



## 臨床試驗調整設計之統計審查重點

蔡貴鳳\*

### 一、前言

臨床試驗的設計通常依據前期試驗結果或已收集的資訊來設計下一個臨床試驗。然常因所收集的資料有限，以致於在試驗設計之初所做的假設存在諸多不確定性。傳統的臨床試驗當試驗進行中若發現其原先設計假設錯誤時，往往無法在試驗過程中修正其設計，致使試驗失敗的機率上升。相較於傳統臨床試驗設計，靈活的調整設計允許依據試驗進行期間所蒐集的資訊，發現並修正原始試驗設計之不正確假設，提升新藥試驗成功的機率，縮短研發時間及降低研發成本。

調整設計與群集逐次設計(group sequential design)執行期間分析是密切相關的。群集逐次設計利用試驗期間分析(interim analysis)結果，評估試驗藥物的有效性及安全性，以決定是否因療效或安全性因素提早終止試驗。當試驗採用群集逐次設計時必須在試驗計劃書中載明是為療效或安全的緣故執行期間分析，訂定適當提早停止試驗的原則(如 Pocock, 1977; O'Brien-Fleming, 1979; Lan-DeMets, 1994)，確保整體型一誤差控制在 0.05 以下。

調整設計比群集逐次設計更具靈活性與可變動性。調整設計除了利用試驗期間分析結果，決定是否因療效或安全性因素提早終止試驗外，亦允許在不破壞試驗的完整性(integrity)與有效性(validity)的前提下，可依據期間分析結果適時調整後續試驗設計，從而修正一些原始試驗設計不合理的假設。依據 US FDA 於 2010 年 2 月公佈之調整設計臨床試驗草案指引(Guideline for Industry: Adaptive Design Clinical Trials for Drugs and Biologics)，調整設計應於計畫書預先載明調整計畫，包括所有可能依據該試驗分析結果(通常為期間分析)而從事的調整項目(設計或假說)，以及相關調整方案。另外，期間分析的時間點應事先設定，可以採盲性或解盲的方式進行，亦可執行或不執行統計假設檢定。

一般而言，調整設計若是基於受試者未治療前之基礎值資料(baseline data)修正納入條件；在不解盲條件下於期間分析時重新計算樣本數以期達足夠檢定力；或基於療效無關資料分析結果(如嚴重不良事件)修正設計等，其試驗的完整性與有效性較不易破壞，可不調整型一誤差。然而，調整設計若基於解盲條件下，依據期間療效分析結果，決定是否終止療效不佳的試驗組別停止；重新估算樣本數；或修正受試者族群等，則如何避免試驗操作偏差與評估偏差，以及如何控制整體型一誤差於 0.05 或以下則是重要的課題。以下將針對調整設計中常見的調整項目之統計考量重點分別加以討論。

---

\* 財團法人醫藥品查驗中心 新藥科技組



## 二、調整設計中常見調整項目

### 1. 隨機分配調整(adaptation in randomization)

隨機分配方法可區分為非機動隨機分配(non-dynamic randomization)與機動隨機分配(dynamic randomization)。非機動隨機分配包括簡單隨機分配、區塊隨機分配與分層隨機分配等，其特徵為操作簡單、受試者分配到各組機率須事先載明且固定不變。

機動隨機分配則為受試者分配至各組的機率可以根據一定條件在試驗中進行調整。常用機動隨機分配包括基礎值-調整分配(baseline-adaptive randomization)與反應-調整隨機分配(response-adaptive randomization)。基礎值調整分配係依據已納入受試者之重要共變數(covariate)基礎值在各治療組中分佈不平均程度，機動調整後續受試者進入各治療組的機率，以期將重要共變數在各治療組分布的差異減至最低。由於基礎值-調整分配是基於受試者未治療前之基礎值資料修正隨機分配機率，不易產生選擇偏差與評估偏差。

相反地，反應-調整隨機分配(又稱 outcome dependent randomization 或 play the winner approach)則是依據前面病人的治療結果，來機動調整後續受試者進入各治療組的機率，以期使較多的受試者分派至治療結果較好試驗組的分配方法。此種隨機分配方法適用探索性臨床試驗，如劑量療效反應(dose-response)試驗，可以分派較少受試者至成功機率較低的試驗組別。然而，由於此方法是依據治療效果進行隨機分配，隨機性不夠，可能導致治療組間影響療效重要預後因子分布不均衡，易產生評估偏差。另外，這一次的治療效果可能使人知道下一對象分配的組別，亦易產生選擇偏差。因此，若使用於療效確認性試驗，應謹慎評估其對試驗執行所產生的偏差程度。

### 2. 樣本數再估計(Sample size re-estimation)

樣本數的再估計可分為盲性作業下再估計(blinded re-estimation)與解盲作業下再估計(unblinded re-estimation)。盲性作業下再估計通常是在不解盲條件下，於期間分析時重新評估 nuisance 參數(如共同變異數 variance 或不分組別的事件發生率 pooled event rate)並重新計算樣本數。盲性作業下重新估計樣本數的好處是可由試驗執行團隊執行，不須委託獨立資料監視委員會(Independent Data Monitoring Committee, IDMC)執行，且一般可不調整整體型一誤差。

解盲樣本數再估計則是依據期間分析所觀察到的療效大小重新估計樣本數，此方法的好處是可獲得更精確的樣本數。然而與盲性作業下樣本數重新估計不同的是，研究顯示根據解盲期間分析之療效大小之去盲再估計將會增加試驗的型一誤差，所以必須採用一些統計方法來控制整體型一誤差 0.05 以下。基本上，期間分析如執行解盲樣本數再估



計，不可由試驗執行團隊來分析，宜委託獨立監測委員會執行。儘管樣本數的再估計可充分利用試驗中所收集的資訊，提高試驗成功的機率，然而最小臨床有意義療效仍應於試驗執行前確定，否則任意擴大樣本數而達到統計上有顯著療效，並不表示為臨床有意義之療效。

### 3. 試驗期間停止試驗組別(interim dropping treatment arms)與無縫設計 II/III 期合併臨床試驗(Phase II/III Seamless Trial)

藥物臨床試驗的目標是在尋找風險/效益比可接受的藥物，若期間分析結果顯示某些治療組別療效太差或風險太高，經由調整設計，廠商可根據期間分析結果停止風險/效益比不可接受的治療組別，如此，不僅可降低病人不必要風險，符合倫理，亦可減少試驗費用。當最後分析所有比較檢定所根據的型一誤差分配，不因期中停止試驗組別而有所改變，則整體型一誤差仍可控制。然而，當試驗執行期間停止某試驗組別後，將原分配到該停止試驗組別之療效比較的型一誤差，重新分配至其他保留的療效比較時，需採用適當統計方法以控制整體型一誤差於 0.05 以下。

試驗期間停止試驗組別的設計常見於無縫設計 II/III 期合併臨床試驗(Phase II/III Seamless Trial)。此項設計主要是合併第 II 與 III 期臨床試驗，如此可減少臨床試驗費用與縮短試驗時間。連續性 II/III 期試驗通常分成兩個階段，第一階段執行第二期探索性劑量選擇試驗，並根據第一階段試驗結果選擇合適劑量進入第二階段。所以第一階段後將會停止某些風險/效益比不可接受的劑量。試驗第二階段為療效確認階段，通常此階段的樣本數決定須考慮檢定力。故第二階段往往會納入更多受試者。一般而言，最後統計分析時可合併第二期與第三期臨床試驗之所有受試者，然而根據期間分析解盲之療效結果選擇劑量，將導致型一誤差擴增，所以必須採用適當統計方法控制型一誤差。另外，值得注意的是，最後療效分析前應檢驗第一階段受試者與第二階段所納入的受試者是否一致。若無縫設計 II/III 期合併臨床試驗可單獨根據第二階段(即第三期部份)納入的受試者資料即顯示試驗用藥與對照藥統計上顯著療效差異，則此療效證據是最可被信任的。

### 4. 統計假說的變更(adaptation in statistical hypothesis)

統計假說的變更包括由原不劣性試驗更改為較優性試驗，或由原較優性試驗(superiority)更改為不劣性試驗(non-inferiority)。通常不劣性試驗於試驗設計之初已預先定義臨界點(margin)，並評估所定義的臨界點是否合理，且主要分析群體一般會兼採「依據計畫書」(per protocol analysis, PP)與「意圖治療」(Intention-to-treat, ITT)分析群體，當兩種分析群體的推論結果顯示一致性時，研究結果才是比較可靠的。因此，若計畫書預先載明統計假說將由原不劣性試驗更改為較優性試驗，且採用封閉檢定策略(closed test procedure)，則此項變更設計是較容易被接受的。

相反的，若由原較優性試驗更改為不劣性試驗則是較複雜的。若原較優性試驗並未



事先定義不劣性的臨界點，則法規單位通常是不接受根據去盲期間分析的結果再來定義臨界點。若統計假說更改為不劣性，則其樣本數有時需要根據去盲期間分析結果再估計，如此將擴增試驗的型一誤差，所以必須採用一些統計方法來控制整體型一誤差 0.05 以下。基本上，此項調整性設計需事先定義合理臨界點(margin)以及主要分析群體，於計畫書預先載明調整計畫，包括樣本數的再估計以及適當的統計方法控制型一誤差，則原較優性試驗變更為不劣性試驗仍是可行的。

### 5. 療效指標變更(adaptation in endpoint selection)與主要療效分析方法變更( change in the primary analysis)

療效指標指的是對臨床試驗受試者治療反應的一種評估或測量，可以是定性或定量的測量。受試者對藥物的反應，雖可有多個療效指標分析，但一般要選取最能確切反應藥物治療的臨床效益(須有實證依據)為主要療效指標，其餘的則可列為次要療效指標。通常主要評估指標與次要評估指標必須在臨床試驗執行前定義清楚，不宜隨意變更。目前各國法規單位皆對主要疾病擬出廣為接受的主要療效指標規範，以供試驗設計參考。然而，有些特定疾病中評估藥物治療效果的最佳評估指標仍未被完全確認，因此在試驗設計階段選擇適當指標有實質上困難。在這種情況下，採用調整設計，依據解盲期間分析結果，變更評估指標順序(包括主要與次要指標順序變更)，可能有其價值。

評估指標調整(endpoint adaptation)需考慮多重指標選擇所產生多重檢定(multiplicity)的問題，宜採用適當的統計方法，避免型一誤差的擴增。值得注意的是，如果期間分析樣本數過少，不足以評估指標間之評估靈敏度差異時，貿然採用這種調整方式，可能會錯選到一個靈敏度較差評估指標，加大試驗失敗的風險。

基本上主要療效指標分析方法，包括主要療效分析群體、統計分析方法及缺失資料處理方式等，應於計畫書中預先定義，且應當嚴格遵守不宜更改。然而當強而有力的證據(如期間分析結果)顯示預先定義的分析方法不適當且可能導致不正確的推論時，修正主要療效分析方法是可接受的。

### 三、結語

相較於傳統臨床試驗設計，調整設計具更高靈活性，可有效降低新藥研發費用，縮短研發時間，提升新藥試驗成功的機率。然而面對臨床試驗執行時的各種不確定性如改變病人族群，重新估計樣本數，停止無效試驗組別，修正主要假說或變更主要療效指標等，常常需要更複雜相對應的統計方法，有些情況下是比傳統臨床試驗設計更難執行的。

從法規角度來看，藥物發展的早期探索性臨床試驗若採用靈活的調整設計，可以獲



得更多藥物正確的訊息，設計更完整而適當的療效確認性試驗。然而晚期療效確認性試驗採用調整設計時，則需要更小心而完善的規劃，以避免試驗執行偏差，並確保型一誤差控制於 0.05 以下。

由於調整設計在試驗執行過程中需進行多次期間分析，廠商必須成立獨立監測委員會以執行期間分析，並建立完整、透明化的標準作業流程來維持期間試驗分析結果的機密性與保持試驗完整性，避免試驗操作偏差與評估偏差。另外，不論採取何種調整策略，試驗計劃書必須事先詳載欲調整的項目，提出對應之統計方法控制整體型一誤差，並於設計之初尋求法規單位建議，審慎評估採用調整設計的適當性。

## 參考文獻

1. FDA Draft Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (2010)
2. CHMP Reflection Paper on Methodological Issues in Confirmatory Clinical Trials planned with an Adaptive Design. EMEA: London, 2007 (CHMP/EWP/2459/02)
3. CPMP Points to Consider on Switching between Superiority and Non-inferiority. EMEA: London, 2000 (CPMP/EWP/482/99).
4. DeMets, D.L. and Lan, K.K.G. (1994), Interim analysis: The alpha spending function approach. *Statistics in Medicine*, 13, pages 1341-1352.
5. O'Brien, P. C. & Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics* 35, 549-556.
6. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika* 64, 191-199.