誤差率都能控制。由表三，當樣本數加大到 80 時，所有的檢定統計量的誤差率皆能控制得當。另一方面，在設置 I，可以由表一得知，我們所提出數據驅動機動檢定統計量 $K_{0,1}^*(\hat{\beta})$ 和 $K_{1,1}^*(\hat{\beta})$ 的檢定力略小於 $K_{0,1}^*(0.5)$ 和 $K_{1,1}^*(0.5)$ ，但隨著樣本數的加大， $K_{0,1}^*(\hat{\beta})$ 和 $K_{1,1}^*(\hat{\beta})$ 的檢定力是與 $K_{0,1}^*(0.5)$ 和 $K_{1,1}^*(0.5)$ 不相上下的。而在配置 II 至 IV 來看，我們所提出數據驅動機動檢定統計量 $K_{p,q}^*(\hat{\beta})$ ，皆明顯表現得比 Chi and Tsai (2001) 所提出的 $K_{p,q}^*(0.5)$ 優異。所以，我們所提出數據驅動機動檢定統計量在面對各種不同的狀況，幾乎可以說擁有最強的檢定力，能敏銳地檢定出兩組存活分配是否相同。此外，我們計算了在各個設置下， $K_{p,q}^*(\hat{\beta})$ 對於 $K_{p,q}^*(0.5)$ 的檢定力比值，在 $n_1 = n_2 = 30$ 之下，其範圍為 0.964 至 1.102；在 $n_1 = n_2 = 50$ 之下，其範圍為 0.991 至 1.101；在 $n_1 = n_2 = 80$ 之下，其範圍為 0.993 至 1.078。由表四、五及六可說明，在不同的狀況下，我們所提出數據驅動機動檢定統計量之權重參數會有所改變。因此，將 WLR 統計量與 WKM 統計量合併在一起，其權重固定各給 0.5 是不適宜的。

第五章、實例分析

本章，我們利用兩筆實際資料說明所提方法的應用。第一筆實際資料（實例一）是由 Nahman 等人 (1992) 在一項評估腎功能不全的患者接受治療後，首次發生出口部位感染的時間（月）的研究中所獲得的，患者的出口部位若在試驗結束時未感染或患者失去追蹤，則視之為設限資料。此研究中，患者總數為 119 位，43 位患者接受了手術治療 (Surgically)，其中 65% (28 位患者) 為設限資料；76 位患者接受皮下插管治療 (Percutaneously)，其設限比率為 86% (65 位患者)。圖二為存活函數估計圖及累積風險估計圖，從存活函數估計圖顯而易見，兩組於後期存活函數差異甚大。由各種相關的檢定所得到之檢定統計量及其單邊 p 值列於表七，其中 $K_{0,0}^*(\hat{\beta})$ 、 $K_{1,0}^*(\hat{\beta})$ 、 $K_{0,1}^*(\hat{\beta})$ 與 $K_{1,1}^*(\hat{\beta})$ 的權重估計值皆為 1。在顯著水準 0.05 之下，WKM、 $G^{0,0}$ 、 $G^{1,0}$ 、 $K_{0,0}^*(0.5)$ 、 $K_{1,0}^*(0.5)$ 、 $K_{0,0}^*(\hat{\beta})$ 與 $K_{1,0}^*(\hat{\beta})$ 並無法偵測出此兩組存活分配有差異，而看重中期風險差異的 $K_{1,1}^*(0.5)$ 與 $K_{1,1}^*(\hat{\beta})$ 、看重晚期風險差異的 $K_{0,1}^*(0.5)$ 與 $K_{0,1}^*(\hat{\beta})$ 則能有效地偵測出此兩組存活分配有差異，也就是說，對腎功能不全的患者而言，接受皮下插管治療感染的風險比接受手術治療感染的風險來得低。其中， $K_{p,q}^*(\hat{\beta})$ 的 p 值皆比 $K_{p,q}^*(0.5)$ 的 p 值來得小，即我們所提方法能提供比 Chi and Tsai (2001) 所提方法更強烈的證據。

接著，第二筆實際資料（實例二）是由 Ichida 等人 (1993) 在美國中西部的一所大型大學醫學中心評估燒燙傷患者使用消毒劑後，首次發生傷口感染的時間（日）的研究中所獲得的，患者的傷口若在試驗結束時未感染或患者失去追蹤，則視之為設限資料。傷口的感染是延長住院的患者及死亡的嚴重燒燙傷患者常見

的併發症。感染的控制對燒燙傷管理來說，是非常重要的。其研究目的是比較正常沐浴護理方法（先用 10%的碘伏進行表面消毒，再使用液狀抗菌肥皂正常沐浴）與身體清潔方法（使用 4%的葡萄糖酸洗必泰溶液清潔）的療效。此研究中，患者總數為 154 位，70 位患者接受了身體清潔方法 (Body Cleansing)，其中 44% (31 位患者) 為設限資料；84 位患者接受正常沐浴護理方法 (Routine Bathing)，其設限比率為 29% (24 位患者)。圖三為存活函數估計圖及累積風險估計圖，從存活函數估計圖顯而易見，兩組於中期存活函數差異甚大。由各種相關的檢定所得之檢定統計量及其單邊 p 值列於表七，其中 $K_{0,0}^*(\hat{\beta})$ 、 $K_{1,0}^*(\hat{\beta})$ 、 $K_{0,1}^*(\hat{\beta})$ 與 $K_{1,1}^*(\hat{\beta})$ 的權重估計值分別為 0、1、0 與 0。在顯著水準 0.05 之下，除了看重晚期風險差異的 $G^{0,1}$ 無法偵測出此兩組存活分配有差異，其他的檢定統計量皆能有效地偵測出此兩組存活分配有差異，也就是說，患者欲預防燒燙傷傷口感染，正常沐浴護理方法是比身體清潔方法有效的。其中， $K_{p,q}^*(\hat{\beta})$ 的 p 值皆比 $K_{p,q}^*(0.5)$ 的 p 值來得小，即我們所提方法能提供比 Chi and Tsai (2001) 所提方法更強烈的證據。

第六章、結論與未來研究方向

在本文，我們提出一依據 WLR 統計量與 WKM 統計量的線性組合之數據驅動機動檢定統計量，選取使其達到最大的權重，再利用排序技巧造其抽樣分配。由模擬研究的結果可得知，在各種不同的狀況下，我們所提出的數據驅動機動檢定，在大部分情況下，其檢定力皆比 Chi and Tsai (2001) 所提出的檢定優越。由兩筆實例資料的分析也可知悉，為了檢定兩組存活分配是否有差異，我們所提出的數據驅動機動檢定所得到的 p 值一直是最小的。如果特定的對立假設是研究者可以預期的，那麼 WLR 檢定與 WKM 檢定或其他適當的檢定能夠獲得更強的檢定力。無論如何，當研究者缺乏相關的臨床知識或者對真實的存活函數一無所知，我們所提出的方法提供了在各種狀況下，都能檢定出兩組存活分配是否有差異的優越準則。

我們的方法不僅可以延伸在確認處理效應的多重比較問題，甚至可以直接應用在成對或群聚右設限資料。

參考文獻

1. O.O. Aalen, Nonparametric inference for a family of counting processes, *Ann. Statist.* **6** (1978), pp. 701–726.
2. Y.C. Chi and M.H. Tsai, Some versatile tests based on the simultaneous use of weighted logrank and weighted Kaplan–Meier statistics, *Comm. Statist. Simulation Comput.* **30** (2001), pp. 743–759.
3. T.R. Fleming and D.P. Harrington, Counting Process and Survival Analysis, *Wiley, New York* (1991).
4. E.A. Gehan, A generalized Wilcoxon test for comparing arbitrarily single-censored samples, *Biometrika.* **52** (1965), pp. 203–223.
5. J.M. Ichida, J.T. Wassell, M.D. Keller, and L.W. Ayers, Evaluation of Protocol Change in Burn-Care Management Using the Cox Proportional Hazards Model with Time-Dependent Covariates, *Statistics in Medicine* **12** (1993), pp. 301–310.
6. E.L. Kaplan and P. Meier, Nonparametric estimator from incomplete observation, *J. Amer. Statist. Assoc.* **53** (1958), pp. 457–481.
7. J.W. Lee, Some versatile tests based on the simultaneous use of weighted logrank statistics, *Biometrics* **52** (1996), pp. 721–725.
8. N. Mantel, Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Rep.* **50** (1966), pp. 163–170.

9. N.S. Nahman, D.F. Middelndorf, W.H. Bay, R. McElligott, S. Powell, and J. Anderson, Modification of the percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: Clinical results in 78 patients, *J. Am. Soc. Nephrol.* **3** (1992), pp. 103–107.
10. W. Nelson, Theory and applications of hazard plotting for censored failure data, *Technometrics* **14** (1972), pp. 945–965.
11. M.S. Pepe and T.R. Fleming, Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data, *Biometrics* **45** (1989), pp. 497–507.
12. R. Peto and J. Peto, Asymptotically efficient rank invariant test procedures (with discussion), *J. R. Stat. Soc. B* **135** (1972), pp. 185–206.
13. P.L. Prentice, Linear rank tests with right censored data, *Biometrika* **65** (1978), pp. 167–179.
14. Y. Shen and J. Cai, Maximum of the weighted Kaplan–Meier tests with applications to cancer prevention and screening trials, *Biometrika* **57** (2001), pp. 837–843.
15. S. Yang and R. Prentice, Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data, *Biometrika* **92** (2005), pp. 1–17.
16. S. Yang and R. Prentice, Improved logrank-type tests for survival data using adaptive weights, *Biometrics* **66** (2010), pp. 30–38.

附錄

表一、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 30$ 及 $\alpha = 0.05$ 下，誤差率及檢定力之估計

	I	II	III	IV	Null
WKM	0.659	0.634	0.242	0.421	0.057
$G^{0,0}$	0.707	0.533	0.351	0.503	0.059
$G^{1,0}$	0.654	0.585	0.226	0.359	0.055
$G^{0,1}$	0.592	0.262	0.460	0.658	0.060
$G^{1,1}$	0.648	0.339	0.436	0.692	0.061
$K_{0,0}^*(0.5)$	0.690	0.603	0.291	0.463	0.058
$K_{1,0}^*(0.5)$	0.658	0.616	0.234	0.382	0.054
$K_{0,1}^*(0.5)$	0.693	0.512	0.394	0.619	0.062
$K_{1,1}^*(0.5)$	0.683	0.538	0.357	0.604	0.060
$K_{0,0}^*(\hat{\beta})$	0.693	0.595	0.305	0.497	0.054
$K_{1,0}^*(\hat{\beta})$	0.658	0.610	0.239	0.421	0.054
$K_{0,1}^*(\hat{\beta})$	0.668	0.548	0.387	0.637	0.054
$K_{1,1}^*(\hat{\beta})$	0.670	0.562	0.368	0.641	0.057
第一組的 設限比率	0.244	0.547	0.307	0.367	
第二組的 設限比率	0.433	0.729	0.431	0.529	0.431

表二、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 50$ 及 $\alpha = 0.05$ 下，誤差率及檢定力之估計

	I	II	III	IV	Null
WKM	0.853	0.830	0.377	0.666	0.053
$G^{0,0}$	0.872	0.735	0.481	0.706	0.056
$G^{1,0}$	0.830	0.798	0.321	0.515	0.048
$G^{0,1}$	0.766	0.363	0.621	0.820	0.057
$G^{1,1}$	0.823	0.470	0.597	0.879	0.051
$K_{0,0}^*(0.5)$	0.866	0.796	0.425	0.696	0.052
$K_{1,0}^*(0.5)$	0.842	0.816	0.350	0.592	0.051
$K_{0,1}^*(0.5)$	0.858	0.685	0.552	0.822	0.054
$K_{1,1}^*(0.5)$	0.858	0.721	0.505	0.825	0.054
$K_{0,0}^*(\hat{\beta})$	0.869	0.805	0.451	0.702	0.051
$K_{1,0}^*(\hat{\beta})$	0.844	0.814	0.361	0.644	0.052
$K_{0,1}^*(\hat{\beta})$	0.850	0.754	0.571	0.821	0.051
$K_{1,1}^*(\hat{\beta})$	0.856	0.766	0.539	0.854	0.053
第一組的 設限比率	0.244	0.547	0.307	0.367	0.431
第二組的 設限比率	0.433	0.729	0.431	0.529	

表三、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 80$ 及 $\alpha = 0.05$ 下，誤差率及檢定力之估計

	I	II	III	IV	Null
WKM	0.962	0.947	0.562	0.858	0.049
$G^{0,0}$	0.970	0.898	0.683	0.862	0.048
$G^{1,0}$	0.952	0.929	0.445	0.686	0.049
$G^{0,1}$	0.913	0.471	0.819	0.937	0.048
$G^{1,1}$	0.948	0.628	0.802	0.973	0.049
$K_{0,0}^*(0.5)$	0.968	0.929	0.626	0.865	0.047
$K_{1,0}^*(0.5)$	0.958	0.938	0.506	0.779	0.048
$K_{0,1}^*(0.5)$	0.966	0.853	0.759	0.942	0.048
$K_{1,1}^*(0.5)$	0.964	0.883	0.715	0.946	0.048
$K_{0,0}^*(\hat{\beta})$	0.968	0.934	0.651	0.869	0.046
$K_{1,0}^*(\hat{\beta})$	0.960	0.938	0.535	0.840	0.050
$K_{0,1}^*(\hat{\beta})$	0.959	0.911	0.774	0.943	0.046
$K_{1,1}^*(\hat{\beta})$	0.959	0.917	0.745	0.963	0.047
第一組的 設限比率	0.244	0.547	0.307	0.367	0.431
第二組的 設限比率	0.433	0.729	0.431	0.529	

表四、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 30$ 及 $\alpha = 0.05$ 下，數據驅動機

動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差，標準差列於括號內

	I	II	III	IV	Null
$K_{0,0}^*(\hat{\beta})$	0.675 (0.460)	0.324 (0.435)	0.719 (0.443)	0.592 (0.486)	0.503 (0.490)
$K_{1,0}^*(\hat{\beta})$	0.518 (0.497)	0.350 (0.466)	0.449 (0.496)	0.333 (0.471)	0.484 (0.499)
$K_{0,1}^*(\hat{\beta})$	0.454 (0.400)	0.221 (0.316)	0.694 (0.391)	0.701 (0.387)	0.500 (0.451)
$K_{1,1}^*(\hat{\beta})$	0.492 (0.447)	0.238 (0.335)	0.733 (0.404)	0.800 (0.354)	0.502 (0.466)

表五、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 50$ 及 $\alpha = 0.05$ 下，數據驅動機

動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差，標準差列於括號內

	I	II	III	IV	Null
$K_{0,0}^*(\hat{\beta})$	0.707 (0.449)	0.251 (0.399)	0.782 (0.407)	0.569 (0.488)	0.509 (0.490)
$K_{1,0}^*(\hat{\beta})$	0.423 (0.487)	0.309 (0.454)	0.296 (0.455)	0.114 (0.317)	0.488 (0.498)
$K_{0,1}^*(\hat{\beta})$	0.415 (0.381)	0.140 (0.242)	0.743 (0.363)	0.707 (0.363)	0.506 (0.453)
$K_{1,1}^*(\hat{\beta})$	0.476 (0.441)	0.155 (0.263)	0.795 (0.365)	0.859 (0.293)	0.501 (0.470)

表六、各個設置下，設限分配為 $U(0,2)$ ， $n_1 = n_2 = 80$ 及 $\alpha = 0.05$ 下，數據驅動機

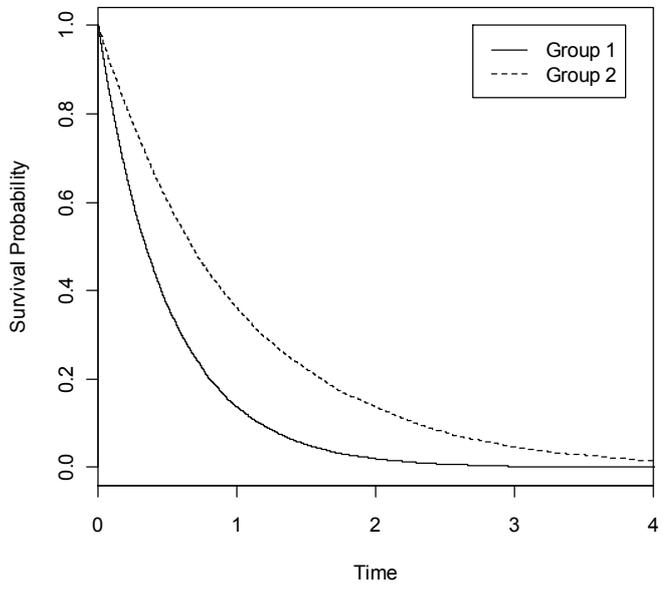
動檢定權重估計值 $\hat{\beta}$ 的平均數、標準差，標準差列於括號內

	I	II	III	IV	Null
$K_{0,0}^*(\hat{\beta})$	0.751 (0.429)	0.182 (0.348)	0.862 (0.339)	0.572 (0.487)	0.512 (0.490)
$K_{1,0}^*(\hat{\beta})$	0.342 (0.458)	0.272 (0.439)	0.129 (0.333)	0.027 (0.163)	0.500 (0.497)
$K_{0,1}^*(\hat{\beta})$	0.368 (0.350)	0.082 (0.170)	0.799 (0.312)	0.733 (0.320)	0.500 (0.453)
$K_{1,1}^*(\hat{\beta})$	0.437 (0.430)	0.098 (0.197)	0.873 (0.287)	0.927 (0.202)	0.483 (0.470)

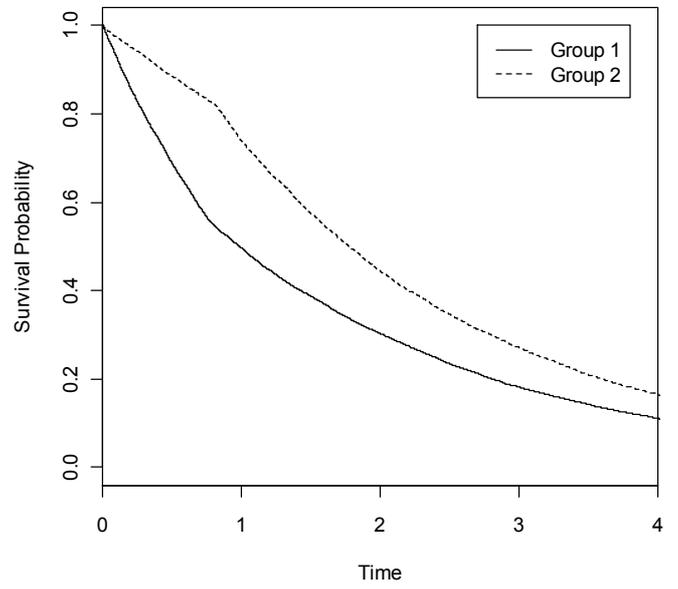
表七、實例分析，其各種檢定之檢定統計量及單邊 p 值

	腎臟手術資料		燒燙傷消毒資料	
	檢定統計量	p 值	檢定統計量	p 值
WKM	0.821	0.205	2.992	0.001
$G^{0,0}$	1.590	0.055	2.691	0.003
$G^{1,0}$	1.177	0.119	3.253	0.001
$G^{0,1}$	3.109	0.001	0.936	0.174
$G^{1,1}$	3.135	0.001	1.999	0.022
$K_{0,0}^*(0.5)$	1.213	0.112	2.814	0.0024
$K_{1,0}^*(0.5)$	0.997	0.159	3.081	0.0010
$K_{0,1}^*(0.5)$	2.234	0.012	2.120	0.0169
$K_{1,1}^*(0.5)$	2.193	0.014	2.571	0.0051
$K_{0,0}^*(\hat{\beta})$	1.590	0.052	2.992	0.0016
$K_{1,0}^*(\hat{\beta})$	1.177	0.093	3.253	0.0006
$K_{0,1}^*(\hat{\beta})$	3.109	0.001	2.992	0.0023
$K_{1,1}^*(\hat{\beta})$	3.135	0.001	2.992	0.0016

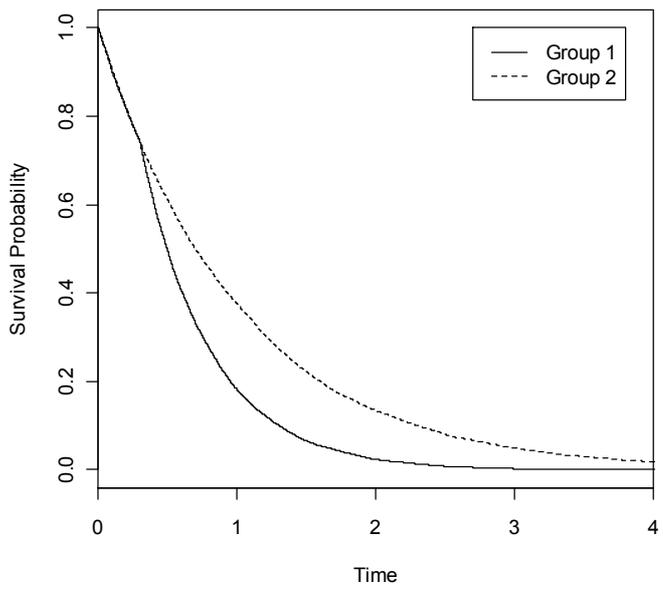
I $\lambda_1(t) = 2$
 $\lambda_2(t) = 1$



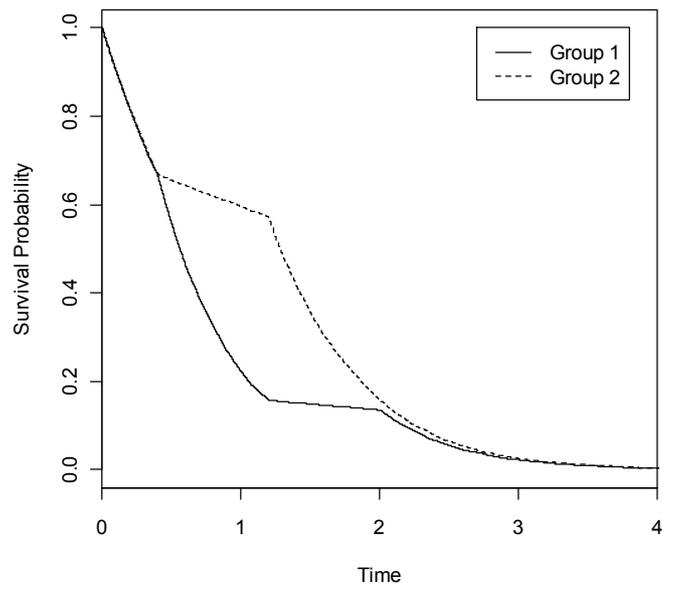
II $\lambda_1(t) = 0.75I\{t < 0.8\} + 0.5I\{t \geq 0.8\}$
 $\lambda_2(t) = 0.25I\{t < 0.8\} + 0.5I\{t \geq 0.8\}$



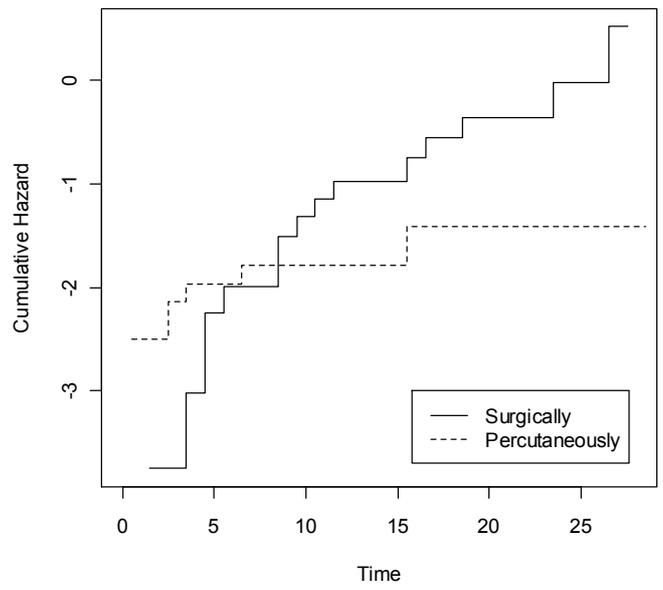
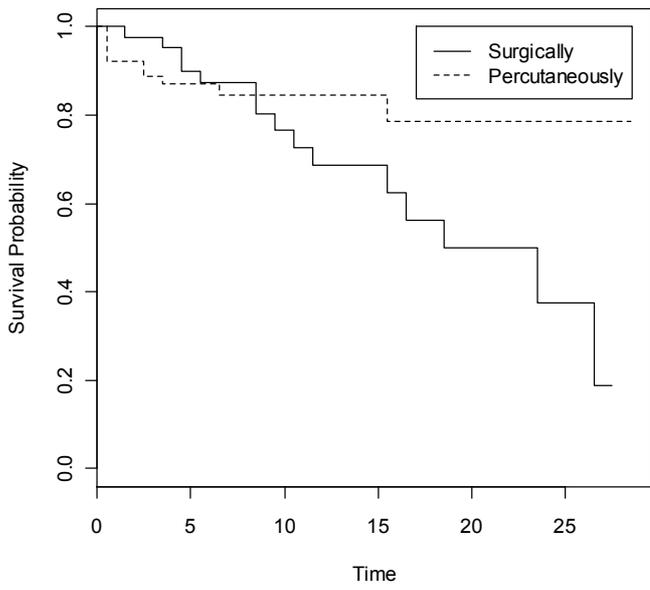
III $\lambda_1(t) = 1I\{t < 0.3\} + 2I\{t \geq 0.3\}$
 $\lambda_2(t) = 1$



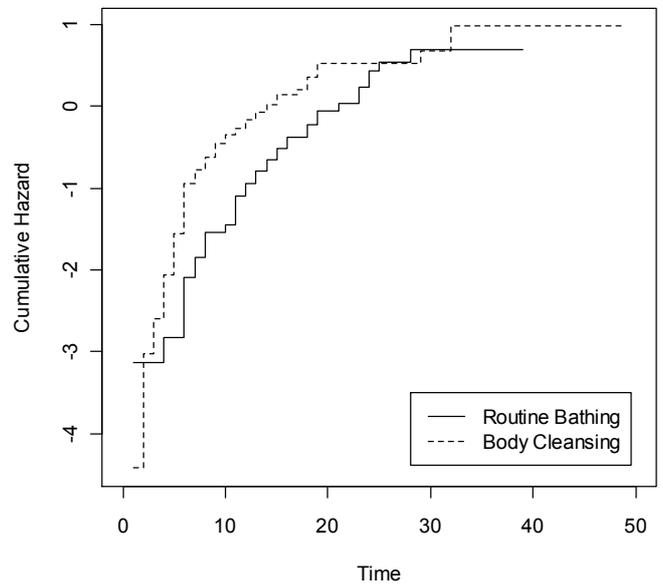
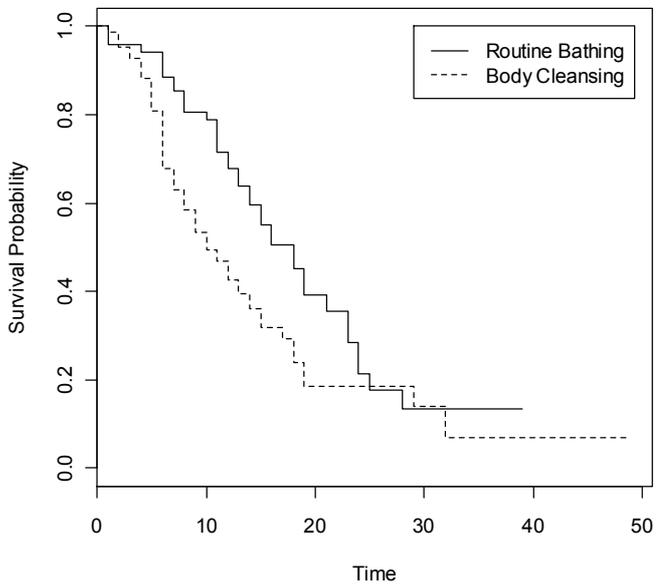
IV $\lambda_1(t) = 1I\{t < 0.4\} + 1.8I\{0.4 \leq t < 1.2\}$
 $+ 0.2I\{1.2 \leq t < 2.0\} + 1.8I\{t \geq 2.0\}$
 $\lambda_2(t) = 1I\{t < 0.4\} + 0.2I\{0.4 \leq t < 1.2\}$
 $+ 1.6I\{1.2 \leq t < 2.0\} + 1.8I\{t \geq 2.0\}$



圖一、各個設置下的風險函數圖



圖二、腎臟手術資料的存活函數估計圖及累積風險估計圖



圖三、燒燙傷消毒資料的存活函數估計圖及累積風險估計圖