



Statistical issues in the prospective monitoring of health outcomes across multiple units

出處：*J. R. Statist. Soc. A (2004) 167, Part 3, pp. 541–559*

作者：Clare Marshall, Nicky Best, Alex Bottle
and Paul Aylin

報告學生：田馥華

指導老師：童超塵教授

Content

- Introduction
- Motivating example
- Statistical process control charts for monitoring performance of a single unit over time
- Modeling the in-control process: adjusting for multiple sources of within and between-unit variation
- Acknowledging multiple testing over units as well as over time
- Application to monitoring the performance of 169 National Health Service hospital trusts
- Discussion

Keywords: *Clinical governance; False discovery rates; Hospital episode statistics; Overdispersion; Routine surveillance; Statistical process control charts*

Introduction (1/4)

- 健康保健發佈(Bristol皇家療養院調查，2000,2001；Horton，2001年；Ramsay，2000年；國家健康所，1997)呼籲應更嚴格監測如死亡率結果。某些國家已制定措施以方便實行監控(國民健康服務執行機構，1999；衛生署，2000)包含最近宣布的辦公室衛生保健績效資訊(衛生署，2002)。需明確說明確立目標。

- Bristol皇家療養院委託調查目的乃是調查其不尋常表現(單位或個人)在特定預選臨床專業的典型綜合分析(Aylin等人，2001; Spiegelhalter等人，2002)，另外關注臨床表現進行健康統計監測檢測。著重於後者並考慮關於例行健康結果資訊監控的方法和實際問題。

- 強調提出例行健康結果資訊分析可以形成表現的'篩選'工具，然後用來針對臨床審核做進一步調查。對於給定單位成果率(如外科醫生，全科醫生，醫院或衛生部門)可被看做基本過程實現。從統計角度，任何監測系統目的是檢測問題發生後進程重要改變。

Introduction (2/4)

- 監測方法已應用於卡爾曼過濾器(Smith等人, 1983), 貝葉斯結構變動方法(Carter和Blight, 1981)和標準假設檢驗程序應用典型 I 錯誤率決定觀察的統計顯著性和預期結果。
- 主要問題出在所有數據被收集後最後執行‘一次性’測試。與此相反, 預期監控系統獨特特點是經過一段時間每個時間點重複分析。
- SPC管制圖用於多重測試結果發表因為易於實施和介紹簡單, 並是其中最廣泛使用如時間數列分析。
- 提出在產業發展控制內過程監測健康結果使用。Mohammed等人(2001)主張使用修華特管制圖在臨床治理, 其中Lovegrove等人(1997)和Poloniecki等人(1998)使用CUSUM管制圖藉由觀察和調整預期死亡病症使用術前危險評分繪製累積差異監測心臟手術死亡率。

Introduction (3/4)

- 在公共健康監測應用包括檢測美國各種呈報疾病的暴發 (Vanbrackle和 Williamson, 1999)和挪威唐氏症出生缺陷的變化 (Lie等人, 1993)。監測隨時間推移的單一單位。
- 除了最近監測死亡率一般做法 (Aylin等人, 2003)和有限問題討論 (Spiegelhalter等人, 2003), 不知道任何同時監測健康結果解決問題在多單位及時間點的文獻。
- 擴展來在 $M > 1$ 單位同時監測結果(其中‘單位’參考如個人外科醫生或普通科醫生, 國家健康服務(NHS)醫院、初級保健或衛生局)。
- 尤其以下問題。
- (a) 要求當被監視過程可能被認為是可接受時定義或管制內被視為已不可接受、超控。監測各單位數據多單位不同來源變異是固有觀察結果, 特別在適當規範下, 管制內分配, 承認單位間不可避免的變化(過度離散)是真實但可接受的表現水準。

Introduction (4/4)

- (b)專為調整多重檢測所設計處理多重顯著性檢測問題。
- (Aylin等人，2003年)提出監測每年死亡率的實際應用，此文提供更詳細討論統計方法問題。
- Section2，介紹1995~2000 169間NHS醫院委託使用數據監測各種原因年度死亡率。
- Section3，勾勒單一單位監測管制圖。
- Sections4和5分別考慮關於監控多單位問題的方法，Section 6運用 NHS委託數據。
- 得出結論：在Section7 NHS建議建立正式表現監控系統並簡單討論一些實作和政策影響。

Motivating example

- 考慮監測NHS委託169家1995~2000所有原因死亡率。該委託代表非專科中心，主要是地區綜合醫院。信託數據來自Dr Foster Unit倫敦帝國學院醫院事件統計數據。
- 觀察超過80個診斷死亡率計數總結在所有醫院死亡共佔80% (Jarman等人，1999)。
- 透過間接標準化調整年、年齡、性別，診斷獲得預期數字 (三位數國際疾病分類代碼)並承認方法(選擇性或緊急事件)。由於在病人離開醫院後死亡發生未記錄於醫院統計內，故調整之。
- 為了簡化把住院治療作為分析單位，事實上是總再住院的19%。可能是個問題，因此稀釋其死亡率。總體而言再住院率和死亡率間幾乎無相關(- 0.14)。
- 此文目的是假設這些數據收集前瞻性，而感興趣的是檢測死亡率的所有原因重要變化(增加和減少)。

Statistical process control charts for monitoring performance of a single unit over time (1/4)

- 在衛生監測方面幾種不同類型的SPC管制圖已建議(看Sonnesson和Bock(2003)最近回顧)，但都有以下主要特點
- (a)計算每個單位時間檢測統計量 S_t ($t = 1, 2, \dots$)：
根據標準化殘差 R_t 公式和預期管控下分配觀測結果。 S_t 取決於之前的殘差值，導致累積總和。
- (b)一個預定義的門檻 h ：
如果檢測統計量 S_t 超過 h ，警報被觸發，管制圖信號失控。
- (c)當基本過程是真的且超控管制圖長期檢測能力表現測量：
如‘表現衡量’在普通顯著性測試採取更熟悉的型 I .型 II 錯誤率。
在此，使用正常累積加總(CUSUM)管制圖檢測統計量定義成

$$S_t = \max(0, S_{t-1} + R_t - K/2)$$

Statistical process control charts for monitoring performance of a single unit over time (2/4)

- 其中 $R_t - K/2$ 與對數概似比成正比， $\log\{f_1(y_t)/f_0(y_t)\}$ ，其中 f_0 和 f_1 分別表示管制內和管制外的分配，而 y_t 為觀測結果(如死亡人數)。在此我們假設 $f_0 \sim N(\mu, \sigma^2)$ 、 $f_1 \sim N(\mu + K\sigma, \sigma^2)$
- $K\sigma$ 代表藉由管制圖的檢測基本過程平均 μ 增加。
- 不妨指定被檢測的兩個不同超控偏移：一個代表警告率(例如 $K=1$ 或 $K=2$)，一個代表警報率(例如 $K=3$ 或 $K=4$)。延遲討論 R_t 計算，觀察結果標準化殘差假設常態分配管制內與管制外過程，直到 Section 4.1。
- 在CUSUM管制圖使用檢驗統計量用來檢測結果下降率，給定
$$S_t = \min(0, S_{t-1} + R_t + K/2)$$
- 特別選擇SPC管制圖動機是CUSUM統計量概似比有各種優良特性和其他類型的 SPC管制圖相比。

Statistical process control charts for monitoring performance of a single unit over time (3/4)

- 特別在某種意義上已被證明是最佳(其他類型CUSUM管制圖間)，給定真正變化在結果檢測最大機會給定假陽性率(Moustakides, 1986)。
- 另一CUSUM檢測統計量優點是界線低於零，管制圖在密切相關的序列概率比檢定(Wald, 1947)可能為正或負值。
- 如果進程在管制內序列概率比檢定統計量變越來越負，使建立過去表現良好單位幫助掩蓋惡化表現，至少初期後變化。但似乎不具吸引力在健康監測系統。
- 另一選擇是指數加權移動平均管制圖(Hunter, 1986)。檢驗統計量降低指數函數的權重殘差 R_1, \dots, R_{t-1} ，對比的CUSUM統計量給予相等權重。雖然前者可能看起來很有吸引力並不清楚如何選擇降低權重所以我們偏好相等的權重，雖然在目前情況下承認值得更進一步的研究。

Statistical process control charts for monitoring performance of a single unit over time (4/4)

- 簡單的CUSUMs替代在臨床性能監控主張使用修華特管制圖（例如Mohammed等人，2001）。
- 然而，每個監測修華特管制圖統計量獨立計算使用最近觀測值(或者說，從最後監測時間觀測平均)，因此不會利用有關單位所有可用資訊研究。
- 雖然基本過程快速反應變化將顯示檢測小而溫和的持續改變(更可能代表定義目標製程偏移)觀察結果往往特別容易波動 (Vanbrackle 與 Williamson,1999;Williamson 和 Weatherby-Hudson,1999)。

3.1. Setting the alarm threshold and deriving chart performance measures (1/3)

- 通常統計概念在序列測試下型 I 和型 II 錯誤是不恰當的，因為錯誤率取決於時間，特別是型 I 錯誤趨近 1 時間趨近 ∞ 。
- Frisen(1992) 提出替代用來總結 SPC 管制圖監控性能，最常見的是平均運行時間(時間點平均數量檢定統計量首先超過警報門檻值 h)。
- 管制內平均運行長度定義為 ARL_0 ；信號應為長期代表假警報。
- 超控的平均運行長度定義為 ARL_1 ，信號應盡可能短然後代表真警報。
- RL 分配取決於門檻值參數 h 和檢測管制外偏移大小 K 。
- ARL_0 和 ARL_1 製表值 h 和指定管制外偏移 K 選擇適當值 h ，反映快速檢測藉由快速檢測真警報太早得到假警報風險權衡。

3.1. Setting the alarm threshold and deriving chart performance measures (2/3)

- 選擇圖表門檻值使用 ARL_s 並總結表現，Frisen(1992)指出RL分配通常顯著傾斜，因此平均未必最合適。批評作為衡量管制圖表現的 ARL_1 因從管制內到外假設偏移監測後同一時間內。
- 實踐變化更可能發生在監測開始後一些未知時間。文獻中提出其他替代評估標準，最相關為
- (a) 時間 t 假警報機率(PFA_t)—改變大的假警報和連續長期監控時間增加函數。
- (b) 時間 t 內檢測成功機率變化(PSD_t) - 給定改變前無假警報在某個時間間隔內管制內到超控實際發生改變後系統信號機率(真警報)。機率取決於時間在哪個失控發生改變和大小變化。
- 表1總結CUSUM管制圖資料測量表現或假設管制內標準化殘差服從標準常態分配(模擬100000次; Monte Carlo標準誤差計算微不足道)。

3.1. Setting the alarm threshold and deriving chart performance measures (3/3)

- 檢驗表現真正改變提升警報門檻盡量減少假警報風險。如，假設想要檢測 1個標準差($K = 1$)增加監測每年數據。
- 表1第二排顯示選擇管制圖門檻值 $h = 3$ 在6年內(\widehat{PFA}_6)假信號機會有3%(如激勵時間序列長度是6年)。
- 如果單位真正超控(此例子中對應單位有一個平均是1標準差以上)，在第1到第12年有90%機會管制圖有正確警報信號(基本第5和第95百分位數超控運行長度分配 RL_1)；如果單位開始錯誤其後超控(均勻離散區間 $[0, 100]$ ，管制圖有67%機會正確檢測到變化其發生6年內(\widehat{PSD}_6)。
- 如果選擇較低門檻(如 $h = 2$ ；表1第一行)則有更大機會正確檢測到超控，但較高假警報率。給定任何門檻值 h 增加 k 提高區別真正、負值結果。

Modelling the in-control process: adjusting for multiple sources of within and between-unit variation (1/2)

- 監控行動第一目標是了解必然變化；能評估觀察績效表現該單位利益是否以合理假設以符合其他單位模式變化。
- 變異可能由於以下原因：
 - (a) 隨機(偶然)波動：將此稱為型A變異。
 - (b) 已知變異來源(型B)如不同單位間或隨時間在年齡性別分佈或社會經濟地位低下級別病患人口：往往稱為'病症'。
 - (c) 不可避免單位間變化因為許多小淨影響，不可測量因素(型C1)：如數據品質和更不可測量病人病例組合差異。大多數單位死亡率受很多因素影響；任何個別特殊因的影響可能很小，但許多小而不可測因素合併後效果可能導致出現不尋常。隨時間推移所有單位的持續變化也可能導致如持續時間趨勢死亡率。

Modelling the in-control process: adjusting for multiple sources of within and between-unit variation (2/2)

- (d) 不明來源更系統的變化只與一或少數幾個單位有關(型C2)，但也可能導致不尋常死亡率：非常好或非常壞的表現因為管制範圍屬於這類系統因素影響大而不可測量病症因素如存在的養老院或收容所。
- (e) 季節變化和極端外部力量影響(型D)：如季節性溫度變化影響是已知的影響死亡率，傳染病或特別嚴重的冬天也可能增加每年死亡率。

4.1. Adjusting for type A, B and D variation (1/3)

- 通常欲調整變數每單位預期(管制內)結果率因為偶然(型A)，已知風險因素(型B)和任何觀察時間趨勢(型D)適用於所有單位。可通過回歸或標準化定義適當抽樣分配和一般季節性趨勢已知風險影響因素協同調整實現。
- 調整病例主要困難(已知風險因素)是除非臨床可接受風險調整計劃已存在，可能使用性能分析關注風險調整特定方法有效性。如，如果預計風險調整方案有危險觀察過度補償性能差異從正在監測數據(Goldstein, 2002)。
- 間接標準化參考利率選擇也需仔細考慮。如在早已確定流行病學文獻中提到如果涵蓋廣泛地理區域，利用國家參考利率導致大範圍檢測區域死亡率或發病率標準化差異。
- 當地參考利率可能因此更適合用於檢測個別單位間性能變化，提供足夠強大假設已知沒有錯誤。

4.1. Adjusting for type A, B and D variation (2/3)

- 嚴格地說，每單位預期死亡率估計所使用參考利率不包括該單位數據，如果該單位死亡率實際上超控則可能偏差。然而，任何單位影響可能是總數據量最小相對超過所有區域匯集，所以對於實際原因是簡單計算單一參考利率計算數據，從各單位而不是試圖全面cross-validatory依次為每單位估計單位特定率。
- 大多數SPC管制圖假設平穩平均值和變異數在獨立控制過程，顯然在大多數健康結果不能成立，變化隨時間推移如病人病例組合必然導致非固定預期成果率。
- 這問題可很容易處理藉由監控相對標準化殘差 R_t ($t = 1, 2, \dots$)—如果結果計數數據初步轉變正常。強調之，但基於表現近似統計比較可能錯誤如果近似假設(例如人口眾多風險要求)未得到充分滿足 (Christiansen和Morris, 1997)。

4.1. Adjusting for type A, B and D variation (3/3)

- 未來監測，Lovegrove等人(1997)討論監控成人心臟外科手術死亡率。假設結果服從Poisson抽樣分配(型A變異)和使用Parsonnet評分系統(Parsonnet等人，1989)得到術前風險調整死亡期望機率(型B變異)。
- Rossi等人(1999) 監控在北托斯卡納男性呼吸系統死亡率，假設每月計數呼吸死亡常態近似Poisson抽樣分配。
- 使用計算預期計數標準化每月年齡分別期間平均率進行風險調整。管制過程更複雜時間模式呈現如定期季節波動或時間相關，可利用時間序列模式(如Williamson和Weatherby-Hudson(1999))。
- 分析醫院信任死亡率數據調整病症變數和通過標準化時間趨勢，假設常態近似Poisson抽樣分配。

4.2 Adjusting for type C sources of variation (1/2)

- 效果不明或不可測風險因素(型C變化)很可能導致單位間觀察結果率更大變異，由二項或Poisson隨機變異預期。
- 如果感興趣在檢測單位結果間差異由於因素綜合影響，那相應二項、Poisson模型或常態近似是合理的(已知風險因素和季節變化討論後第一次調整)。
- 然而，如果主要感興趣在大而有系統的因素檢測(型C2變異)導致不尋常(超控)結果率，那理想應該調整結果率為了變異型C1元件結合許多小而不可測量影響因素。可看作調整過度離差；如果不進行會導致太多假警報。
- 對此問題膨脹管制內變數藉由估計過度離散水準 (Farrington等人，1996)。有效使觀測數據更加變異比假設下Poisson抽樣分配(或常態近似)同時仍然視為管制中。其淨影響使用統計圖表形式轉化殘差監測較小(較不極端)如果只有型A，B和D來源變異被允許。

4.2 Adjusting for type C sources of variation (2/2)

- 此方法限制估算過度離散因素。什麼是需要估計計數過程產生死亡率過度離散數量 (額外Poisson變異)。
- 任何過度離差估計自從超控觀察真實過度離差量往往使數據更加變異比產生從管制過程內情況下。
- 由於開始監控前沒法只選擇從管制內數據，所以只能估計所有數據基礎過度離散因素。得到分別估計每年並選擇中位數過度離散估計在ad hoc嘗試彌補部分可能過度膨脹估計。此外，如果數據中超控單位比例小，那麼過度離散估計應是小的。
- 今後方法應用應可利用已知可接受表現水準歷史數據改善過度離散估計。然而，對於目前例子處理歷史數據如果有前瞻性的被收集。

Acknowledging multiple testing over units as well as over time (1/7)

- 以上各種測量管制圖表現討論目的為提供成功檢測和假警報。
- 日常監控中死亡率數據需延長表現測量描述預期假警報率和成功進行監測中 $M > 1$ 進程（單位）。
- 一般多重顯著性檢測方法是控制整體型 I (假警報)率。如廣泛使用Bonferroni校正(Milner, 1981)確保如果M顯著性檢驗進行而且所有虛無假設為真，那麼錯誤率拒絕至少一個虛無假設小於或等於指的整體型 I 誤差率。
- 性能監控方面，許多醫生或健康中心提供者採用序列概率比管制圖，Spiegelhalter等人(2003)提出更嚴格報警門檻值設置使用類Bonferroni調整數量單位進行監測。目前情況下這種方法也有許多困難。
- 首先，其調整只涉及整體型 I 誤差率而不是多種測試程序。最佳化權衡藉由失敗和成功的檢測率。

Acknowledging multiple testing over units as well as over time (2/7)

- 此外，正如事件監控單一單位，型I錯誤率序列分析並不固定，而是隨監測時間長度增加。
- 最後，也許最重要並不是並不那麼關心所有報警都是假的至少一個假警報超出M單位檢測估計得到機率-稱為假發現率(FDR)。
- 目前沒有多重序列分析測試文獻解決這些問題。
- 這是很重要的量化表現，建議同時監控多個單位的SPC管制圖作出參數選擇如警報門檻值h。
- 大量利用估計FDR首次提出藉由Benjamini和Hochberg(1995)而最近延伸藉由Storey (2002, 2003)和Storey等人。
- 假設M是獨立單位藉由使用特定SPC管制圖，並假定自從開始監測已真實超控 m_1 ，與其他 $m_0 = M - m_1$ 保持在控制內真實。
- 表2顯示t時間間隔下m單位可能結果。

Acknowledging multiple testing over units as well as over time (3/7)

- 特別的是， $A_t(h, k)$ 定義為警報總數， $F_t(h, k)$ 定義為假警報總數使用SPC管制圖警報門檻值 h 設計檢測超控偏移 K 標準差。

- 為了符號方便，這些壓制變數依賴 h 和 k 。Storey(2002, 2003)，定義FDR為

$$FDR_t = E \left[\frac{F_t}{A_t} \mid A_t > 0 \right].$$

- 定義FDR， $E [F_t | A_t]$ 是有問題的因為正機率 $A_t = 0$ 。

- 根據假設每單位管制圖檢驗統計量獨立且有相同分配隨機變量(本例中從管制內和外的混合分配產生)，Storey(2003)顯示方程式(1)相等於

$$FDR_t = \Pr(\text{unit truly in control} \mid \text{alarm by time } t)$$

$$= p_0 \Pr(\text{alarm by time } t \mid \text{unit truly in control}) \{ p_0 \Pr(\text{alarm by time } t \mid \text{unit truly in control}) + (1 - p_0) \Pr(\text{alarm by time } t \mid \text{unit truly out of control}) \}^{-1} \quad (1)$$

Acknowledging multiple testing over units as well as over time (4/7)

- $p_0=m_0=M$ ，單位比例真實在管制內。因此FDR在某種意義上類似1 minus正預測值的標準診斷檢測程序。
- 方程式(1)分子機率和右手邊分母概率分別是假警報機率(\widehat{PFA}_t)和成功檢測機率(\widehat{PSD}_t)單一管制圖，討論在section3.1。
- 數量使用Monte Carlo估計量來估計FDR，有條件的固定為 P_0 ，藉由

$$\widehat{FDR}_t|p_0 = \frac{p_0 \widehat{PFA}_t}{p_0 \widehat{PFA}_t + (1 - p_0) \widehat{PSD}_t}$$

- 另一有用考慮數量是時間成功檢測率(即所有超控單位比例時間t警告信號，給定至少一個單位信號)。
- 類似敏感性標準診斷測試，定義為 $SDR_t = E \left[\frac{T_t}{m_1} \middle| m_1 > 0 \right]$

Acknowledging multiple testing over units as well as over time (5/7)

- 根據管制圖檢驗假設每單元檢定統計量都是獨立且同分配隨機變數，方程式(2)可表示相等機率成功檢測單一管制圖 PSD_t 。

- 因此，使用Monte Carlo估計獲得 $\widehat{SDR}_t = \widehat{PSD}_t$

- 定義類似數量為1 minus負預測值—稱為假無發現率藉由Genovese和Wasserman (2002)和Storey (2003)—和監控系統變異性。

- 注重FDR和敏感性FDR，因此設想被主要用作單位鑑定'篩選'工具可能引發進一步品質、病例組合和照顧品質組織問題調查數據。主要關心的是設計監控系統，信號有高機率真實超控。

- 主要關注情況下應避免未能識別的真正超控單位(如未能發現Harold Shipman或Bristol 皇家療養院事件)，將是有興趣在特異性和系統假非發現率。

Acknowledging multiple testing over units as well as over time (6/7)

- 注意 FDR_t 不是 SDR_t 也不取決於監控總數單位(方法對比如Bonferroni校正)，但 FDR_t 單位比例是真正超控。
- FDR_t 和 SDR_t 管制圖表現評估取決於監控長度隨 t 增加。
- 圖 1顯示監控表現估計 $t=6$ 個時間點(如6年) $M>1$ 單位固定值範圍 $P1(= 1- P0$ ，單位真實機率超控)。
- 圖 1(a)-1(c)顯示如何減少 FDR_t 增加 $P1$ (注意通常希望單位相對較小機率真實超控監測死亡率數據，因此應重點關注在 FDR_s 顯示某點)；圖 1(d)-1(f)顯示 SDR_t 保持不變不論數字是否真實超控。
- 每行三個不同點對應CUSUM管制圖設計去檢測不同程度超控性能($|K| = 1,2,3$)顯示隨超控偏移檢測變更加極端，管制圖更可能正確區分單位是真實管制內或外。

Acknowledging multiple testing over units as well as over time (7/7)

- 每點三個曲線對應管制圖警報門檻值 $|h|$ 為2, 3或4;比較每列頂點和底點說明如何增加警報門檻值降低DFR, 但成功檢測率相對低。
- \hat{FDR}_t 和 \hat{SDR}_t 數值選定由組合 $|K|, |h|$ 和 P_1 也總結於表3。
- 注意管制圖表現衡量同時應用管制圖設計檢測增加結果和下降結果。

Estimating the number of in-control units

(1/2)

- 估計FDRs，Storey (2002)提出方法估計真實虛無假設數字 m_0 。
- 目前情況下， m_0 數字對應於真實管制內單位，大綱是Storey的方法如何被使用來估計。假設已監測 m 單位 T 時間點固定觀察期間，設 T_{ii}^{\max} 定義為CUSUM統計量最大值 i th單元($i=1, \dots, M$)。
- Storey方法利用事實即在正確指定控制內分配有最大CUSUM統計量高機率低於精心挑選門檻值 λ 將與管制內單位有關。根據這假設，保守估計 $P_0(=m_0=M)$ 是

$$\hat{p}_0(\lambda) = \min\left(1, \frac{\text{observed proportion of test statistics} < \lambda}{\text{expected proportion of test statistics from the in-control distribution} < \lambda}\right)$$
$$= \min\left\{1, \sum_{i=1}^M \frac{I(T_{ii}^{\max} < \lambda)/M}{1 - \widehat{PFA}_t(\lambda)}\right\}$$

- 其中 $\widehat{PFA}_t(\lambda)$ 估計機率假警報信號給予警報門檻值 $h=\lambda$ (得到通過Monte Carlo模擬)，因此對應於期望機率最大CUSUM統計值管制內單位超過 λ 。選擇合適 λ 值。

Estimating the number of in-control units

(2/2)

- Storey(2002)提出最佳化 λ 最小均差錯誤 P_0 估計結果使用引導方法，提出模擬結果建議 P_0 是保守非常接近真實。

Application to monitoring the performance of 169 National Health Service hospital trusts

6.1. Specifying the in- and out-of-control processes (1/2)

- 假設死亡率計數 Y_{it} 單位 i 和年 t 服從過度離散Poisson分配，平均 E_{it} 是調整年齡,性別,診斷(三位數國際疾病分類代碼)停留及分配方法(選擇性或緊急情況)。
- 既要滿足SPC管制圖平穩性條件並承認額外Poisson變化固有數據，選擇採用Z-轉變
$$R_{it} = \frac{y_{it} - E_{it}}{\sqrt{(cE_{it})}}$$
- 其中 C 過度離散估計如標準類概似法(McCullagh和Nelder, 1989)。標準化殘差 R_{it} 假設服從 $N(0,1)$ 分配。管制外單位定義為平均死亡率標準化殘差標準差 K 高於期望(對於某些預先設定 K 值)以便假設 R_{it} 服從 $N(K,1)$ 分配。

6.1. Specifying the in- and out-of-control processes (2/2)

- 委託觀察死亡率計數為112至3132不等，期望為110.7到3361.0。
- 對於過度離散Poisson分配平均變數，上述轉換提供良好近似標準常態。
- 委託數據year-by-year (類概似)過度離差估計範圍為8.6到11.2，中位數為9.7，顯示高度變異性原因不明—預期超過9次會得到變化由獨自Poisson抽樣變數。這是經調整後過度離差大小固有例行健康數據測量病例組合變數。

6.2. Results (1/2)

- 委託數據信號不同 K 和 h 值給定於表4。忽視過度離差影響是明確的。
- 在所有 K 和 h 值，越來越多委託信號沒有考慮到過度離差—某些情況下超過五倍多。額外信號產生於低估管制內分配變數。
- 關注於CUSUM管制圖檢測運行結果增加 $K=2$ 標準殘差死亡率(R_{it})，如果警報門檻值設置 $h=3$ 則六單位(約4%)信號。
- 圖2(a)和2(b)顯示CUSUM管制圖這六單位和死亡率標準殘差。
- 6年來所有情況下殘差始終高於0，足夠證據表明死亡率為單位每殘餘不能被假定從管制內分配已出現。
- 為了10個委託信號圖2(c)和2(d)管制圖上類似結果在 $h=-3$ 檢測低於期望標準死亡率標準差 $K=2$ 。

6.2. Results (2/2)

- 提及多重管制圖表現衡量給定在表3，六個信號期望8.6%左右或以下在 $K = 2$ ， $h = 3$ 管制圖是假警報(取決於真實，未知，委託的機率是超控的)。
- 重要的是強調，儘管這些‘真實’信號可以被認為是真正‘統計’警報，標準殘差死亡率不能被假定在管制內，可能是假‘醫療警報’，因素解釋如錯誤數據或病症組合而不是真正質量委託真實超控。
- 還需考慮可能錯過多少真實超控單位。表3顯示在 $K=2, h=3$ 管制圖將錯過剛剛超過全部超控單位17%。這些數字相符情況圍繞七個真正超控單位，6年內已成功檢測有六個沒有假警報。
- $m_0=162$ 估計是一致得到藉由使用Storey方法，表示 $169-162=7$ 委託是真實超控。

Discussion (1/7)

- SPC管制圖已被廣泛使用於監測醫療成果，並已明確主張在臨床使用管理(Mohammed等人，2001)。以往應用考慮在醫療監測單一過程。
- 本文目標是探討試圖擴大SPC管制圖影響健康監測結果方法同時在多單位。雖然研究結果表示這種管制圖可能有價值在部分國民保健服務範圍監測表現系統，應用和解釋並沒有直接方法，有些問題仍有待解決。
- 關鍵問題是解決SPC管制圖問題真正量化和誤報率，同時考慮多重測試單位。
- 在這種情況下，估計FDRs更有用比控制整個機率獲得至少一個假警報(SPC管制圖將迅速提高到1中度敏感增加單位數量和時間點)。
- 表明FDR如何估計多重單位和固定監測期。
- FDR_t 和 SDR_t 估計報告不應被視為最終。

Discussion (2/7)

- 特別是計算所有超控單位並未考慮更現實情況單位超控後不同時間監測工作已開始。也沒考慮到指定單一超控分配並不完全能發現所有模式不尋常表現可能出現問題;現實更可能是連續因委託受到不同程度型C1(不可測量病症組合)變數來源，加上可能型C2(大有系統的不可測量因素)來源。
- 假的和成功檢測率也只適用於固定長度監測期;延長可能性概念在管制內和外運行長度分配為單一聯合分配多重單位運行長度，應加以探索。
- 其他管制圖表現措施如分配運行長度直到第一個報警或者機率是第一單位信號真正超控也值得進一步研究。同樣重要的是效果，就管制圖表現估計重新啟動—管制圖信號可能會重新啟動，通常為 $h=2$ (Lucas和Crosier, 1982)。
- 另一有趣領域將進一步研究探討理論，正式決定選擇門檻 h 。

Discussion (3/7)

- Bickle(2002)，例如定義可取功能notatio相當於

$$D(h) = \{1 - \text{FDR}(h)\}b - \text{FDR}(h)c$$

- 其中b和c分別表示檢測真正超控單位和假定義單位成本超控。然後，選擇門檻h最大化。
- Genovese和Wasserman(2002)採取類似方法，但正式平衡假警報成本在有遺漏真實警報和結果最小化損失函數。
- 第二個關鍵方法問題是已考慮如何調整必然單位間變化表現結果因為許多小淨影響無法衡量因素，特別是不可測量病症。特意區別由於這種變化從更系統因素可能反映某些方面所提供服務服務。
- CUSUM管制圖設計主要動機是已嘗試檢測單位選擇超控分配，表現可能受一或兩個主要系統因素(型C2變異來源)導致死亡率平均標準殘差大偏移。

Discussion (4/7)

- 然而實踐中，管制圖也用於檢測單位殘差死亡率由於系統不同期望淨影響在無法衡量病例組合及其他小錯誤數據，但誰可能被認為是'接受'如果原因是我們已知的型C1變數。
- 試圖防止一定程度誇大變異在管制內分配以反映每單位更多不確定性預期死亡率由於不可測量因素。
- 如果未測量病例組合每年隨機波動—如在某些年份可能治療患者往往比一般病情加重，而在其他年份患者可能會少生病，那麼應該提供合理描述在管制內分配。
- 然而，專家相信可能有未測量情況組合保持相對穩定誘導時間依賴性標準化死亡率，因此不能假定隨機發生在過度離散邊際管制內分配(雖然也不一定產生於指定超控分配)。言外之意是CUSUM管制圖最終信號將逐步逐年上升。

Discussion (5/7)

- 因此，檢測不只產生超控分配極端殘差，而且結果表明表現差異小但持續時間超過預期。如果管制圖被視為一種篩查工具以確定進一步調查單位更容易理解造成表現變化，那麼實際上可能是有吸引力的特點。
- 另一種方法處理型C1變異對各單位和時間點指定分層模型結果。這種做法明確塑造單位變化間型C1來源歸屬及適當分配假設，可預期大小和反映來源性質。
- 實際上，這種分層模型調整每計數單位預期死亡率估計以反映由於型C1來源各單位間真正差異。
- 然而，分層建模方法可用於這方面解決方法問題。
- 問題類似估計過度離差因素 (隨機效應) 理想結果應使用數據估計而不是整個數據。

Discussion (6/7)

- 另外，如果適當風險調整模式已存在(如Parsonnet成人心臟手術評分結果(Parsonnet, 1989))，則可使用加上預期分配在患者接受治療機率風險因素單位管制內預測分配。
- 關於如何分配應用來幫助指定每單位SPC管制圖統計量。隨機整合—影響分配並假設每單位標準殘差從共同邊際管制內分配—實際上是一樣過度離差的方法。
- 另外，以下建議由Farewell(2002)，可能是明智的隨機效應預測每單位從上管制tail(或下管制tail如果目的是檢測好的表現)分配，並假設如果沒有更極端比預期單位該單位在管制內觀察結果隨機影響。
- 保守做法相當於假設單位是只要沒有惡化(或更好如果檢測表現特別好)比預期情況下如果無法衡量該單位組合特別極端。
- 設計SPC管制圖監控估計單位—特別是隨機影響而不是觀察數據。

Discussion (7/7)

- 這些建議值得進一步調查，雖然普遍關注分層建模可能有損於簡單基本SPC管制圖使其更多參與且簡單非統計。
- 本文關注在死亡結果但同樣可應用CUSUMs處理結果如停留長度和等待時間。
- 常規使用CUSUM管制圖或任何表現測量方法，無論是個別醫生或整體系統如醫院—需從根本改變態度。適當報告將有助於重新確定重點從相互指責理解和處理系統改進。
- 最後，如果通過指導NHS提供高品質成本效益服務導致健康狀況改善、審核和最佳實踐，需要高品質和及時資訊。
- 任何方法無論多麼複雜如果無法使用，監控比較表現將失敗。

Table 1 單一績效衡量指標CUSUM管制圖標準常態管制內進程和N(K, 1)管制外進程，採用對數概似比檢驗統計量(100000筆模擬)

Shift K	Alarm threshold h	Percentiles of RL_0 distribution		\widehat{PFA}_6 (%)	Percentiles of RL_1 distribution		\widehat{PSD}_6 (%)	\widehat{PSD}_1 (%)
		5th	95th		5th	95th		
1	2	4	>100	12.2	0	9	83.7	13.4
1	3	9	>100	3.0	1	12	67.3	4.4
1	4	21	>100	0.7	1	16	48.6	1.4
2	2	15	>100	2.0	0	6	93.7	17.6
2	3	98	>100	0.2	1	9	82.6	4.2
2	4	>100	>100	0.02	1	13	67.8	1.0
3	2	>100	>100	0.24	0	5	96.5	24.1
3	3	>100	>100	0.01	0	8	89.0	5.4
3	4	>100	>100	<0.01	1	11	78.5	0.9

K ，超控偏移被檢測到的大小(測算多重管制內的標準差); h ，管制圖統計量門檻值將產生報警， \widehat{PFA}_t ， t 年內假警報的估計機率， \widehat{PSD}_t ， t 年內超控改變真警報估計機率;
 RL_0 ，估計管制內運行長度(從開始監控開始直到假警報的時間點數量); RL_1 ，估計超控運行長度(從開始直到進程超控開始真警報訊號的時間點數量)。

Table 2

固定長度 t 監控期間 M 獨立單位SPC
管制圖運行後可能出現的結果

	<i>Number not signalling an alarm by time t</i>	<i>Number signalling an alarm by time t</i>	<i>Total</i>
Number of units truly in control	$m_0(K) - F_t(h, K)$	$F_t(h, K)$	$m_0(K)$
Number of units truly out of control	$m_1(K) - T_t(h, K)$	$T_t(h, K)$	$m_1(K)$
Total	$M - A_t(h, K)$	$A_t(h, K)$	M

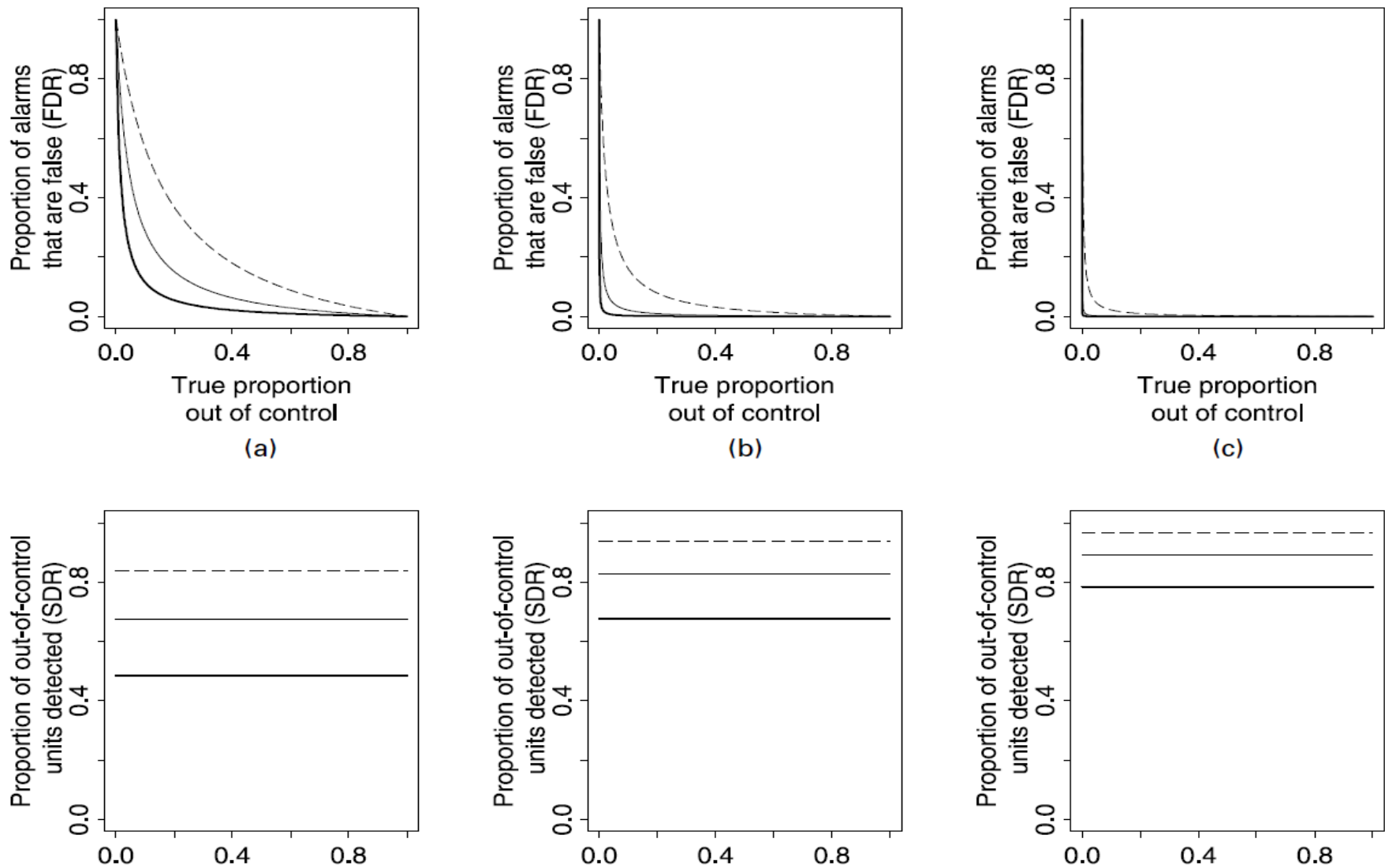


Fig. 1. Multiple chart performance measures for monitoring $M > 1$ units for six time points (years), shown as a function of the proportion of units that are truly out of control: (a) ----, $K = 1, h = 2$; ———, $K = 1, h = 3$; ———, $K = 1, h = 4$; (b) ----, $K = 2, h = 2$; ———, $K = 2, h = 3$; ———, $K = 2, h = 4$; (c) ----, $K = 3, h = 2$; ———, $K = 3, h = 3$; ———, $K = 3, h = 4$; (d) ----, $K = 1, h = 2$; ———, $K = 1, h = 3$; ———, $K = 1, h = 4$; (e) ----, $K = 2, h = 2$; ———, $K = 2, h = 3$; ———, $K = 2, h = 4$; (f) ----, $K = 3, h = 2$; ———, $K = 3, h = 3$; ———, $K = 3, h = 4$

Table 3

1 2 3

假警報率假設下給定 3%,6%,12%,18%和24%單位是真時超控-相對應的用五, 10,20,30和40分別被信任。K, 超控偏移被檢測的大小(測量管制內的多重標準差); h, 門檻值管制圖統計量將產生警報; $\widehat{FDR}_6|p_0$, 觀察真實超控單位六年假警報估計百分比給定為100(1-p0)% ; \widehat{SDR}_6 , 六年內成功檢測超控單位估計百分比。169的超控比例等於100(1 -P0)%

Magnitude of shift to be detected, K	Magnitude of threshold, h	Multiple chart performance measure					
		\widehat{SDR}_6 (%)	$\widehat{FDR}_6 p_0$ (%) for the following numbers of units truly out of control‡:				
			5 (3%)	10 (6%)	20 (12%)	30 (18%)	40 (24%)
1	2	83.7	82.4	69.5	51.7	39.9	31.6
	3	67.3	59.2	41.3	24.8	17.0	12.5
	4	48.6	30.8	17.7	9.2	5.9	4.2
2	2	93.7	40.4	24.7	13.3	8.7	6.2
	3	82.6	8.6	4.4	2.1	1.3	0.9
	4	67.8	1.1	0.6	0.3	0.2	0.1
3	2	96.5	0.7	0.4	0.2	0.1	0.08
	3	89.0	0.4	0.2	0.1	0.05	0.04
	4	78.5	<0.01	<0.01	<0.01	<0.01	<0.01

Table 4

委託警報信號百分比在1995年和2000年不同的
h值(警報門檻值)和K(檢測死亡率增加標準殘差)

<i>K</i>	<i>h</i>	<i>Number of trusts signalling (%)</i>	<i>K</i>	<i>h</i>	<i>Number of trusts signalling (%)</i>
<i>No adjustment for overdispersion (c = 1)</i>					
1	3	87 (51.5)	-1	-3	73 (43.2)
1	4	76 (45.0)	-1	-4	69 (40.8)
2	3	71 (42.0)	-2	-3	63 (37.3)
2	4	62 (36.7)	-2	-4	55 (32.5)
<i>Adjusted for overdispersion (c = 9.7)</i>					
1	3	18 (10.7)	-1	-3	24 (14.2)
1	4	11 (6.5)	-1	-4	16 (9.5)
2	3	6 (3.6)	-2	-3	10 (5.9)
2	4	2 (1.1)	-2	-4	7 (4.1)

Fig. 2

(a) CUSUM charts and (b) standardized residual all-cause mortality by year for the six NHS trusts signalling on a $KD2, hD3$ chart at any time between 1995 and 2000, and (c) CUSUM charts and (d) standardized residual all-cause mortality by year for the six NHS trusts signalling on a $KD2, hD3$ chart at any time between 1995 and 2000 (, year when the threshold was first crossed by each trust)

