



臨床試驗調整設計之統計指導原則

第一版

中華民國 111 年 10 月 25 日

財團法人醫藥品查驗中心

序言

為彌補傳統試驗設計的不足，近年來調整設計 (adaptive design) 快速發展，它允許試驗執行過程中，根據試驗中受試者累積的資料，前瞻地規劃去修改試驗設計，提升新藥試驗成功的機率。因此，查驗中心參考 US FDA 指引 Adaptive Designs for Clinical Trials of Drugs and Biologics (2019)，制定「臨床試驗調整設計之統計指導原則」，作為臨床試驗調整設計規劃的參考依據。

【撰寫團隊】

陳瑱芳審查員、蔡貴鳳小組長、宮玫芬資深審查員、吳雅琪資深審查員、林資荃資深審查員、李雅巽審查員、方偉泉審查員、楊雅淇專案經理、詹明曉組長、賴怡君主任、陳可欣主秘、徐麗娟副執行長、林時宜執行長

目錄 (Table of Contents)

1	引言與適用範疇	5
2	調整設計的描述和動機	5
2.1	定義	5
2.2	重要概念	5
2.3	潛在的優勢和範例	6
2.4	限制	8
2.5	選擇調整	9
3	調整設計的原則	9
3.1	控制錯誤結論的機會	9
3.2	估計療效	10
3.3	試驗規劃	10
3.4	維持試驗的進行和完整性	11
4	基於非比較性資料之調整設計	11
5	基於比較性資料之調整設計	12
5.1	群組序貫設計	13
5.2	樣本數調整	14
5.3	病人群體的調整 (例如，調整性富集設計)	15
5.4	治療組選擇的調整	16
5.5	病人分派的調整	17

5.6	選擇指標的調整	17
5.7	多個設計特性的調整	18
6	特別的考量和主題.....	18
6.1	規劃調整設計時的模擬 (Simulation)	18
6.2	貝氏調整設計	19
6.3	事件發生時間設置下的調整 (Adaptations in Time-to-Event Settings) 20	
6.4	基於潛在替代或中間指標的調整 (Adaptations Based on a Potential Surrogate or Intermediate Endpoint).....	20
6.5	次要指標	21
6.6	安全性考量	21
6.7	早期探索性試驗之調整設計.....	22
6.8	基於比較性期中結果之計畫外設計變更	22
6.9	基於試驗外來源的訊息進行設計的變更	22
7	維持試驗的完整性.....	23
8	法規考量	24
8.1	與法規單位的互動	24
8.2	執行調整性試驗前的文件準備.....	25
8.3	評估與報告已完成的試驗.....	26
9	參考文獻	27

本指導原則係參考 US FDA 指引 Adaptive Designs for Clinical Trials of Drugs and Biologics (2019)之規範，亦代表醫藥品查驗中心 (Center for Drug Evaluation, CDE) 對此議題的當前想法。如果未來有相關的科學證據，則會進一步修訂此指導原則。本指導原則非審查基準，若有不同的研發方式，可透過諮詢管道與查驗中心討論。凡涉及政策方向及法規解釋與適用，仍應依中央衛生主管機關之指示為準。

若對此指導原則有任何疑問，歡迎來信至 feedbackbox@cde.org.tw

1 引言與適用範疇

本文件為申請新藥臨床試驗 (INDs)、新藥查驗登記 (NDAs)、生物製劑許可 (BLAs)、或其他補充申請 (supplemental)之廠商和申請者提供指導，使他們適當地使用調整設計於臨床試驗中，以提供藥品或生物製劑有效性和安全性的證據。本指導原則針對調整性臨床試驗，描述設計、執行、及結果報告的重要原則，亦對於具調整設計的臨床試驗，包含設計須仰賴電腦模擬之貝氏調整 (Bayesian adaptive)和複雜的試驗，須提交法規單位以供評估之資料的類型給予廠商建議。

本指導原則主要著重於旨在支持藥物有效性和安全性之臨床試驗的調整設計，其包含的概念亦適用於早期或探索性臨床試驗，以及為滿足上市後承諾或要求的試驗。

2 調整設計的描述和動機

2.1 定義

就本指導原則而言，調整設計定義為一種臨床試驗的設計，允許根據試驗中受試者累積的資料，前瞻地規劃去修改試驗設計。

2.2 重要概念

以下為本指導原則使用的重要概念之描述：

- **期中分析**¹為試驗進行中，對由試驗受試者獲得的資料進行任何的檢視，並不限於正式的組間比較。期中分析使用之觀察資料可包含一個或多個類型，例如基線資料、安全性結果資料、藥物動力學、藥效學或其他生物標記資料、或療效結果資料。

¹ 「臨床試驗統計指導原則」中對期中分析的定義為『在試驗正式結束前，於任何時間點為了比較療效或安全性結果，所執行的任何分析。』本指導原則對於期中分析使用更廣泛的含義，以包含可用於試驗調整決定以累積資料進行的廣泛分析。

- *非比較性分析*為對試驗累積資料的一種檢視，其中受試者分派的治療組別不會以任何方式使用於分析。*比較性分析*則為對試驗累積資料的另一種檢視，其中受試者分派的治療組別或以真正分派的治療或以代碼（如標示為 A 和 B，而不透露那一個為研究治療）來識別²。*解盲分析*和*盲性分析*這兩個詞有時也用來區分可識別和不可識別治療分派的分析。本指導原則避免使用*解盲分析*和*盲性分析*這兩個詞，因為它們可能會造成誤導，將“知曉隨機分派的組別”和將“分派的組別納入調整運作之中”二者混為一談。期中分析不管受試者、研究者、和其他人員如廠商和數據監測委員會（data monitoring committee, DMC）成員是否知道每個人分派的治療，是否有權取得治療組別間比較的結果，都可能為比較性分析或非比較性分析。舉例來說，即使在一個開放設計的試驗，亦可能根據一個非比較性分析來做調整，然而確保調整完全不受到所知比較性資料的影響，這會帶來額外的挑戰。有關限制取得比較性期中分析結果的重要性，將於本指導原則章節 7 做詳細的討論。
- 就本指導原則而言，*前瞻*一詞的意思為以試驗累積資料進行任何比較性分析前，已規劃調整並詳細設定。幾乎在所有的情况，潛在的調整性設計修改，都應在試驗開始前規劃，並在臨床試驗計畫書（和一個分開的統計分析計畫書）中描述。
- 本指導原則將旨在提供實質有效性證據的試驗，和其他所謂的*探索性試驗*之間作一個區別。這個區別取決於臨床試驗的多項特性，例如主要評估指標的臨床關聯性、試驗執行的品質、控制錯誤結論可能性的嚴謹度、以及估計的可靠性。
- *固定樣本數試驗*為一項臨床試驗，設計階段即指定目標總樣本數或目標總事件數³，且不會受到前瞻規劃調整的影響。
- *非調整性試驗*為沒有任何前瞻規劃對試驗設計做修改可能之臨床試驗。
- *偏差*為療效的估計傾向系統性地偏離其真實值。
- *可靠性*為由臨床試驗所得的統計推論，評估療效準確及精確的程度。
- 臨床試驗證明藥物有效性和安全性（在某些情况）的關鍵要件為虛無假說的檢定。於一個特定的顯著水準（通常為單邊水準等於 0.025），假若虛無假說被推翻，證明藥物具臨床上有意義的效果，則此證據通常可以支持有效性的結論。然而，即使藥物是無效的，有時也會推翻虛無假說，這就稱為*型一誤差*。通常有多個虛無假說為真的情境，我們使用*型一誤差機率*這個詞，代表所有這些情境中推翻虛無假說的最大機率。

2.3 潛在的優勢和範例

與非調整設計相比，調整設計可提供多個優勢。這些優勢來自於臨床試驗具調整設計的基

² 非比較性分析和比較性分析二詞是多組別臨床試驗設置下的定義。於單一組別臨床試驗，任何以試驗累積資料進行的分析都包含治療分派的訊息，因此，對本指導原則而言，視為相當於比較性分析。

³ 當感興趣的主要結果為發生事件（如，死亡）的時間，試驗統計檢定力由觀察的事件數而非樣本數來決定。

本特性：它們容許試驗去校正試驗開始時尚未具有的訊息。具體的優勢本質取決於科學背景與所考慮的調整類型，潛在的優勢不外乎以下幾大類：

- 統計上的效率：與類似的非調整設計相比，在某些情況，調整設計提供較多的機會偵測到藥物真正的效果（即，較大的統計檢定力）⁴。例如群組序貫設計（group sequential design）（章節 5.1）與調整性修改樣本數的設計（章節 5.2）。另一方面，調整設計與類似的非調整設計相比，可用較少的預期樣本數⁵或較短的試驗預期執行時間，提供相同的統計檢定力。
- 倫理上的考量：有多種情況調整設計可提供比非調整設計較高的倫理優勢。例如，當試驗無法證明有效性的情況變得明朗時，調整設計能提早停止試驗，如此暴露於無效研究性治療之不必要風險的病人數可以減少，且使受試者有機會探索較有希望的治療選擇。
- 增進對藥物效果的瞭解：相較於非調整設計，調整設計較可能回答更廣泛的問題。例如，調整性富集設計（adaptive enrichment design）（章節 5.3）可能在一個給定的病人群體，或該病人群體之目標次群體證明藥物的有效性；然而若代之以非調整設計，則可能需要極不可行的大樣本數。調整設計也可增進對實驗性治療效果的瞭解。例如，具調整性劑量選擇的設計（章節 5.4），可能得到較佳的劑量反應關係的估計，此亦可能導致後續更有效率的試驗。
- 利益相關者的接受性：與類似的非調整設計相比，調整設計由於增加了靈活性，較可能為利益相關者所接受。例如，廠商可能較願意承諾進行依累積訊息去規劃設計的修改之試驗；病人也許較願意加入採用依反應調整隨機分派（response-adaptive randomization）（章節 5.5）的試驗，因為這些試驗能增加受試者被分派至較有效治療的機會。

以下具調整設計的臨床試驗範例，說明了某些潛在的優勢：

- 執行一項評估 Eliprodil 用於治療患有嚴重頭部損傷病人的臨床試驗（Bolland et al. 1998）。試驗主要評估指標為治療六個月後病人功能狀態定義的三個類別結果。由於試驗設計階段，對於安慰劑對照組病人預期在三個不同功能結果的比例相當不確定。為了可能增加樣本數，預先設定進行一次基於合併、非比較性資料的期中分析，以更新此等比例的估計。選擇此方法為避免試驗的統計檢定力不足，從而幫助確保試驗有效、可靠地達到其目標。該期中分析最終導致試驗樣本數由 400 名病人增至 450 名。
- PARADIGM-HF 為一項執行於慢性心臟衰竭且心室射出分率降低病人的試驗，旨在比較 LCZ696（為 neprilysin 抑制劑 sacubitril 與 renin-angiotensin system 《RAS》抑制劑 valsartan 的合併藥）與 enalapril（RAS 抑制劑）對於心血管死亡或心衰竭住院複合指標的風險（McMurray et al. 2014）。試驗設計規劃 3 次期中分析，分別在總規劃事件數的三

⁴ 非調整設計的一個範例為固定樣本數設計，其樣本數與調整設計預期樣本數相等。

⁵ 預期樣本數為假若試驗重覆多次的平均樣本數。

分之一、二分之一、及三分之二時進行，且根據比較的結果，若 LCZ696 療效顯著優於 enalapril，可能停止試驗。相對於觀察到固定總事件數之後才進行單一次分析的試驗，具有療效停止規則之期中分析，亦可減少試驗預期樣本數和預期持續時間，而能維持相似的試驗成功機率。PARADIGM-HF 於第三次期中分析後停止試驗，該次分析顯示 LCZ696 極顯著地優於 enalapril，且超越了預先設定的停止邊界值 (boundary)。由此可見群組序貫設計比固定樣本數設計較可能快速地確定效益。

- 為評估九價人類乳突病毒 (HPV) 疫苗的安全性和有效性，執行一項具調整性劑量選擇的臨床試驗 (Chen et al. 2015)。試驗受試者經隨機分派分別接受九價 HPV 疫苗三個劑量配方之一或 HPV 四價有效對照疫苗，試驗執行一次期中分析以選擇三個劑量配方之一進入第二階段。試驗的目的為選出一個合適的劑量，並及時確認該劑量的安全性和有效性。
- STAMPEDE 為一項臨床試驗，藉由比較標準雄激素剝奪療法 (ADT) 和幾種不同的治療處方 (ADT 合併一個或多個已核准的療法)，來瞭解一般醫療常規，同時評估前列腺癌的多種治療方法 (Sydes et al 2012)。試驗設計包含多次的期中分析，根據比較結果可能放棄表現不佳的治療組。當同時評估多種治療時，使用共同對照組和以逐次分析終止治療組的設計，會比多項單獨試驗更能達到較高的效率。
- PREVAIL II 為一項臨床試驗，旨在評估 ZMapp 加上現行標準照護，與單獨使用現行標準照護相比，用於治療具有伊波拉 (Ebola) 病毒疾病的患者 (PREVAIL II Writing Group et al. 2016; Dodd et al. 2016)。試驗採用了一種新穎的貝氏 (Bayesian) 調整設計，期中分析和最終分析有效性結論 (即，標準照護加上 ZMapp 能降低 28 天死亡率) 的決策規則，乃基於貝氏後驗機率 (posterior probability)。試驗的期中分析，規劃於每二名病人完成試驗時進行，惟在收錄最少病人數 (每組 12 名) 前，皆不可能採取任何作為。試驗設計亦容許增加試驗藥物為新的治療組，以及可能以試驗執行期間確定為有效的藥物補充或取代現行的標準照護。

2.4 限制

以下列出臨床試驗採用調整設計的一些可能的限制：

- 調整設計需要特定的分析方法，以避免增加錯誤結論的機會和產生偏差的估計。對於複雜的調整設計，此類方法可能不易獲得，而模擬 (simulation) 往往是關鍵 (章節 6.1)。
- 在某些方面效率的提高可能會被其他方面的損失所抵銷。例如，相對於非調整設計，調整設計最少樣本數和預期樣本數可能會減少，然而最大樣本數卻會增加⁶。此外，預先規劃的調整設計修改，可能在設計階段須要付出更多的努力，從而導致規劃與開始試驗之間的前導期時間變長。

⁶ 最少和最多樣本數分別為在調整設計下若試驗重複多次可能選擇的最小和最大樣本數。

- 調整設計的使用增加了後勤方面的挑戰 (logistical challenges)，以確保試驗適當地進行和試驗的完整性 (integrity)。尤其是，適當限制取得比較性期中分析結果的方法可能非常複雜，並且會增加試驗的費用。此外，確保能及時提供高品質的期中資料，以使調整的決策基於最新且可靠的結果具有挑戰性。
- 經由調整以提高效率的機會，可能因重要科學上的約束或在某種臨床環境下會被限制。例如，為了得到關於安全性可靠的評估，可能會有最少樣本數的要求。若感興趣的主要結果，比收錄大多數或全部病人進入試驗，需要更長的時間才能確定，則使用某些類型的調整也可能有所限制。
- 對試驗設計做調整性的變更，可能導致調整後的結果不同於調整前的結果，如此對結果的可解釋性，可能帶來挑戰。

2.5 選擇調整

一般而言，臨床試驗中是否使用調整設計，取決於很多因子，包括前一節所描述潛在的優勢和劣勢，亦可能有各種非科學的考量。簡而言之，設計一項臨床試驗為一個複雜的過程，本指導原則非意圖在一般或特殊狀況下要求或限制調整設計的使用。

3 調整設計的原則

一般而言，提供實質有效性證據之調整性臨床試驗，其設計、執行與分析應滿足四個關鍵性原則：錯誤結論的機會應充分地控制、療效的估計應足夠地可靠、設計的細節應完全預先設定、以及試驗的完整性應適當地維持。雖然所有旨在提供實質有效性證據的臨床試驗皆應滿足此四項原則，以下各節特別針對調整設計概述其考量。

3.1 控制錯誤結論的機會

由於臨床試驗在上市前決策中扮演核心的角色，因此對任何考慮中的試驗設計評估以下機率是重要的：導致安全或有效之錯誤結論、導致缺乏安全或有效之錯誤結論、或其整體效益風險評估之誤導性估計。例如，調整特性可以透過許多方式擴增試驗的型一誤差機率，最明顯的範例為執行多重統計假說檢定的情況。試想一項群組序貫設計，於規劃的總受試者 50% 完成試驗後執行初步的檢定，以便因療效可能停止試驗。假若試驗沒有提早停止，一旦規劃的總受試者 100% 完成試驗時，就會進行最終分析。假若此兩次檢定每一次都執行於常規單邊 0.025 的顯著水準，且藥物是無效的，則試驗產生的整體型一誤差機率將超過 2.5%。這是一個眾所周知的問題，且存在各種方法去決定適當的期中和最終分析之顯著水準，以確保加總後試驗整體型一誤差機率控制於 2.5% (Jennison 與 Turnbull 1999)。

明確的多重假說檢定並非導致錯誤結論的唯一調整設計特性。試想單純的調整性選擇病人群體的方法，在試驗中途會檢視整體試驗群體與次群體的資料，且在那時間點選擇具有較高療效的群體繼續研究。假若在 0.025 的顯著水準於選擇的群體執行最終分析，且分析包含用於選擇那個病人群體相同的資料，則型一誤差機率將會超過 2.5%。其他調整設計特性也許會引入更多可能的型一誤差機率的擴增。

因此，試驗包含虛無假說檢定的調整設計方案，應解決型一誤差機率擴增的可能性。在某些情況，如簡單的群組序貫設計 (章節 5.1)，統計理論可用於推導顯著水準，以確保型一誤差機率控制於預期的水準。在其他的情況，如基於非比較性期中分析的樣本數重估計 (章節 4)，可以證明在常規 0.025 顯著水準下執行分析，對於型一誤差機率沒有影響或影響有限。另外，例如許多貝氏調整設計 (章節 6.2)，使用模擬去評估 (章節 6.1)錯誤結論的機會是至關重要。

3.2 估計療效

臨床試驗重要的是能產生足夠可靠的療效估計，以提供效益風險評估，且於新藥仿單中適當地標示，從而使實證醫學 (evidence-based medicine)的實踐成為可能。某些調整設計的特性可能導致療效和相關數量估計的統計偏差。例如，上述章節 3.1 所提及型一誤差機率擴增的兩個案例，每一個都有可能產生偏差的估計。特別是傳統於試驗結束時的療效估計如樣本平均值，若未將調整納入考量，將傾向於高估母群體真正的療效。這不僅發生於主要指標，也發生於與主要指標相關的次要指標。此外，主要和次要指標的信賴區間，可能對真正的療效沒有正確的覆蓋機率 (coverage probability)。

某些調整設計已知有方法可以校正估計，以減少或去除調整所造成的相關偏差，並改善如均方差 (mean squared error)⁷測量的表現 (如，Jennison 與 Turnbull 1999；Wassmer 與 Brannath 2016)。這些方法當它們可用時，應前瞻地規劃，並使用於結果報告。目前對於調整設計估計的偏差不如對型一誤差機率擴增的研究了解那麼透徹。然而，其他設計可能沒有可用的偏差處理方法。對於這些其他的設計，應該評估其估計偏差的程度，且呈現療效估計與相關的信賴區間時，解讀應適當且謹慎。

3.3 試驗規劃

一般而言，與任何臨床試驗相同⁸，調整設計的細節，於試驗開始執行前應完整地設定，並依此保有紀錄。而前瞻的計畫，應包含預先設定期中分析預期的次數和時間、調整的類型、使用之統計推論的方法、及操控調整決策的特定規則。出於各種原因，完整的預先設定很重要。首先，對於許多類型的調整，假若調整決策的某些方面沒有規劃好，則一旦資料收集完畢，可能沒有適當的統計方法可以控制錯誤結論的機會和產生可靠的估計。其次，完

⁷ 均方差為對估計表現的測量，包含偏差和變異性。

⁸ 「臨床試驗統計指導原則」建議所有臨床試驗預先設定試驗設計和分析計畫。

整的預先設定有助於增加信心，確認調整並不是由計畫外方式累積的知識所下的決策。例如，試想一項臨床試驗，根據合併的、非比較性期中分析的變異數估計（章節 6）來規劃樣本數重估計，而其參與調整決策的人員（如監測委員會）可取得比較性期中分析的結果。預先設定包含修改樣本數的確切規則，將可減少人們對調整決策已受到知曉比較性結果影響的疑慮，並排除根據比較性期中結果而使用統計校正來修改（章節 5.2）。最後，完整的預先設定可以激發設計階段的仔細規劃，排除廠商不必要地取得比較性期中資料，且有助於確保數據監測委員會（若其參與調整設計的實施）有效地關注於維護病人安全和試驗完整性之主要責任（章節 7）。

雖然我們建議預先設定調整的規則，惟監測委員會根據全部的資料所給的建議，偶而會偏離預期的規則。若預期此類的靈活性，則應於預先設定的計畫承認偏離預期規則的可能性，概述可能導致此種偏離的因素，並提出無須仰賴嚴格遵守既定規則的檢定與估計方法。當發生完全無法預見的情況時，我們建議盡快與法規單位討論任何可能的設計變更。

3.4 維持試驗的進行和完整性

調整設計可能會產生額外的試驗操作的複雜性。知曉累積資料會影響試驗的過程和進行，以及廠商、試驗主持人、和參與者的行為。而這些影響是難以預測且無法校正的。因此，強烈建議所有的臨床試驗（調整性和非調整性），應僅限於具有相關專業知識且獨立於試驗進行和管理的人員，才有權取得比較性期中結果⁹。當試驗為調整設計時，維持比較性期中結果的機密性特別具有挑戰性。調整性試驗可能出現問題之兩個範例為：

- 假若試驗主持人不恰當地取得期中分析的比較性結果，根據不可靠的資料得知估計的療效很小或不樂觀，可能錯誤地解讀為治療無效的可靠證據，從而造成順從性的降低和留住病人的努力減少，導致剩餘試驗缺失資料數量的增加。
- 一項根據比較性結果進行樣本數重估計的設計（章節 5.2），期中分析後試驗主持人與潛在的試驗受試者得知試驗目標樣本數增加，可能解讀為期中療效低於預期，如此可能影響未來受試者的收納且危及試驗的成功。

諸如此類的問題，一旦資料收集完畢通常無法校正，因此規劃調整設計，應針對試驗執行的問題，其可能的來源和所導致的後果有所考量，並規劃去避免此等問題。其規劃應描述控制資料取得的程序，和在整個試驗過程記錄所有資料的取得。此於章節 7 有更詳細的討論。

4 基於非比較性資料之調整設計

⁹ 「臨床試驗統計指導原則」也傳達此項建議。

本章節介紹之調整性臨床試驗設計，其調整完全基於非比較性資料的分析，即不使用治療組別分派的資訊。此等分析有時稱之為盲性 (blinded) 或遮蔽 (masked) 分析。一般而言，基於非比較性資料且充分預先設定的調整，對型一誤差機率沒有影響或影響有限。這使得它們在許多情況下成為一個具吸引力的選擇，特別是事件機率或指標變異性之不確定性很高時。

結果資料 (outcome data) 的累積，對於試驗的調整可以提供有用的基礎。結果資料的分析未使用治療分派資訊，有時稱之為合併分析 (pooled analysis)。基於合併結果資料的調整設計 (有時稱之為盲性樣本數重估計) 為樣本數調整最廣泛使用的類別之一。臨床試驗樣本數的計算，取決於幾個因素：預期的顯著水準、預期的檢定力、治療分派產生的結果差異之假設或目標、及其他非感興趣參數 (nuisance parameters) – 非主要關注的數值卻可能影響統計比較。於二分法結果 (例如，反應或不良事件) 的試驗，對照組的反應或事件的機率通常被視為非感興趣參數。而在如症狀分數之連續結果的試驗，分數的變異數為非感興趣參數。藉由使用關於非感興趣參數的累積資訊，可根據預設的演算法來校正樣本數，以確保預期的檢定力得以維持。在某些情況，由於參數本身依賴治療分派的知識，這些方法會以統計模式來估計非感興趣參數值 (Gould 與 Shih 1992)。此等調整通常不會擴增型一誤差機率；然而，於不劣性或相等性假說檢定的試驗，則有少許型一誤差機率的增加 (Friede 與 Kieser 2003)。廠商應評估於此等情境下其擴增的程度。

調整基於合併結果資料的另一個範例為預後富集策略 (prognostic enrichment strategy) 下，規劃於期中重新評估一種生物標記 (biomarker) 或其他基線特徵的預後強度 (prognostic strength)。例如，一項試驗的目標可能是收納較多具某種生物標記的病人，以增加指標事件數。然而，期中合併結果資料可能顯示該生物標記對於合併事件率的影響未如預期，這可能會導致招募策略的改變。

5 基於比較性資料之調整設計

本章節討論不同類型的臨床試驗設計，具有預先設定的基於比較性資料期中分析結果停止試驗或修改試驗設計的規則。此期中分析有時稱之為解盲分析。通常有一些重要的觀念適用於以下各節。首先，與基於非比較性資料的調整不同，基於比較性資料的調整通常會直接增加型一誤差機率，並導致療效估計的偏差。因此，採用的統計方法須將調整性試驗設計納入考量。其次，當調整是基於比較性期中分析時，採取額外的步驟以確保試驗執行適當至關重要。此點會於章節 7 中詳細討論。最後，停止試驗或調整試驗設計的規則，可以各種不同的尺度來設定。例如，療效的估計、固定樣本的 p 值、試驗成功的條件機率、藥物有效之貝氏後驗機率、或試驗成功之貝氏預測機率 (Bayesian predictive probability) 等。只要能充分地評估設計的操作特徵 (operating characteristics)¹⁰，尺度的選擇相對並不重要。

¹⁰ 試驗的操作特徵為具有給定設計之試驗的性質。例如，較感興趣的性質包括：型一誤差機率、檢定力、預期最小和最大樣本數、療效估計偏差、及信賴區間覆蓋率 (即，若試驗重複多次，信賴區間包含真正療效的機率)。

5.1 群組序貫設計

群組序貫設計試驗，容許前瞻地規劃一次或多次的比較性資料期中分析，具有預先設定的停止試驗標準。試驗包含逐次分析，藉由減少預期樣本數和試驗時間，以及加速安全且有效之新治療的核准，可以提供倫理和效率上的優勢。例如，一項具有單次期中分析和慣用療效停止邊界值 (boundaries) 的群組序貫設計，相較於固定樣本數之比較性試驗，可減少約 15% 的試驗預期樣本數¹¹。

群組序貫設計的停止試驗規則，可用於當試驗具備充分療效證據足以支持法規決策時，或當證據顯示無法證明療效時，後者通常稱為因無效停止試驗 (stopping for futility)。群組序貫設計，以常規單邊 0.025 顯著水準，執行多個針對療效的統計假說檢定，將擴增型一誤差機率，從而增加錯誤結論的機會。決定期中與最終分析的停止邊界值，以適當地控制型一誤差機率的方法有許多種。例如，O'Brien-Fleming 方法傾向於要求極具說服力的早期結果才能因療效停止試驗 (O'Brien 與 Fleming 1979)。其他如 Pocock 所提出的方法，要求較低說服力的早期結果，而有較高的機率提早停止試驗 (Pocock 1977)。這些和其他方法均依賴前瞻地規劃期中分析的次數和分析執行時的目標特定樣本數或事件數。

Lan-DeMets 之 alpha 消耗函數¹²方法，藉由設定整個試驗如何消耗型一誤差機率函數，可容許各種不同級別的證據以提早停止試驗，同時還允許靈活決定期中分析的次數和時間點 (Lan 與 DeMets 1983)。靈活的時間點有助於監測會議排定於特定的日曆時間，而非於期中樣本數或事件數達到特定目標時。分析次數的靈活性有助於配合病人收納速度快於或慢於預期。然而，期中分析時間點的選擇若基於累積的比較性結果，可能會擴增型一誤差機率。例如，由於最近一次期中分析的結果接近停止邊界值，就調整下次期中分析的時間比原規劃更早發生是不合適的。由於 Lan-DeMets 之 alpha 消耗函數方法有此潛在的問題，廠商應採取額外的保障措施，例如擬定期中分析的目標次數和它們發生的大約時程，以及在試驗開始後修改分析次數或時間之決策框架。決策框架應基於統計上與療效估計無關的資訊 (例如，收納速度或後勤的調度)。舉例來說，決策框架可以設定半年一次的期中分析，假若病人收納遠低於預先設定的目標，則規劃進行額外的分析。

為了確保具有適當的設計、執行和分析，群組序貫試驗還有許多額外的考量。首先，為使群組序貫方法有效，遵守前瞻分析計畫且僅於滿足停止標準時才可因療效終止試驗是非常重要的。其次，因無效提早停止試驗的方針應執行適當。試驗設計通常採用無約束力 (nonbinding) 之無效規則，依此方針無效停止標準可能會也可能不會被遵循，具體取決於可用的期中結果的整體情況。將此無約束力的無效方針附加於固定樣本的試驗，或具有適當的群組序貫療效停止規則的試驗，不會增加型一誤差機率且通常是適當的。反之，群組序

¹¹ 相較於所有病人收錄完畢且結果都已確定時才執行單一次分析之設計，一項群組序貫設計包含一次期中分析，於最高病人數一半其結果訊息可用時進行，且使用 O'Brien-Fleming 療效停止邊界，若對立假說 (具 90% 檢定力) 為真，試驗預期樣本數大約可減少 15%。

¹² 臨床試驗型一誤差機率通常以希臘字母 α (alpha) 表示。

貫設計可能包含具約束力 (*binding*) 的無效規則，如果滿足無效標準時，試驗一定要停止。具約束力的無效規則可提供試驗療效分析的一些好處 (例如，放寬確定療效的門檻)，但只有在遵循停止規則時才能控制型一誤差機率。因此，一項試驗儘管符合預先設定的具約束力之無效規則卻仍繼續進行，不管最終分析的結果如何，法規單位可能會認為該試驗無法提供療效證據。亦要注意有些數據監測委員會可能較喜歡不具約束力無效方針的靈活性。

第三，試驗因療效提早停止，將會有較少的樣本數用於評估安全性和潛在重要的次要療效指標。因此，因療效提早停止試驗的設計通常保留給存在令人信服的倫理原因之情況 (例如，主要指標為存活或不可逆的發病事件)，或停止試驗規則要求估計的療效大小和其證據強度具有高度說服力的情況。在某些情況，為了可靠地評估安全性而有預期最低樣本數的要求，因此群組序貫期中分析可以如何早進行或能否進行可能有所限制。例如，預防性疫苗試驗通常會有此限制。

最後，傳統固定樣本的療效估計如樣本平均值，當使用群組序貫設計時，往往會有效果比其真實值來得大的偏差。同樣地，信賴區間也不會有預期名義上的 (*nominal*) 覆蓋機率。因此，存在各種不同計算估計值和信賴區間的方法，俾能針對群組序貫停止規則適當地校正 (Jennison 與 Turnbull 1999)。為確保試驗結果於科學上和統計上的可信度，並有助於重要的效益風險考量，計算估計值與信賴區間的方法將群組序貫設計適當地納入考量，此應前瞻性地規劃並用於結果的報告。

5.2 樣本數調整

樣本數調整方法之一為前瞻地規劃，於期中分析使用治療組別的訊息，得到非感興趣參數的估計值，來修改樣本數。例如，結合個別治療組變異數的估計值，來估計連續結果的變異數，或僅根據對照組的資料來估計該組二分法結果的機率。這些方法通常對型一誤差沒有影響，或影響有限。然而，不同於基於非比較性合併資料所得之非感興趣參數期中估計值 (章節 4) 所進行的調整，此類調整方法需要有治療組別的訊息，也因此要求額外的步驟來維持試驗的完整性 (章節 4)。

另一種樣本數調整方法為前瞻地規劃，基於比較性期中分析結果 (即，療效之期中估計值)，來修改樣本數。這通常稱為解盲樣本數調整 (*unblinded sample size adaptation*) 或解盲樣本數重估計 (*unblinded sample size re-estimation*)。樣本數的決定，取決於多個因素，例如對照組的事件率或主要結果的變異、型一誤差機率、假想之療效大小、以及偵測此療效大小預期的檢定力。於章節 4，我們描述了基於非比較性期中結果的調整方法，以應對設計階段對於結果的變異或對照組事件率的不確定性。反之，若對於療效大小的真實值不確定時，可使用基於比較性期中結果之樣本數調整設計。此方法與群組序貫設計試驗相同，在合理的療效大小範圍內，可提供足夠的檢定力，也因此若真實療效低於預期但仍具有臨床意義時，可確保試驗維持足夠的檢定力。再者，預先設定樣本數的修改規則，可提供某些操作特徵效率上的優勢。

在沒有適當校正的情況下，任意修改試驗樣本數，可能會擴增型一誤差機率。試想一個具

有一次期中分析的設計，使用期中療效估計值來修改最終樣本數。若於試驗結束時以常規的 0.025 顯著水準執行假說檢定，其型一誤差機率可能增加一倍以上 (Proschan 與 Hunsberger 1995)¹³。因此，此類調整設計應使用可行方法，以適當地控制型一誤差機率。例如，目前的假說檢定方法，以預先規劃的方式，合併試驗不同階段的檢定統計值或 p 值，或保留條件型一誤差的機率 (如，Bauer 與 Kohne 1994; Fisher 1998; Cui et al. 1999; Denne 2001; Müller 與 Schäfer 2001; Chow 與 Chang 2011)。此等方法除了針對最大樣本數的調整外，亦包含其他樣本數規劃方面 (例如，未來期中分析的次數與間隔) 的調整。

本指導原則章節 5.1 針對群組序貫設計所討論的調整計畫的順從性、安全性評估、及療效估計之額外考量，亦適用於基於比較性資料之樣本數調整設計。值得注意的是，前瞻性的規劃不僅應包括預先設定的統計檢定方法，亦應包括樣本數修改的規則。最後，在有樣本數調整情況下，維持試驗的完整性還有其他的挑戰。例如，樣本數修改的規則，通常基於試驗結束時療效具統計上顯著之條件機率 (通常稱之為條件檢定力《conditional power》) 能維持在或接近於預期的水準。在此情況下，知曉調整規則並依此調整性地選擇樣本數，將使得回推估算 (back-calculation) 期中的療效估計相當容易。因此，應採取額外的措施來限制相關工作人員有此方面詳細的資訊，以維持試驗的完整性。章節 7 將有額外的討論。

本章節所討論的原則，亦適用於指標為事件發生時間 (time-to-event) 的試驗，其調整設計容許前瞻地規劃基於比較性期中結果修改試驗總事件數。然而，於此設置下，有一些特殊額外的考量，將於章節 6.3 進一步地討論。

5.3 病人群體的調整 (例如，調整性富集設計)

於許多的情況，可能預期療效在試驗群體的某些子集會較好。這些次群體可由人口學特性、或由與藥物作用機轉相關的遺傳或病理生理標記來定義。於此情況下，可以考慮一種允許根據比較性期中結果，對病人群體做調整性修改的設計。例如，一項試驗收納試驗整體受試者群體直至期中分析，在此時間點，根據預先設定的標準，將決定是否繼續收納整體受試者群體，或僅收納侷限於目標次群體。期中分析前與後所累積的資料可以合併，以得出目標群體療效的推論。此類型的設計，通常稱之為調整性富集設計，可提供比非調整性設計更多的優勢。特別是，在相同樣本數下，此項調整設計相較於執行於整體受試者之非調整性固定樣本設計，可提供更高的檢定力¹⁴。再者，與僅收納目標次群體的試驗不同，此調整設計亦可在非目標(互補)次群體上，進行試驗治療的療效評估。

調整性修改病人群體的設計通常包含 (1) 根據比較性期中結果，修改收納的群體或主要分析群體等設計特性；與 (2) 於多個群體，如目標次群體和整體群體，執行假說檢定。因此，統計假說檢定方法應將兩種來源的多重性納入考量。例如，有一種方法以事先規劃的方式合併不同階段的檢定統計值或 p 值，同時也使用適當的多重檢定程序 (Wassmer 與 Brannath 2016)。此方法亦可適用於樣本數調整或特殊次群體病人比例的調整 (如，增加子集收納的

¹³ 意思是即使用 Bonferroni 方法去校正所執行的兩次分析仍然不適當。

¹⁴ 於此情況下檢定力可被定義為，例如，於目標次群體或整體群體成功識別出一個真正療效的機率。

比例而非將收納完全侷限於那個子集)。

除了之前針對群組序貫設計和樣本數調整設計的討論外，還有許多重要的考量。首先，於調整性富集設計，所擬對病人群體調整性修改的動機，應來自於先前 (如，早期) 試驗的結果，或很強生物學上的合理性，認定利益風險概況 (benefit-risk profile) 對於特殊次群體最為有利。其次，若認為療效可能受益者之基線特徵不是顯而易見的二分法，則用以定義次群體的一個或多個閾值，應說明其合理性。第三，目標次群體的識別可能依賴體外診斷器材的使用。在這種情況下，診斷器材應有適當的效能特性 (performance characteristics)。最後，試驗應設計到何種程度以描述互補次群體 (complementary subpopulation) 的療效，可能取決於許多因素。例如富集所依據的病理生理學上或實證學上的合理性，藥物的毒性，定義次群體之基線標記的分布，定義次群體之閾值選取的理由，以及若僅核准於目標次群體則互補次群體於處方外使用的可能性。

5.4 治療組選擇的調整

另一個調整方法為前瞻地規劃，基於比較性期中結果修改納入試驗的治療組，其修改可能包括添加或終止某個或某些治療組。此種設計通常用於早期探索性劑量範圍試驗 (dose-ranging trials)，調整性劑量範圍試驗可能由幾個劑量開始，並結合基於比較性資料進行的期中分析，以選擇繼續評估的劑量。其目的為提供相較於非調整設計對劑量反應關係較佳的描述，並容許選擇一個或多個最佳劑量用於未來確認性試驗。例如，連續重新評估法 (continual reassessment method, CRM) 即為基於觀察的毒性，調整性地遞增早期試驗中所評估劑量的方法，以可靠有效地估計新藥的最大容忍劑量 (maximum tolerated dose) (Le Tourneau et al. 2009)。調整性治療組選擇亦可用於旨在提供實質有效性證據的試驗。例如，當兩個劑量任一個或兩個具有可接受的利益風險時，若採用序貫分析 (sequential analyses) 的調整設計，則可容許提早終止其中一個劑量組，如此相較於其他非調整性設計，可以較有效地達到其科學目標。此調整設計原則上亦容許期中修改試驗設計的其他方面，例如收納額外數目的病人 (樣本數) 或治療組間隨機分派比例。

對於旨在提供實質有效性證據的試驗，統計假說檢定方法應考量試驗由多組評估劑量中，調整性地選擇一個或幾個最佳劑量，以及任何其他調整性的修改，諸如可能提早停止試驗或修改未來的樣本數等。就以簡單情況而言，在多於一個劑量且包含期中分析的試驗中，某一特定劑量可因療效或無效而停止收錄該劑量的病人。此設計於評估多個劑量時，可以使用典型的群組序貫檢定方法，併用多重檢定方法，以控制型一誤差。假若試驗設計還容許其他調整如修改樣本數，則應使用之前針對樣本數和病人群體調整所描述的方法。如同其他的調整設計，前瞻規劃很重要，不僅應預設檢定方法，且應預設治療組選擇和所有可能的期中調整規則。一般而言，如果遵循章節 3 所概述的原則，可考慮採用結合劑量選擇和確認所選劑量療效 (基於整個試驗的資料) 的無縫設計 (seamless design)。

調整性選擇治療組的一種特殊情況為針對一種疾病設計多個試驗組和單一對照組比較之調整性平台試驗 (platform trial) (如，Woodcock 與 La Vange 2017)。為了提高效率，此類試驗

通常包括使用共同對照組和前瞻地規劃於期中分析時選擇有希望的治療組繼續研究等兩項特性。由於此類試驗可能包含來自多個廠商的研究藥品、試驗執行時間的長短可能未明、及通常包含複雜的調整，廠商應與法規單位討論。

5.5 病人分派的調整

本章節討論兩類針對病人分派的調整：基於比較性基線特徵資料的調整和基於比較性結果資料的調整。第一類為依共變數調整之治療分派 (*covariate-adaptive treatment assignment*)，其病人的治療分派部分或全部取決於此人的基線特徵，和之前收錄病人的基線特徵及所分派的治療。使用此方法可促進治療組間基線共變數的平衡。廣為人知的依共變數調整之隨機分派的範例為最小化 (*minimization*)隨機分派 (Pocock 與 Simon 1975)，其分派每一位接續病人至各個治療組，係以將治療組間潛在預後共變數的差異減至最小的方式。當使用適當的方法分析時，例如隨機化檢定 (*randomization test*)或置換檢定 (*permutation test*)，依共變數調整之治療分派技巧不會直接增加型一誤差機率。此項技巧相對於簡單的隨機分派，易增加治療分派的可預測性；惟透過加入額外的隨機性，以避免完全確定的治療分派，可減輕可預測性。

第二類為依反應調整之隨機分派 (*response-adaptive randomization*)，其調整特性為：新收錄的受試者被分派至某一治療組的機會，於試驗過程中會隨著之前收錄的受試者累積的結果資料而改變。依反應調整之隨機分派有許多不同技巧，如勝者優先 (*play the winner*)設計。有時統計、倫理、和務實的考量都可能成為使用依反應調整之隨機分派的理由。在統計方面，依反應調整之隨機分派在某些情況可將檢定統計值的變異數減至最小，從而導致試驗期縮短、較少的樣本數、和/或較高的統計檢定力。就倫理的論點，依反應調整之隨機分派其設計的特性可導致較多的試驗受試者被分派至較有希望的治療組。最後，就務實的論點，具此項設計特性的臨床試驗對潛在受試者可能較有吸引力，進而使收案較快且較容易。值得注意的是，對於依反應調整之隨機分派有其爭議的論點，有些研究者認為不應該使用尚未定論的期中結果，來改變進行中試驗的隨機分派，且雙臂試驗之統計效率並未實質地改善，無法支持隨機分派比例調整的合理性 (Hey 與 Kimmelman 2015，及相關的評論)。

當使用適當的統計分析技巧時，依反應調整之隨機分派本身通常不會增加型一誤差機率。重要的是要確保分析方法將試驗設計適當地納入考量。最後，與其他許多基於結果資料調整的方法相同，依反應調整之隨機分派，對於需要相對較短時間確定結果的試驗效果最佳。

5.6 選擇指標的調整

這項設計容許基於期中分析的比較性結果，調整性修改主要指標。此設計的動機可能來自於，法規單位認可作為主要指標的多個病人結果變數中，其療效的不確定性。與其他調整設計相同的是，應預先設定調整規則，並且統計假說檢定方法應將指標選擇的調整納入考量。由於指標選擇涉及重要的臨床考量，因此若考慮採用此項設計時，建議儘早與法規單

位討論。

5.7 多個設計特性的調整

若結合本指導原則所討論的兩個或多個調整設計特性，此臨床試驗可能較為複雜；惟適用於簡單調整設計的一般性原則，也同樣適用於此等複雜的設計。對於包含多個調整特性的設計，型一誤差機率和其他操作特徵的估計可能特別困難。臨床試驗模擬（章節 6.1）通常對於此類試驗設計的評估至關重要。

6 特別的考量和主題

6.1 規劃調整設計時的模擬 (Simulation)

臨床試驗模擬通常於規劃和設計臨床試驗時扮演關鍵的角色，模擬對於調整性試驗尤其重要，可用於如選擇期中分析的時間和次數，或用於決定檢定統計值之適當臨界值以宣稱療效或無效。模擬對於比較不同試驗設計的表現也很有用。於調整性試驗設計，模擬最重要的用途為估計試驗的操作特徵，並證明這些操作特徵達到預期的水準。

傳統的非調整性臨床試驗，通常依賴統計理論來確保型一誤差機率控制於一個預期的水準，並獲得試驗檢定力的估計。最簡單的例子，於固定樣本數試驗設計檢定單一指標，通常顯示最終分析的檢定統計量具有某個漸近的機率分布 (asymptotic probability distribution)¹⁵，且推論和操作特徵可基於這個分布的性質。許多調整設計，諸如傳統的群組序貫設計，同樣地可由數學推導其漸近的機率分布，並根據此分布來做推論與規劃。

然而，有些調整設計，不是無法推導出檢定統計量的相關分布，就是分布本身在計算上不易處理。這往往是較複雜的調整設計情況，例如調整多個元素，或者使用預測機率模式 (predictive probability models) 去決定分析時間點的設計。於此等情況，試驗操作特徵通常可由臨床試驗模擬的方法來估計。例如，針對型一誤差機率和檢定力，此方法的基本邏輯為根據各種不同的假設模擬出許多試驗實例，並評估在各個假設下之所有模擬試驗，達到預先設定支持有效性結論之百分比。

對於旨在估計型一誤差機率的模擬，假想的臨床試驗將基於虛無假說為真的前提下，進行一系列的參數假說模擬。對於每一個假設集合，模擬試驗導致偽陽性結論的百分比將被視為在那些假設下型一誤差機率的估計。幾乎所有情況都有無數個情境潛在地與虛無假說相符。在估計型一誤差機率時，識別出那些情境應被考慮頗具有挑戰性，可能須依賴醫學與數學結合的考量。

¹⁵ 檢定統計量的漸近分布為當樣本數變大時那個統計值的近似機率分布。

這些情境可能包含關於非感興趣參數不同的假設，而這些非感興趣參數可包括統計參數（例如，症狀分數的變異數或對照組的反應機率）、以及操作參數（例如，受試者收納速度）。舉例來說，考量存活率很低（例如，中位數為 6 個月）的癌症用藥試驗，比較試驗療法與安慰劑組之兩年死亡率。虛無假說為兩組死亡率相等，與這虛無假說一致的可能情境包含相等死亡率為 5%、50%、99%、99.01% 等。雖然不可能去模擬每一個與虛無假說相當的情境，但決定一個能適當代表潛在偽陽性的合理範圍有限情境集合仍然是有可能的。在這個例子，醫學專家可能根據文獻和疾病的臨床經驗，覺得應排除安慰劑組兩年死亡率低於 75% 的情境。在決定須模擬那些情境以估計型一誤差機率時，數學上的考量也可能扮演一定的角色。他們可能主張某些情境必然比其他基於單調性 (monotonicity) 的情境有較低的型一誤差機率。

在許多的情況，即便將臨床與數學都納入考量後，也不可能將每個虛無假說集都去估計型一誤差。通常須在合理的參數組合 (grid) 上進行模擬，在參數組合範圍內，最大型一誤差機率不會超過預期的水準下，根據來自模擬的全部證據進行論證。在上述案例中，可能以安慰劑組和實驗治療組的死亡率均為 75%、80%、85%、90%、95% 及 99% 來進行模擬。若在此等情境下，每一個估計的型一誤差機率都低於 0.025，則被視為有充足證據顯示當安慰劑組的死亡率介於 75% 與 99% 之間，所有情境的型一誤差機率都能充分地控制。然而，無論採用任何方法，試驗結束時的評估，都應考慮統計推論是否適當，及根據非感興趣參數累積訊息所得的結論是否合理。在上述案例中，若安慰劑組觀察到的死亡率意外地為 50%，就需要額外的模擬。

另一個複雜的因子則是存在多個指標。假若廠商預期檢定多個臨床指標，並控制這些指標的家族型一誤差機率 (familywise Type I error probability)，則應執行虛無假說情境下每位受試者所有指標的模擬，並瞭解多個指標間相互的關係結構。這可能是臨床試驗模擬無法解決的複雜問題。然而，在某些情況，可以主張假設多個指標彼此獨立將會得到型一誤差機率的上限。譬如針對多重檢定的控制使用 Bonferroni 或 Holm 方法¹⁶。

6.2 貝氏調整設計

貝氏調整設計係指使用貝氏統計推論或計算的臨床試驗設計 (Berry, et al. 2010)。貝氏調整設計特性的範例如下：

- 使用可能結合試驗外訊息的預測統計模式，來管理期中分析的時間與決策規則。
- 使用假設的劑量-反應關係來控制劑量的遞增和選擇。
- 明確借用外部來源的訊息，例如之前的試驗、既存的歷史研究、以及病人註冊，透過具訊息的先驗分布 (prior distributions) 來提高試驗的效率。

¹⁶ 關於 Bonferroni、Holm、和其他多重檢定方法更多的討論，可於指導原則「臨床試驗多重指標之統計指導原則」尋得。

- 使用後驗機率分布 (posterior probability distributions)來形成試驗成功的標準。

一般而言，貝氏調整設計與不具貝氏特性之調整設計皆適用相同的原則。使用貝氏調整特性的試驗設計可能依賴頻率學派 (frequentist)或貝氏學派推論程序來支持藥物有效性的結論。頻率學派推論的特點為藉由已知的檢定力和型一誤差來執行假說檢定，且通常與依賴不具訊息先驗分布的貝氏運算技巧一起使用。貝氏學派推論的特點為使用後驗分布所得之結論來說明藥物效果，並與頻率學派推論有很重要的差異 (Berger 與 Wolpert 1988)。對於使用具訊息先驗分布 (例如明確借用外部訊息)的貝氏推論之試驗，貝氏統計性質相較於型一誤差機率提供更多的訊息。

許多貝氏調整設計的共通點為使用模擬 (章節 6.1)去估計試驗的操作特徵。由於許多貝氏方法本身依賴大量的運算 (馬可夫鏈蒙地卡羅《MCMC》方法和其他技巧)，因此對於貝氏調整設計而言，試驗模擬可能特別耗費資源 (resource-intensive)。

6.3 事件發生時間設置下的調整 (Adaptations in Time-to-Event Settings)

對於主要指標為某個事件發生的時間 (例如，發生死亡的時間或發生腫瘤反應的時間)之調整性試驗，有些額外的考量。於此等試驗，檢定力取決於事件數而非受試者人數。因此通常目標為達到固定事件數，而非固定受試者人數。此類試驗的樣本數調整以修改事件數為目的，因此，可能採取的形式有修改受試者人數、修改每位受試者追蹤期長度、或者兩者兼而有之。另外，於事件發生時間設置下，期中分析可能會採用替代或中間結果的訊息，分析時應適當地納入考量 (參見下一節有進一步的討論)。

6.4 基於潛在替代或中間指標的調整 (Adaptations Based on a Potential Surrogate or Intermediate Endpoint)

大多數的調整設計依賴對主要指標持續地監測。然而，當存在與主要指標相關的潛在替代或中間指標，並且主要指標本身很難或很久才能確定時，則可以基於替代或中間指標進行調整設計。舉例來說，一項癌症治療試驗，其主要指標為整體存活期 (overall survival)，存活時間中位數遠超過 2 年，並且腫瘤反應或可預測臨床利益。於這個案例，樣本數重估計或其他調整特性基於腫瘤反應而非死亡事件可能是合理的，最終療效評估仍須基於主要指標 (於此例為整體存活期)。同樣地，當主要指標為 6 個月的病人症狀測量，調整設計可以基於相同症狀 2 個月的結果。有些方法假設替代或中間指標與主要指標的相關性，則型一誤差或其他試驗操作特徵的評估應考慮錯誤設定此項相關性的可能影響。其他方法則不依賴替代或中間指標與主要指標相關性的假設 (Jenkins et al. 2011；Irle et al. 2012；Magirr et al. 2016)。

在指標為事件發生時間或長期追蹤結果 (longitudinal outcomes)之調整性臨床試驗，使用期中分析之替代或中間指標訊息可能會增加錯誤有效性結論的機會，除非使用適當的統計分析技巧。例如，Bauer 和 Posch (2004)指出，在指標為事件發生時間的試驗中，期中分析時

使用尚未觀察到事件的受試者之替代指標來預測未來的事件時間，會導致型一誤差的擴增。額外的防護措施，例如限制取得比較性期中結果和僅依賴事先計畫的主要指標調整規則，可以幫助增加信心，確定沒有執行這些計畫外的方法。。參見章節 7 之額外討論。

6.5 次要指標

除了主要指標外，大多數的臨床試驗皆設定一個或多個次要指標，且調整設計可能對這些次要指標的分析產生影響。試想群組序貫設計：一般普遍曉得對主要指標進行多次的分析，可能會擴增型一誤差機率，且導致對該指標的療效估計產生偏差。然而，較少人知道型一誤差機率的擴增和偏差的估計也可能適用於與主要指標相關的任何指標 (Hung et al. 2007)。臨床試驗中大多數的次要指標與主要指標有關，且通常是高度地相關。有些設計譬如群組序貫方法，亦有對次要指標調整之校正分析方法 (Glimm et al. 2010)。若沒有這樣的校正，在解讀次要指標結果時，應當適當地保守。

6.6 安全性考量

雖然調整設計臨床試驗的規劃通常著重於證明有效性的結果，惟安全性目標也扮演重要的角色。首先，在某些情況調整規劃係針對安全性指標而非療效指標。其中一個案例為癌症早期劑量範圍試驗，使用連續重新評估法 (CRM) 或其他調整方法試圖去決定最大耐受劑量。另一個案例為「輪狀病毒療效與安全試驗 (Rotavirus Efficacy and Safety Trial, REST)」，為 2006 年輪狀病毒疫苗 RotaTeq 核准的主要基礎 (Heyse et al. 2008)。REST 為一項群組序貫試驗，以高達 100,000 個嬰兒評估一項嚴重的胃腸道疾病腸套疊 (intussusception) 的風險，僅使用其中的一個子集來評估療效。

其次，針對療效指標調整的試驗，獲得足夠的安全性訊息以支持產品的核准，往往是一個主要的考量。因強而有力的有效性證據而提早停止的試驗，仍須收集足夠的安全性資料，以容許對研究藥物進行可靠的利益風險評估，並為仿單提供訊息。因此，當規劃期中分析的次數、時間點、及停止試驗的邊界值時，應將安全性資料的數量納入考量。特別是，期中分析的時間點可能會受限於所期待的最低受試者人數和最短暴露時間，以確保可靠的安全性評估。

最後，重要的是要考量是否某些調整可能潛在地使試驗受試者處於不必要的風險中，特別是針對早期劑量遞增試驗。

調整規則中容許連續幾組的受試者接受快速遞增的劑量，可能導致受試者接受不安全的高劑量，而這種情況可透過較為漸進的劑量遞增設計來避免。例如當研究藥物可能具有作用延遲 (delayed onset of action) 之嚴重不良事件時，選擇特定的劑量遞增調整規則時應考量遞增的速度。

6.7 早期探索性試驗之調整設計

藥物開發的探索性試驗旨在獲得藥物使用之廣泛訊息，以便引導日後對於如何研究該藥物(例如，劑量、處方、病人群體、併用治療、或指標的選擇)做出最佳的決定。可以為一系列個別的早期試驗，以循序檢視不同方面的藥物效果，或為試圖同時評估多個不同面向藥物效果的較複雜試驗。調整設計容許對廣泛的選擇進行初步的評估，所提供的靈活性在藥物開發的探索性階段特別有用。早期開發試驗運用調整設計去探索關於劑量、暴露、藥效學、病人反應的變異性、或反應調節 (response modifiers)等不同的面向，提供廠商改善設計的機會，進而可能提高晚期試驗成功的機會。

雖然探索性試驗與旨在提供實質有效性的試驗，通常在統計嚴謹度和操作特徵上法規的期望有所不同；然而，於探索性試驗注意到潛在錯誤的結論仍然是重要的。例如，探索性多劑量比較試驗的錯誤，可能導致選擇不理想的劑量用於後續的確認性試驗，結果無法顯示有效性或產生不必要過度的毒性。因此，探索性試驗能夠遵循調整性試驗設計的良好原則，可以降低對開發計畫不利影響的風險。

6.8 基於比較性期中結果之計畫外設計變更

當比較性期中分析時檢視試驗的資料，有些未前瞻規劃做為調整基礎的資料分析可能意外地指示某些設計的變更(例如，分析限於某些群體的子集、放棄一個治療組、校正樣本數、修改主要指標、或變更分析方法)於倫理上是重要的，或者可能增加最終試驗結果統計顯著的可能性。舉例來說，多個治療組的試驗，其中一組意外地缺乏治療的順從性，可能促使放棄該治療組。此種基於非前瞻規劃的修正，可能在控制型一誤差機率和解讀試驗結果上造成困難。在不損害受試者的安全性考量下，建議廠商事先與法規單位討論考慮的變更情況再施行這樣的變更。

6.9 基於試驗外來源的訊息進行設計的變更

在藥物開發計畫過程，發生於進行中試驗外之不可預測的事件，可能提供與進行中試驗相關的重要新訊息，且可能會引起對試驗設計的修訂。例如，可能有意想不到的安全性訊息來自於不同的研究(也許執行於不同的病人群體)、新的有關識別疾病亞型之疾病病理生理學或病人特徵訊息、有關藥物之藥物動力學或藥效學反應的新訊息、或其他在試驗設計時已知可能導致不同試驗設計的訊息。發生這種情況時，可能有理由以某種方式修訂試驗設計，而不是終止現有的試驗，以修改的設計開始新的試驗。於存在嚴重安全性顧慮情況下，特別是大型的試驗，修訂試驗設計以容許試驗繼續進行可能是至關重要的。動機良好且僅根據試驗外訊息之設計變更，不會影響統計推論的有效性，通常法規單位認為是可接受的。事實上除非廠商對於比較性期中結果完全不知情，確保修改試驗的決定完全基於外部的信息是非常具有挑戰性的。這就是為什麼限制取得比較性期中結果如此重要的原因之一。參

見章節 7 的額外討論。

7 維持試驗的完整性

一般來說，會強烈建議，僅限於相關專業人士才能獲得比較性期中結果，且這些人士不能參與試驗的執行或管理。確保病人、試驗主持人與其職員、以及廠商無法取得比較性期中結果，有兩個重要的目的：首先，它提供了最大程度的信心，確定任何非事先規劃的試驗設計修改，非源自於任何累積的資料。例如，若試驗管理人員可獲知比較性期中結果，將使法規單位難以確定，一項非事先規劃的計畫書變更，是否受到累積的比較性資料影響。假若認為設計的變更可能已受到比較性期中結果的影響，用以控制錯誤結論的機會或產生可靠估計的適當統計方法可能無法獲得、或執行上可能具有挑戰性、或可能大幅降低試驗的效率。

其次，限制比較性期中結果的取得，為試驗執行的品質提供最佳的保證。試驗主持人一旦知曉累積的資料，可能對病人的收納、順從性、持續留於試驗、或指標的評估產生不利的影響，從而降低試驗能及時可靠地達到其目標的能力。試驗執行的問題難以預測，且通常無法以統計分析來校正。因此，具調整設計之臨床試驗應該嚴謹地規劃，審慎地執行，並全面紀錄所採取的方法，以維持比較性期中結果的機密性和保持試驗的完整性。

為限制取得調整性試驗之比較性資料，廠商可有多種方式運用於試驗執行中。如果試驗已設立有數據監測委員會，可建立一個（不包含數據監測委員會）獨立的調整委員會。或者，數據監測委員會也可兼任調整性決策的任務，雖然它主要的責任是確保病人的安全性與試驗的完整性。群組序貫設計或其他具有簡單調整演算法之設計，建議採用單一委員會的設計。兩種方式各有其優點，例如，使用兩個分開的委員會，可以促使每一個委員會納入更多相關的專業人士，且使數據監測委員會更有效地專注於其主要職責。另一方面，將數據監測委員會用於此兩個目的，則可避免在多個監測委員會間的訊息共享與互動所帶來的後勤挑戰。

無論選擇何種方法，負責提出調整建議的委員會，應有合適的專業成員，包括一名或多名熟知調整方法學、數據監測計畫、和決策規則的統計學家。此外，該委員會的職責應該是在一項經審慎設計和預先設定之調整計畫的基礎上，提出適當的建議或決定；而不是在審視比較性期中結果後，才找出須調整的設計面向。因此，非常重要，數據監測委員會或調整委員會在設計階段，即應與廠商就假想的情境，以及調整計畫所擬定的行動，是否所有相關方都認為合理，進行廣泛的討論。

應有保護措施，以確保負責準備並向數據監測委員會或調整委員會報告期中分析結果的人員，無論這些人員是否屬於廠商組織內的人員，或是另一個組織（例如，受託研究機構《CRO》）人員；應與負責試驗管理或執行是不同的人員。這樣的實行，將有助於確保參與日常試驗管理與執行的人員，即使是無意的，也不會取得治療組別或比較性的結果。同樣，基於上述的理由，數據監測委員會或調整委員會回報給廠商的建議，通常不應包括期中分

析結果的任何細節。

雖然通常不建議廠商代表能取得比較性期中結果，但在有些情境下，特定的廠商人員取得有限的資料結果是合理的。例如，有些調整性試驗可能包含決策，如劑量的選擇，這在非調整性試驗中通常為廠商的責任，並且對藥物開發計畫具有重要的長期影響。廠商在滿足以下狀況時，取得有限的資訊是可以接受的：僅有少數廠商代表參與、被允許取得資訊的廠商人員不得參與試驗的執行或管理、且制定適當的程序確保其他關鍵方（例如，病人、試驗主持人、與試驗指導委員會《trial steering committee》）均不知道比較性期中結果。完全限制廠商取得比較性期中結果，可將試驗完整性被破壞的風險降至最低。此舉在大多數情況下，可以透過廠商和數據監測委員會在設計階段廣泛的規劃與討論而達成。

適當地限制資料的取得，須有精心計畫的程序來維持和驗證機密性，以及監測與遵守操作程序的文件紀錄。其方法通常包括對於可取得期中資料的人員採用保密協議；或使用防火牆以避免試驗人員取得可推論治療組別訊息的資料；並制定資料取得計畫，確定誰可以取得機密性資料、何時可使用，與涉及那些類型的資料和結果。重要的內容將在 8 章節中更詳細的討論。

於調整性試驗中，也有可能因為知道調整的決策，而獲得了關於期中結果的訊息。例如，知曉樣本數修改的計算方法以及調整後的樣本數，便可反向計算療效的期中估計值。因此，應盡可能採取措施，將試驗相關人員能做推論的訊息降至最低。預先設定調整規則是至關重要的，但可以於計畫書中只概述一般的方法，而特定的演算法，僅載明於給少數人的文件中，如數據監測委員會章程 (charter) 或調整設計章程。慎重考慮和規劃期中分析後可以散布訊息的範圍也很重要。一般而言，調整性變更的資訊應盡可能不讓試驗主持人和試驗參與者知道。舉例來說，假若期中分析後要增加樣本數，試驗中心可以被告知目標收納的人數尚未達到，而無須通知他們最終目標樣本數。此外，使用離散性而非連續性的調整決策閾值 (adaptation decision threshold)，是另一種可行的方法來限制可以推斷的資訊，有助於降低試驗完整性的風險。

8 法規考量

8.1 與法規單位的互動

試驗廠商與法規單位間互動的目的與性質，會因在不同的開發階段而異。某些調整性試驗增加的複雜度和對其操作特徵的不確定性，可能與法規單位的互動須要較通常更早、更廣泛。

在藥物開發的早期，法規單位對於試驗計畫書的審查通常著重於試驗受試者的安全性，而非關於藥理活性或療效推論的有效性。在藥物開發的晚期，法規單位在評估設計和分析計畫上扮演更廣泛的角色，以確保試驗提供足夠可靠的結果以作法規單位決策之用。

法規單位對於廠商所擬晚期調整性臨床試驗的審查，包括考量其設計與分析計畫是否滿足概述於本指導原則的關鍵性原則。亦即，廠商應預先設定調整設計的細節，並說明錯誤結論的機會如何能充分地控制、療效估計如何能足夠可靠、以及試驗完整性如何能適當地維持。此外，一個良好的做法是，廠商在規劃階段即已探索各種不同的調整和非調整設計選項，並與法規單位討論所擬調整設計的選取考量。

雖然試驗過程中任何擬定之計畫外變更試驗設計應該告知法規單位（通常透過計畫書變更），法規單位通常不會參與前瞻規劃的調整決策。這是廠商的責任，通常是透過指定的委員會（例如數據監測委員會）來實施調整設計。試驗進行間，法規單位可能要求監測委員會提供開放會議的會議紀錄，惟閉門會議的會議紀錄，或任何其他關於比較性期中結果的溝通紀錄或資訊，除非病人有安全性風險之特殊情況，否則應保持機密性直到試驗有了結論。

8.2 執行調整性試驗前的文件準備

為了使法規單位能透徹地評估，調整性臨床試驗在規劃文件的準備上勢必比非調整性試驗來得複雜。除了非調整性臨床試驗的計畫書和統計分析計畫書如「*臨床試驗統計指導原則*」所論述之基本項目外，調整性臨床試驗在開始執行前應制定的文件包括：

- 選擇該設計的理由。正如在其他章節所討論的，一個良好的做法是，評估擬定的設計與其他調整和非調整設計在重要的操作特徵上的比較，並且將這些資訊提交給法規單位可能是有用的。然而，最終設計的選擇仍是廠商的責任。
- 調整性計畫的詳細描述，包含期中分析預期的次數與時間點、可能修改的特定設計面向、以及使用的調整決策規則。
- 負責實施調整設計的機構，如數據監測委員會或指定的調整委員會（如果適用），其任務相關的資訊。
- 預先設定的統計方法，作為產生期中結果、引導調整決策、進行假說檢定、估計療效、及試驗結束時對療效估計值之不確定性估計等使用。應預先設定用於進行期中與最終分析的軟體。資訊可能包含所有相關電腦程式碼（當適用時）。
- 評估和討論設計的操作特徵，其中應包含型一誤差機率、檢定力、樣本數（預期、最小和最大）、療效估計的偏差、及信賴區間的覆蓋範圍。此等評估可能須透過分析計算或電腦模擬來達成。假若以分析方法評估操作特徵，應提供方法學適當的細節（例如，參考文獻或證明）。
- 誠如之前所定義的，當模擬成為主要或唯一評估試驗操作特徵的方法時，可提供詳細的模擬報告，包含：
 - 試驗設計的整體描述。
 - 範例試驗，描述少數幾個具有不同結論的假想試驗。例如，以原本樣本數得到正

面結果之試驗，第一次期中檢視後因無效停止之試驗，樣本數增加後獲得正面結果之試驗等等。

- 模擬情境所使用的參數配置集 (set of parameter configurations)，包含其選擇適當性的理由。
 - 每一模擬情境評估所使用的模擬試驗數目，和此數目適當性的理由。
 - 在不同模擬情境下估計的操作特徵之模擬結果細節。
 - 模擬程式碼。若提供程式碼，應具有可讀性並且應充分地解說。程式碼應包含生成模擬結果的隨機種子，提供的程式碼以廣泛使用的統計程式語言來撰寫也很有幫助。即使以另一種語言生成模擬結果 (通常是為了計算效率的理由)，提供以廣泛使用的統計程式語言撰寫的可執行版本之程式碼可能有助於模擬的審查。在某些情況，包含額外詳細的資訊 (例如，公式和模擬程式碼的使用說明) 也很重要。
 - 提供整體結論摘要。
- 一份全面性的資料取得計畫，是在預先規劃的調整情境下，定義如何維持試驗的完整性。此份文件應包含以下資料：(1) 執行期中分析的人員；(2) 有權取得期中結果的人員；(3) 如何控制資料的取得；(4) 如何做出調整的決策；以及 (5) 做出調整決策後，何種類型的資料將被散佈及散佈給誰。資料取得計畫應該描述在甚麼情況下那些資料可以容許傳達給廠商或試驗主持人。此外，建議廠商建立程序，以評估遵循資料取得計畫的情況，並紀錄負責做出調整決策的委員會 (即，數據監測委員會或調整委員會) 所有的期中會議。例如，期中會議所審視和討論的內容及其決定，應有書面的會議記錄來描述。

此書面文件可包含於臨床試驗計畫書，或包含於如統計分析計畫書、數據監測委員會章程、或調整委員會章程等個別的文件中。

8.3 評估與報告已完成的試驗

以調整性試驗向法規單位提出上市申請，應包含足夠的資料與文件，以使法規單位能透徹地審查結果。除了一般上市申請基本內容外，申請內容建議包含以下部分：

- 所有前瞻的計畫、任何相關委員會的章程 (如，數據監測委員會或調整委員會章程)、和上述的任何支持性的文件 (如，參考文獻、程式碼、和模擬報告)。
- 遵循規劃的調整規則和遵循概述於資料取得計畫的程序以維持試驗的完整性之資料。
- 調整過程參與的任何委員會於任何期中討論之審議和參與者記錄 (例如，數據監測委員會或調整委員會閉門和開放會議的會議記錄、指導或執行委員會的會議記錄)。
- 期間分析或用於調整決策之分析的結果。

- 於擬定仿單的臨床試驗資料段落中適當地報告調整設計和試驗結果。例如，試驗摘要應描述所用的調整設計。此外，療效估計值應充分地將調整設計納入考量，或者若使用單純的估計值 (naïve estimates) (例如，未校正之樣本平均值)，應評估偏差的程度，並且對療效估計的解讀應謹慎保守。

9 參考文獻

- 9.1 Bauer, P and K Kohne, 1994, Evaluation of Experiments with Adaptive Interim Analyses. *Biometrics*, 50(4):1029–1041.
- 9.2 Bauer, P and M Posch, 2004, Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections, by H. Schäfer and H.-H. Müller. *Stat Med*, 23(8):1333–1334.
- 9.3 Berger, JO, RL Wolpert, 1988, MJ Bayarri, MH DeGroot, BM Hill, DA Lane, and L LeCam, The likelihood principle, Institute of Mathematical Statistics, Volume 6:iii–v, vii–xii, and 1–199.
- 9.4 Berry, S, BP Carlin, JJ Lee, and P Muller, 2010, Bayesian Adaptive Methods for Clinical Trials, CRC Press.
- 9.5 Bolland, K, MR Sooriyarachchi, and J Whitehead, 1998, Sample Size Review in a Head Injury Trial with Ordered Categorical Responses, *Stat Med*, 17(24):2835–2847.
- 9.6 Chen, YJ, R Gesser, and A Luxembourg, 2015, A Seamless Phase IIB/III Adaptive Outcome Trial: Design Rationale and Implementation Challenges, *Clin Trials*, 12(1):84–90.
- 9.7 Chow, SC and M Chang, 2011, Adaptive Design Methods in Clinical Trials, CRC Press.
- 9.8 Cui, L, HM Hung, and SJ Wang, 1999, Modification of Sample Size in Group Sequential Clinical Trials, *Biometrics*, 55(3):853–857.
- 9.9 PREVAIL II Writing Group; Multi-National PREVAIL II Study Team, Davey Jr, RT, L Dodd, MA Proschan, J Neaton, JN Nordwall, JS Koopmeiners, J Beigel, J Tierney, HC Lane, AS Fauci, MB Massaquoi, F Sahr, and D Malvy, 2016, A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection, *N Engl J Med*, 375(15):1448–1456.
- 9.10 Denne, JS, 2001, Sample Size Recalculation Using Conditional Power, *Stat Med*, 20(17-18):2645–2660.
- 9.11 Dodd, LE, MA Proschan, J Neuhaus, JS Koopmeiners, J Neaton, JD Beigel, K Barrett, HC Lane, and RT Davey, 2016, Design of a Randomized Controlled Trial for Ebola Virus Disease Medical Countermeasures: PREVAIL II, the Ebola MCM Study, *J Infect Dis*, 213(12):1906–1913.
- 9.12 FDA-NIH Biomarker Working Group, 2016. BEST (Biomarkers, EndpointS, and other Tools) resource.
- 9.13 Fisher, LD, 1998, Self-designing Clinical Trials, *Stat Med*, 17(14):1551–1562.

- 9.14** Fleming, TR, K Sharples, J McCall, A Moore, A Rodgers, and R Stewart, 2008, Maintaining Confidentiality of Interim Data to Enhance Trial Integrity and Credibility, *Clin Trials*, 5(2):157–167.
- 9.15** Friede, T, and M Kieser, 2003, Blinded sample size reassessment in non-inferiority and equivalence trials, *Stat Med*, 22(6):995-1007.
- 9.16** Glimm, E, W Maurer, and F Bretz, 2010, Hierarchical Testing of Multiple Endpoints in Group- sequential Trials, *Stat Med*, 29(2):219–228.
- 9.17** Gould, AL and WJ Shih, 1992, Sample Size Re-estimation Without Unblinding for Normally Distributed Outcomes with Unknown Variance, *Communications in Statistics – Theory and Methods*, 21(10): 2833–2853.
- 9.18** Hey, SP and J Kimmelman, 2015, Are Outcome-adaptive Allocation Trials Ethical? *Clin Trials*, 12(2):102-106.
- 9.19** Heyse, JF, BJ Kuter, MJ Dallas, P Heaton, and REST Study Team, 2008, Evaluating the Safety of a Rotavirus Vaccine: The REST of the story, *Clin Trials*, 5(2):131–139.
- 9.20** Hung, HMJ, S-J Wang, and R O’Neill, 2007, Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials, *J Biopharm Stat*, 17(6):1201–1210.
- 9.21** Irle, S, and H Schäfer, 2012, Interim design modifications in time-to-event studies, *Journal of the American Statistical Association*, 107(497):341-348.
- 9.22** Jenkins, M, A Stone, and C Jennison, 2011, An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints, *Pharm stat*, 10(4):347-356.
- 9.23** Jennison, C and BW Turnbull, 1999, *Group Sequential Methods with Applications to Clinical Trials*. CRC Press.
- 9.24** Lan, KG and DL DeMets, 1983, Discrete Sequential Boundaries for Clinical Trials, *Biometrika*, 70(3):659–663.
- 9.25** Le Tourneau, C, JJ Lee, and LL Siu, 2009, Dose Escalation Methods in Phase I Cancer Clinical Trials, *Journal of the National Cancer Institute*, 101(10):708–720.
- 9.26** Magirr, D, T Jaki, F Koenig, and M Posch, 2016, Sample size reassessment and hypothesis testing in adaptive survival trials, *PloS One*, 11(2):e0146465.
- 9.27** McMurray, JJ, M Packer, AS Desai, J Gong, MP Lefkowitz, AR Rizkala, JL Rouleau, VC Shi, SD Solomon, K Swedberg, MR Zile, and PARADIGM-HF Investigators and Committees., 2014, Angiotensin–Neprilysin Inhibition Versus Enalapril in Heart Failure, 2014, *N Engl J Med*, 371(11):993–1004.
- 9.28** Müller, HH and H Schäfer, 2001, Adaptive Group Sequential Designs for Clinical Trials: Combining the Advantages of Adaptive and of Classical Group Sequential Approaches, *Biometrics*, 57(3):886–891.
- 9.29** O’Brien, PC and TR Fleming. *A Multiple Testing Procedure for Clinical Trials*, 1979, *Biometrics*, 35(3):549–556.

- 9.30** Pocock, SJ, 1977, Group Sequential Methods in the Design and Analysis of Clinical Trials, *Biometrika*, 64(2):191–199.
- 9.31** Pocock, SJ and R Simon, 1975, Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial, *Biometrics*, 31(1):103–115.
- 9.32** Proschan, MA and SA Hunsberger, 1995, Designed Extension of Studies based on Conditional Power, *Biometrics*, 51(4):1315–1324.
- 9.33** Sydes, MR, MK Parmar, MD Mason, NW Clarke, C Amos, J Anderson, J de Bono, DP Dearnaley, J Dwyer, C Green, Jovic, AW Ritchie, JM Russell, K Sanders, G Thalmann, ND James, 2012, Flexible Trial Design in Practice-stopping Arms for Lack-of-benefit and Adding Research Arms Mid-trial in STAMPEDE: A Multi-arm Multi-stage Randomized Controlled Trial, *Trials*, 13(1):168.
- 9.34** Wassmer, G and W Brannath, 2016, Group Sequential and Confirmatory Adaptive Designs in Clinical Trials, Springer series in pharmaceutical statistics, New York: Springer.
- 9.35** Woodcock, J and LM LaVange, 2017, Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both, *N Engl J Med*, Jul 6; 377(1):62–70.