



臨床試驗統計指導原則

第一版

中華民國 107 年 11 月 20 日

財團法人醫藥品查驗中心 著

目錄

1. 緒論	3
1.1 背景與目的	3
1.2 範圍與指南	4
2. 臨床試驗整體考量	5
2.1 試驗內容	5
2.1.1 發展計畫	5
2.1.2 確認性試驗	5
2.1.3 探索性試驗	6
2.2 試驗範圍	6
2.2.1 受試者族群	6
2.2.2 主要指標與次要指標	6
2.2.3 複合型指標	7
2.2.4 整體評估指標	7
2.2.5 多個主要指標	8
2.2.6 替代指標	8
2.2.7 類別型指標	9
2.3 避免偏差之設計技巧	9
2.3.1 盲性	9
2.3.2 隨機分派	10
3. 試驗設計之考量	11
3.1 設計類型	12
3.1.1 平行設計	12
3.1.2 交叉設計	12
3.1.3 複因子設計 (Factorial Designs)	13
3.2 多中心試驗	13
3.3 比較的類型	15
3.3.1 較優性試驗	15
3.3.2 相(對)等性或不劣性試驗	15
3.3.3 劑量反應關係試驗	16
3.4 群集逐次設計 (Group Sequential Designs)	17
3.5 樣本數	17
3.6 數據收集與處理	18
4. 試驗執行之考量	18
4.1 試驗監測與期間分析	18
4.2 變更納入及排除條件	19
4.3 受試者收納速度	19
4.4 調整樣本數	19
4.5 期間分析與試驗提早終止	19

4.6 獨立數據監測委員會(IDMC)之角色	21
5. 數據分析之考量	21
5.1 預先設定分析方法	21
5.2 分析群體	22
5.2.1 全分析群體	22
5.2.2 依計畫書群體	23
5.2.3 不同分析群體的角色	24
5.3 缺失值及異常值	24
5.4 數據轉換	25
5.5 參數估計、信賴區間及假說檢定	25
5.6 顯著水準的調整	26
5.7 次群體、交互作用與共變數	26
5.8 數據完整性與電腦軟體效度	26
6. 安全性與耐受性之評估	27
6.1 評估範圍	27
6.2 指標選擇與數據收集	27
6.3 評估群體與數據的呈現	27
6.4 統計評估	28
6.5 綜合總結	29
7. 報告撰寫	29
7.1 評估與報告	29
7.2 臨床試驗資料摘列	30
7.2.1 療效資料	31
7.2.2 安全性資料	31
詞彙表	32

本指導原則係參考 ICH E9 Guideline：Statistical Principles for Clinical Trials 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，並非政府機關所發佈之法規。凡涉及政策方向及法規解釋與適用，仍應依衛生主管機關之指示為準。

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

1. 緒論

1.1. 背景與目的

藥品的有效性和安全性需要由臨床試驗來驗證，而臨床試驗則須遵循 2001 年 9 月所修訂之「藥品優良臨床試驗規範 (Good Clinical Practice, GCP)」來執行。該規範認定統計學在臨床試驗設計與分析扮演重要的角色。隨著統計學研究在臨床試驗領域不斷地發展，加上臨床研究在新藥核准過程與健康照護上的關鍵角色，因此一份簡明的臨床試驗統計相關文件是十分必要。本指導原則主要目的為促使於歐洲、日本和美國申請藥品上市所附臨床試驗之統計方法學原則能夠協調。

首先，本指導原則運用 CPMP (Committee for Proprietary Medicinal Products 專利醫藥品委員會) 於 1994 年 12 月所發布之「醫藥品上市之臨床試驗統計方法 (Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorizations for Medicinal Products)」，並參照日本衛生福利部 (Japanese Ministry of Health and Welfare) 於 1992 年 3 月所發布之「臨床試驗統計分析指引 (Guidelines on the Statistical Analysis of Clinical Studies)」，及美國食品藥品管理局 (Food and Drug Administration, FDA) 於 1988 年 7 月發布之「新藥查驗登記之臨床與統計文件的格式與內容基準 (Guideline for the Format and Content of the Clinical and Statistical Sections of a New Drug Application)」。涉及統計原理與方法學的相關議題也會包含於國際醫藥法規協和會 (ICH) 的其他規範中，尤其是以下的指導原則：

- E1A：參與臨床安全性評估之總體範圍
- E2A：臨床安全性資料管理：快速通報的定義與標準
- E2B：臨床安全性資料管理：個案安全性報告通報要點
- E2C：臨床安全性資料管理：上市藥品定期安全性報告 (PSUR)
- E3：藥品臨床試驗報告之格式及內容基準
- E4：支持藥品查驗登記之藥物劑量反應資訊
- E5：接受國外臨床資料之族群因素考量
- E6：藥品優良臨床試驗規範
- E7：特殊族群研究：老年醫學
- E8：臨床試驗之一般考量
- E10：臨床試驗對照組之選擇
- M1：制訂規章用的醫學術語標準化
- M3：藥物臨床試驗前所需之非臨床安全性研究

本指導原則旨在提供廠商整體藥物研發階段之臨床試驗設計、執行、分析和評估的方向，亦可協助負責準備申請文件或是在臨床試驗晚期評估有效性和安全性證據之科學專家。

1.2. 範圍與指南

本指導原則對於廣泛各學科的專業人員應該都有用處。然而，正如「藥品優良臨床試驗規範」所述，所有臨床試驗相關的統計工作，皆應由具有適當資格及足夠經驗的統計人員負責。試驗統計人員(詳見詞彙表)的角色與責任是與其他臨床試驗專業人員合作，以確保統計學原則能被適當地應用於支持藥物研發的臨床試驗。因此，試驗統計人員須有足夠的專業訓練與教育，並有足夠經驗貫徹本指導原則。

所有為上市申請而執行的臨床試驗，其設計、執行與採用的統計分析方法等重要細節，應於試驗開始前於計畫書中清楚說明。試驗結果可信度的高低將取決於對計畫書執行程序的遵循程度及事先規畫的主要分析計畫。試驗計畫書及後續變更，必須經過負責的人員，包括試驗統計人員的同意。試驗統計人員必須確保計畫書及後續任何的變更中，統計相關問題以適當術語正確並清楚地描述。

本指導原則所闡述的原則主要是針對晚期臨床試驗，大多數是療效確認性試驗。除了療效之外，確認性試驗也可用安全性指標(如：不良事件、臨床檢驗數據、心電圖資料等)、藥效學、藥動學指標(例如：生物相等性的確認性試驗)作為主要評估指標。最後，儘管藥物早期研發階段包括許多的臨床試驗，在本質上是以探索性為目的，本規範統計學原則仍可應用到這類試驗。因此，本指導原則的內容應盡可能應用到各個階段的臨床試驗。

本指導原則中所闡述的原則多是以降低偏差(bias)、提升精準度(precision)至最高為目標。偏差一詞在本指導原則指的是臨床試驗中與試驗設計、執行、分析與結果解讀相關的任何具系統性傾向的因素，使得治療效果(詳見詞彙表)估計值偏離其真值。因偏差的存在將嚴重危及臨床試驗得到正確結論的能力，應盡可能地找出導致偏差所有的潛在來源，才能減少這些偏差。

有些偏差是源自試驗的設計，例如將病情較輕的受試者系統性地分配到某一組。其他偏差如：根據試驗結果認定違反計畫書標準，並將其排除於統計分析之外，即為可能的偏差來源，並會影響到治療效果正確地估計。偏差常在不知不覺中發生且難以直接測量，因此，試驗結果和主要結論的穩健性就顯得格外重要。穩健性的概念是整體結論對於不同資料的局限性、或假設與分析方法的敏感度。穩健性意味著不同的假設或不同統計方法對於試驗療效和主要結論皆沒有實質的影響。至於治療效果與治療間比較的不確定性統計測量的解釋，應包含偏差對 p 值、信賴區間或統計推論的影響

由於頻率學派(詳見詞彙表)為試驗設計和分析的主要方法，本指導原則在討論假說檢定和

信賴區間時，大部分引用頻率法。這並不意味著其他方法不適當，若使用的理由明確且所得結論足夠地穩健，則貝氏方法(詳見詞彙表)及其他方法的使用亦可納入考量。

2. 臨床試驗整體考量

2.1. 試驗內容

2.1.1. 發展計畫

新藥臨床試驗的主要目標是尋找藥物是否存在安全及有效的用法用量。在此範圍內，風險效益比是可接受的，同時還要找出可能由該藥得到好處的特定族群及適應症。

為滿足此一目標需要一系列的臨床試驗，每一個試驗均有其特定的目標(見 ICH E8)，這都需要在臨床研究計畫中或一系列計畫中闡明，並且需具有適當的決策點，能夠有隨知識的累積而進行變更的彈性。一個新藥上市申請需要清楚地描述各試驗的主要內容，以及每個試驗的貢獻。解讀和評估整個藥物發展計畫的證據需整合各個試驗所提供的證據(見 7.2 節)。在每個試驗中採用約定的標準，如醫學術語的標準化、主要測量的定義與時程、處理違反試驗計畫書的方法等，將有助於各個試驗的整合。當多個試驗都論及同樣的醫學問題時，則統計學上的總結、概述或統合分析(見詞彙表)將提供更豐富的資訊。可能的話，計畫中應預先定義出相關的試驗與試驗間共同的特點。若有其他主要統計議題會影響發展計畫中的試驗，亦需在計畫中陳述。

2.1.2. 確認性試驗

確認性試驗(confirmatory trial)是一種事先提出假說並進行檢驗且具有適當對照組的試驗。通常，確認性試驗是提供療效或安全性的有力證據，在這些試驗中，依據試驗的主要目標事先提出主要假說，並在試驗完成後進行檢定。在確認性試驗中，準確估計療效大小和轉化至臨床意義是同樣重要的。

確認性試驗可以針對所欲提出的宣稱提供有力證據，因此，確實依照試驗計畫書及標準操作步驟進行是格外重要，一些不可避免的改變需提出解釋及記錄，並檢驗其所產生的影響。這類試驗設計的合理性及其他重要統計部分，如預計分析的主要特點等，需在試驗計畫書中陳述，每個試驗應僅探討數量有限的問題。

要支持任何宣稱所需要的有力證據，皆需由確認性試驗來驗證所開發的藥物具有臨床效益。因此，確認性試驗必須針對所提出的療效或安全性關鍵臨床問題給予充分的回答。此外，很重要的是，要能把試驗結果推論到目標病人群體。因此，這可能也會影響到所需試驗中心的數目與類型(例如專科或一般醫師)及所需的試驗數。確認性試驗的結果必須穩健。在某些情況下，單一的確認性試驗所提供的證據可能就足夠。

2.1.3. 探索性試驗

確認性試驗的緣由及設計是基於一系列早期臨床探索性試驗。跟所有臨床試驗一樣，這些探索性研究也應有清楚及明確的目標；但與確認性研究相比，探索性試驗的目標並不限於單純地對預設的假說進行檢定。探索性試驗有時需要採用更為彈性的方法進行設計，以因應逐漸積累的結果對試驗進行適當的修改，對資料進行探索性分析，假說檢定可能也會執行，但此假說檢定是根據資料的特點而定的。儘管這類試驗對整體療效驗證有貢獻，但不能作為證明有效性的正式依據。

任何單一臨床試驗可能同時具有確認和探索兩個面向，例如大多數確認性試驗常對資料進行探索性分析，用以對試驗結果提出解釋或支持，並為後續研究提出進一步的假設。試驗計畫書須明確地區分確認和探索這兩方面的內容。

2.2. 試驗範圍

2.2.1. 受試者族群

在藥物開發的早期階段，臨床試驗受試者的選擇在某種程度上，是受到最可能觀察到預期的臨床療效之影響，因此，受試者往往是整體病人族群中很局限、最容易顯現療效的次族群。但在確認性試驗階段，試驗受試者須更能反映目標族群，因此在這些試驗中應盡可能地放寬納入和排除標準，使得受試者與目標族群有足夠的同質性以便精確地估計治療效果。由於不同地理位置、不同研究時間、不同的研究者和醫療院所的醫療處置等因素影響，沒有任何一個臨床試驗可以完全代表未來的使用狀況。儘管如此，我們仍應盡可能減少上述因素的影響，並在闡述試驗結果時，對上述影響加以討論。

2.2.2. 主要指標與次要指標

主要指標(又稱主要變數)是針對試驗主要目標,提供與臨床最有關且具有可信證據的指標。通常主要指標只有一個，這個指標常是療效指標。因為大多數確認性試驗的主要目標是提供與療效相關的強有力的科學證據。安全性與耐受性也是重要的考量，故有時也會是主要指標。有關生活品質和衛生經濟的測量也是進一步可能考慮的主要指標。主要指標的選擇應反映相關研究領域中已被接受的準則和標準。建議採用先前試驗中或發表文獻中已有使用經驗的可信且有效的指標。所選的指標需有充分的證據說明它對符合納入排除條件的受試者，能有效且可信地反映在臨床上的關聯性及重要性的治療效果。主要指標通常是用來估計樣本數的指標(見 3.5 節)。

在很多情況下，評估受試者結果的方法不是很直觀，故需小心的定義。例如將死亡率選為主要指標而無進一步說明或嚴格定義就是不適當的；死亡率的評估可以是比較某一時間點

尚存人數的比例，也可以是比較某一時段內的整體存活曲線分布。另一個常見的例子則是復發，療效指標可以是簡單的二分法指標(在指定時段內的任一復發)、第一次復發時間、或復發率(單位時間內的復發次數)等。針對慢性疾病的療效研究，隨時間而改變的功能性狀態，又給主要指標的選擇增添其他的挑戰。這有很多可能的評估方式，例如比較開始和結束時的觀察值，比較試驗期內所有觀察值所得的斜率，比較超過或低於特定區間的病人比例，或使用重複量測方法來分析。為避免事後定義而造成的多重檢定問題，在計畫書中精確的定義將在統計分析中使用的主要指標是很重要的。此外，所選定的主要指標其臨床意義和相關量測過程的有效性，都應在計畫書中加以闡述。

主要指標的定義及其選擇的理由應在計畫書中加以說明。解盲後重新定義主要指標是無法接受的，因為由此所產生的偏差難以評估。當主要目標所定義的臨床療效有多種測量方法時，只要實際情況可行，在計畫書中應根據測量方法的臨床相關性、重要性、客觀性或其他相關特性選定一種測量作為主要指標。

次要指標是與主要目標相關的支持性指標，或與次要目標相關的療效指標。在計畫書中預先定義這些次要指標以及在解讀試驗結果時，解釋其相對重要性和所扮演的角色也是很重要的。次要指標的數目應有所限制，且對應於試驗中所要回答的限定問題。

2.2.3. 複合型指標

如果與主要目標相關的多種測量中，不能選擇單一個作為主要指標，另一種方法是應用預先定義的方法來整合或組合多個測量值，使其成為一個單一或複合的指標。事實上，主要指標有時是整合多種臨床指標而成(如在關節炎、精神疾病或其他疾病中的評分尺度)，這種方法涉及到多重檢定的部分，是不需對型一誤差進行調整的。整合多個測量值的方法應在計畫書中加以說明，且應就複合指標尺度所對應之臨床療效大小進行說明。當複合指標被選為主要指標時，這個複合指標的單一組成指標若有臨床意義並有效量測時，有時也會進行單獨分析。另，當評分量表被選為主要指標時，其內容效度(見詞彙表)、評估者間和評估者內的信度(見詞彙表)及疾病嚴重程度變化的鑑別度是格外重要的。

2.2.4. 整體評估指標

在某些情況下，整體評估指標(見詞彙表)已發展用來評估藥物的整體安全性、療效和有用性。這種指標整合了客觀指標與研究者對受試者的現況或狀況改變的總印象，它常常是一個序位性類別型指標。在某些治療領域如神經科和精神科，已經建立整體評估指標來評估療效。

整體評估指標通常都有一個主觀成分。使用整體評估指標作為主要或次要指標時，需要在計畫書中對以下方面進行詳細說明：

- (1) 整體評估指標與試驗主要目標的相關性；

- (2) 整體評估指標之效度和信度的基礎；
- (3) 如何將單一受試者搜集到的資料歸類至整體評估指標的一類；
- (4) 如何將有缺失資料的受試者歸類至整體評估指標的一類，或有其他處理方法。

當使用整體評估指標時，若其中組成的客觀指標由研究者評估，則這些客觀指標應作為額外的主要指標或至少是重要的次要指標。

整體評估的有用性在於它整合了效益與風險兩個因素，也反映了醫師的決策過程，醫師在決定用藥時須權衡使用這些藥物的利弊。使用整體評估指標的一個問題是，在某些情況下可能將兩種不同效益風險的藥物判為相同，例如在評估一種藥物的整體有用性等同或優於另一種藥物時，可能掩蓋了其療效較差或無效而同時不良反應較少的事實。因此，不建議採用整體評估指標作為主要指標。如果整體評估指標被選作主要指標時，分別考慮對特定的療效和安全性指標作為附加的主要指標是很重要的。

2.2.5. 多個主要指標

有時，我們會希望使用多個主要指標，每一個(或其中一些)可以支持藥物某部分的療效。使用這類指標的證據須清楚載明，例如應當清楚地定義任何一個指標、以及最少有幾個指標或全部指標被認為是達到試驗目標所必須的。針對已定義的主要指標及統計推論方法，其主要假說與參數(如平均數、百分比、分布)均應清楚地說明。而多重檢定問題可能造成的型一誤差影響也應解釋(見 5.6 節)，亦應在計畫書中訂定控制型一誤差的方法。評估型一誤差影響時，主要指標間的相關程度可能也須考慮。若試驗的目標是顯示所有主要指標的效果，那就不需要調整型一誤差，但是必須考慮型二誤差和樣本數的影響。

2.2.6. 替代指標

如果不能通過觀測實際臨床效果來直接評估受試者的臨床效果，可以考慮採用間接指標(替代指標，見詞彙表)。有些被普遍接受的替代指標已用於一些適應症，這些替代指標被認為是臨床療效上可靠的預測指標。在選用替代指標時，主要考量有兩點，第一，它可能不是臨床感興趣的真正預測值，例如，它可能測定了與一特定藥理學機轉相關的藥物活性，無論是有反應還是沒反應，它可能無法提供藥物作用的範圍與最終療效的完整資訊。已經有很多例子指出，替代指標顯示有高度正效應的藥物，最終卻對受試者的臨床結果有害；相反地，也有一些例子指出具有臨床效應的藥物，替代指標卻沒有可測量到的影響。第二，選用的替代指標並不能產生可直接與藥物不良反應相權衡的定量臨床效果測量。儘管如何驗證替代指標有效的統計標準已經提出，但如何運用這些標準之經驗還相當有限。事實上，替代指標所提供的證據強度取決於：(i)生物學上的合理性；(ii)替代指標對臨床結果預後判斷之流行病學研究證據；(iii)從臨床試驗中獲得的替代指標治療效果與臨床結果相關的證據。一個產品的臨床指標和替代指標之間的關係，並不一定能適用於治療同一種疾病，卻具有不同作用方式的另一個產品。

2.2.7. 類別型指標

二分法或其他分類的連續型或順序型指標有時也是需要的。”成功”和”有反應”的準則就是這種二分法指標的常見例子，但須精準地定義，例如，對一個連續型指標最少改善(相對於基值)百分之多少或對等級型指標等於或超過某一尺度閾值(如”好”)。如舒張壓下降到低於90 mmHg 就是一個常見的二分法案例。對有明顯臨床相關性的變量進行分類最有用，惟由已知的試驗結果很容易對分類標準的選擇產生偏差，計畫書應事先定義分類標準並明確寫明。採用分類指標通常意味著要喪失部分資訊，由此導致分析的檢定力降低，應在樣本數估算時加以考量。

2.3. 避免偏差之設計技巧

在臨床試驗中，避免偏差產生最重要的兩個設計技巧是盲性和隨機分派，這應該是大部分申請上市有對照組之臨床試驗的一般特點。大多數這樣的試驗採用雙盲設計，依據適宜的隨機分派計畫，事先將藥物進行包裝，提供給試驗中心的藥物標明的僅有受試者編號和服用時間，因此參與試驗的每一個人都不知道哪位受試者服用哪一種藥物，甚至不知道編碼字母。這種方法將在 2.3.1 節和 2.3.2 節進行討論，例外狀況將在最後討論。

透過計畫書特定的程序可以在設計階段減少偏差，這個目標是盡可能減少試驗過程中任何可預期的違規事件發生而影響分析，包括各種情況的違反試驗計畫書、提早退出試驗和缺失值。計畫書應考量如何減少這些狀況發生的頻率及發生在資料分析時該如何處理。

2.3.1. 盲性

盲性是為了減少在試驗的過程中和在結果解讀時產生有意或無意的偏差(bias)。這些偏差來自於對治療的瞭解而對受試者的納入、隨機分派、後續照顧、受試者對治療的態度、評估指標的量測、提早退出試驗者的處理、在分析中剔除數據等等的影響。其根本目的是防止辨認出受試者所接受的治療直到這些可能產生偏差的機會都結束為止。

雙盲試驗是所有受試者及所有參與治療或臨床評估的研究人員及廠商人員都不知道受試者接受的是何種治療，包括決定受試者是否符合資格、量測評估指標或評估計畫書依從性的人均須保持盲性。這樣程度的盲性要持續到整個試驗完成，且只有在資料庫清理到可接受的品質時，才可對適當的人員解盲。若廠商有不參與治療或評估受試者的人員需要知道治療組別編碼(如生物分析學家、試驗監測人員、參與嚴重不良事件通報者等)，廠商必須有適當的標準操作程序以防止治療組別編碼不適當地傳出。在單盲試驗中，研究者或其成員知道受試者接受何種治療但受試者本身不知道，亦或相反。在開放性試驗中，所有人都知道受試者接受何種治療。雙盲設計是最佳的方法，試驗用藥在用藥前或用藥時必須無法從外觀、味道等識別出來，且在整個試驗均保持盲性。

要做到雙盲，會遇到很多困難，比如：兩種治療型態可能完全不同，如手術治療和藥物治療；或兩種藥物可能是兩種不同的劑型，雖然用膠囊技術可使兩者無法分辨，但改變劑型可能會改變藥物動力學或藥效學的特性，因此需要建立不同劑型的生物相等性。又或兩種藥物每天的用法不同，在這種情況下，要達到雙盲就要採用「雙虛擬」(見詞彙表)技術。這種技術有時會使用藥時程十分不尋常，以至對病人的參與動機和順從性產生負面影響。倫理方面的困難也會干擾其應用，例如，必須進行無用的手術操作。無論如何，應該要努力克服這些困難。

若採用的治療出現了效果，可能會使雙盲遭到部分破壞。在這種情況下，不讓研究者和相關廠商人員知道某些檢驗結果(如某些臨床實驗室檢驗結果)，有助於盲性的維持。在開放性試驗中，專一的或特別的治療效果可能無法不讓病人知道，則可考慮採用類似的方法使偏差達到最小(見下段)。

如果雙盲設計不可行，則應考慮採用單盲設計。而在有些情況下，只有開放性試驗才可行或符合倫理。單盲和開放性試驗更具彈性，但特別重要的是，研究者知道下一個受試者將接受哪種治療，不應影響對下一位進入試驗的病人的納入；納入病人最好早於隨機分派之前。針對這樣的試驗，應考慮採用中央化的隨機方法，如用電話通知指定之隨機方法。除此之外，臨床評估應該由不參與治療的醫療人員執行且其不知道受試者接受何種治療(evaluator blind)。在單盲或開放性試驗中，應盡最大努力使偏差來源達到最小，且主要評估指標應盡可能地客觀。採用何種程度盲性的理由及透過什麼方法使偏差達到最小的步驟，均應在計畫書中說明。例如，廠商應當有適當的標準作業流程以確保在分析前的資料庫清理過程中，治療組別編碼的存取被適當地限制。

單一受試者的解盲只有在醫生必須瞭解受試者所接受的治療時，才能對此病人解盲。任何有意或無意地破壞盲性，不管是什麼理由，都必須記錄下來並在試驗結束時予以解釋。解盲的過程及時間點亦須記錄。

在本指導原則中，數據的盲性評估(見詞彙表)是指在試驗完成(最後一位病人的最後一次觀察)與解盲之間的資料查核。

2.3.2. 隨機分派

臨床試驗中，隨機分派導入了一個特定的機率因素到受試者的治療分派，在後續的資料分析中，對量化評估療效提供了確實的統計基礎，也使得各治療組間已知和未知預後因子分佈趨於相似。若能與盲性設計合用，隨機分派有助於避免在受試者的選擇和分組時因治療組別可預期而導致的偏差。

臨床試驗的隨機分派表記錄了隨機安排給受試者的治療，最簡單的情況是記錄治療組別序

列(在交叉試驗中則是治療順序)，或者按受試者編號的相應編碼。有些試驗的隨機分派工作較複雜，如有篩選期的試驗，其事先計畫好的隨機分派治療組別或治療序列仍然必須清楚。不同的試驗設計需有不同的方法產生隨機碼，而隨機分派表必須具有重現性。

儘管無限制條件的隨機分派是可接受的方法，但區塊(block)隨機分派法更具優越性。這將有助於增加治療組間的可比性，特別當受試者的某些特徵隨時間而變化，例如，收納受試者的策略做了改變。它還更能保證各組的樣本數幾乎相等。在交叉試驗中，它提供了一個較高效率且更容易解釋的平衡設計方法。值得注意的是區塊大小的選擇，每個區塊要盡可能的小，以防止不平均，又要足夠大，以防止對區塊內末端治療組別的可預測性。研究者及其他相關人員應對區塊大小保持盲性，可選用兩種以上的區塊大小，或每個區塊的大小隨機決定也可達到相同目的。(理論上，在雙盲試驗中，可預測性是不用擔心的，但藥物的藥理學反應常常提供了猜測的機會。)

多中心試驗(見詞彙表)應採中央化的隨機分派過程。建議每一個中心都有分開的隨機分派表，也就是依中心分層隨機分派，或每個中心分到數個完整的區塊。更常見的是，依基礎值資料中的重要預後因子(如疾病的嚴重程度、年齡、性別等)進行分層，有時對各層內各組人數達到平衡很有幫忙，這在小型的臨床試驗很有助益。一般很少需要多於兩個或三個的分層因子，因為很難達到平衡且很麻煩。倘若後續的步驟可以調整並適應的話，使用機動隨機分派(dynamic allocation procedure，見後述)是可同時使數個分層因子中達到平衡。隨機分派的分層因子應於後續分析中加以考量。

下一個進入試驗的受試者接受的是隨機分派表上所對應的治療組別(如果是分層隨機分派，則對應到分層內的隨機分派表)，只有在受試者確定進入試驗的隨機分派後，才提供該受試者適當的編碼及相對應的治療。會使人容易預測的隨機分派細節(如區塊的長度等)不應揭露於試驗計畫書中。隨機分派表須由廠商或獨立的單位安全存放，以確保整個試驗過程維持盲性。隨機分派表的查閱，須考慮到試驗中受試者在緊急情況下需要解盲的可能性。而解盲的步驟、解盲的必要性、後續治療及受試者的評估，均需在計畫書中載明。

機動隨機分派是另一種隨機分派方法，其受試者分派至何種治療取決於當前各治療組的平衡情況，在分層試驗中則是取決於受試者所屬層內的平衡情況。確定性動態隨機分派(deterministic dynamic allocation)應當避免，應對各治療組的分派導入適當的隨機化成分。必須盡一切努力維持試驗的盲性。例如，治療組別編碼資訊僅限於執行機動隨機分派的中心試驗辦公室相關人員才知道，一般是透過電話告知。這也順帶對納入條件再次確認和確定納入試驗，這特點對有些多中心試驗是非常有價值的。一般雙盲試驗事先將藥物包裝並編號的系統仍可採用，但藥物分派並非依序的，最好選用一個合適的電腦演算法，防止試驗中心辦公室人員知道治療組別編碼。當考慮採用機動隨機分派時，操作上的複雜性及對分析潛在的影響必須仔細評估。

3. 試驗設計之考量

3.1. 試驗類型

3.1.1. 平行設計

確認性試驗最常採用平行設計，即將受試者隨機分派到兩組或多組中的其中一組，每組分別施以不同的治療。這些治療組包括藥品的單一或多個劑量、一個或多個對照組，如安慰劑和活性對照組。該設計的假設(assumptions)相較於其他種設計是較不複雜的。然而，與其他設計一樣，仍然可能有其他因素使得分析及解釋變得困難(例如共變數、重複量測、交互作用、違反試驗計畫書、退出試驗及失訪等)。

3.1.2. 交叉設計

交叉設計中的每位受試者被隨機分派至接受兩種或多種治療的不同順序安排，是一種以自我為對照組的設計方法。這種設計受歡迎主要是因為可以減少受試者人數及達到特定檢定力所需要的評估數量，有時減少的幅度會很大。在最簡單的 2×2 交叉設計中，每位受試者在相繼的兩個治療期分別接受兩種治療。治療的順序是隨機指派，而兩個治療期之間常有一個廓清期(washout period)。最常見設計可擴展到在 $n(>2)$ 個治療期比較 n 種不同治療，每個受試者都接受 $n(>2)$ 種治療。這類設計有多種不同類型，如每位受試者接受 $n(>2)$ 種治療的一部分(子集)或受試者接受重複同種治療的設計。

交叉設計有很多問題足以使結論無效，最主要的原因是殘餘效應(carryover)，即治療的殘餘效果存在於後續的治療期。在加成模型中，不等的殘餘效應將使治療組的比較產生偏差。在 2×2 設計中，統計學上無法區分是殘餘效應還是治療與期間的交互作用(interaction between treatment and period)，且其相應的比較是個體間的，故這兩種比較都缺乏檢定力。這一問題在高階的設計(higher order designs)較不嚴重，但也不能完全排除。

因此，在進行交叉設計時，最重要是要避免殘餘效應。最好是在充分瞭解疾病與新藥之下有選擇地精心設計。所研究的疾病應當是慢性病且在穩定期。藥物的療效需在治療期間內完全發揮出來，廓清期必須足夠長，以使藥物的作用完全消退。這些條件是否能滿足，必須利用先前的資訊及數據，並在試驗前確定。

在採用交叉試驗時，還有一些問題須密切注意，最主要的是，若有受試者失訪時，分析和解釋都變得很複雜。另外，可能的殘餘效應使得在後續治療期出現的不良事件，難以判斷是何種治療所導致的。這些問題和其他的議題在 ICH E4 已有詳細論述。交叉設計一般應限於預期僅有少數失訪的情形方可使用。

2×2 交叉設計一個最常且令人滿意的應用是，驗證同一種藥物不同劑型的生物相(對)等性。應用於健康受試者時，如果廓清期足夠長，則藥物動力學指標的殘餘效應不大可能會出現。然而，由所得到的資料來驗證此一假設也很重要，例如，確認在每一個治療期開始時並無

藥物可檢出。

3.1.3. 複因子設計 (Factorial Designs)

複因子設計是透過的不同治療組合，對兩個或多個治療同時進行評估。最簡單的例子是 2 x 2 複因子設計，將受試者隨機分配到兩種治療(如 A 和 B)的組合之一，即只有 A，只有 B，A 及 B，既無 A 又無 B。在很多情況下，該設計主要用於檢定 A 和 B 的交互作用。但如果樣本數是基於檢定主效用(main effect)來計算的，則用於檢定交互作用的檢定力將會不足。若要檢驗 A 和 B 的聯合效應時，這點考量是很重要的，特別是兩種治療要同時使用時。

複因子設計的另一個重要應用是，當聯合使用治療 C 和 D 時，建立其劑量反應關係，特別是各自單獨使用時的固定劑量療效已於先前的試驗確立。先選定治療 C 的 m 個劑量(通常包括 0 劑量，即安慰劑)、治療 D 的相似數目 n 個劑量。完整的設計包括 m x n 個治療組，每個治療組接受治療 C 和 D 不同劑量的組合。最後估計出來的反應曲面將有助於在臨床上決定治療 C 和 D 最適合的劑量組合(見 ICH E4)。

某些情況下，2 x 2 設計比起相同樣本數評估單一種治療設計可以更有效地評估兩種治療的效果，這種策略被證實用於高死亡率的試驗非常有價值。這方法的效率及有用性是基於假設治療 A 和 B 沒有交互作用。因此，治療 A 和 B 在主要療效指標是遵循加成模型，無論有無加上治療 B，治療 A 的療效實際上都是一樣的。如同交叉設計，這特性的證據需根據先前的資訊和數據在試驗前確立。

3.2. 多中心試驗

採用多中心試驗有兩個主要的理由。首先，多中心試驗是被大家所接受，用以評估新藥更有效的方法，在某些情況下，這是在有限的時間內納入足夠受試者的實際方法。原則上，這樣特性的多中心試驗可用在臨床開發的各個階段，它可能有數個試驗中心，但每個中心都有大量的受試者。對罕病而言，它可能有大量的試驗中心，每個中心只有收納到少數幾位受試者。

其次，將一個試驗設計成多中心(及多個研究者)試驗，可為研究結果後續的推廣提供良好的依據。因為納入受試者的範圍愈廣，用藥的臨床條件設定愈寬，則對未來的使用將更具代表性。同時，因參與的研究人員較多，也為新藥療效廣泛的臨床評估提供了可能，這樣的試驗在藥物研發的晚期將成為確認性試驗，且常包含很多研究者和試驗中心。有時試驗在不同國家執行，更能促進新藥未來使用的廣義性(見詞彙表)。

如果想要多中心試驗被有意義地闡釋和外推，試驗計畫書的執行在各中心必須是清楚且相似的。此外，樣本數與檢定力的計算均假設各中心的治療效果是相同的且是無偏估計的。制訂一個共同的試驗計畫書並以此執行試驗是很重要的，執执行程序盡可能完全標準化。透

過召開研究者會議、試驗前對人員進行統一培訓、試驗過程中加強監視，可以減少評估標準和評估方式的差異。完善的設計應使得各中心、各治療組別的受試者分布相同，完善的管理是維持並實現這一設計的目標。倘若之後發現需要考量治療效果在各中心是否一致，則避免各中心樣本數相差懸殊以及有些中心樣本數太少是有好處的。因為，這樣減少了在估計治療效果時，各種不同權重估計值間的差別（這一點對所有中心的樣本數都很少時並不適用，此時中心效應在分析中不考慮）。如果事先沒注意這一點，且對同質性結果有懷疑，嚴重時會降低多中心試驗的價值到不能認同廠商提供了令人信服證據的結論。

在最簡單的多中心試驗中，每一個研究者將負責一個醫院收集的病例，所以中心指的是研究者或醫院。然而，很多試驗的情況就複雜一點，一個研究者可能要負責幾個醫院的病例收集，一個研究者可能代表幾個醫院的一組臨床醫生(協同研究者)，而每個醫生負責從各自的或幾個相關的醫院收集病例。一旦對統計模型中關於中心的定義有疑問，則計畫書中的統計章節(見 5.1 節)需明確定義(例如，依研究者、場所或地區)。在大多數情況下，還是可以用研究者來定義中心，ICH E6 提供了這方面的規範。為避免疑義，應將影響主要指標測量值及治療效果的重要因子達到同質性為目標來定義中心，將各中心合併分析的任何規則均須闡述並事先訂於試驗計畫書中，無論何時均須對治療組別保持盲性，例如在對資料作盲性評估時。

用於估計和檢定治療效果的統計模型需在計畫書中闡述。首先，應選用一個允許有中心差異的模型來研究主要治療效果，但模型不包含中心與治療的交互作用項。若跨中心的治療效果是同質性的，則在模型中常規地納入交互作用項將降低主效果檢定的效率。若跨中心的治療效果是異質性的，則治療效果的闡述是會產生爭議的。

在某些研究中，如需要大量死亡率而每個中心只有少數幾個受試者的試驗，也許不需要預期中心會對主要指標及次要指標有什麼影響，因為它們不太會表現出臨床重要性的影響。在其他研究中，可能一開始就意識到每個醫院只有少數的受試者，在統計模型中包含中心效應是不實際的。這些情況下，模型包含中心項是不合適的，而且也沒有必要依中心分層隨機分派。

若治療結果顯示藥物是有效的，且每個中心均有足夠的樣本數，一般須探討跨中心治療效果的異質性，因為這將影響結論的廣義性。對每個中心的結果可用圖示的方法，或用分析方法，如交互作用項的統計檢定，可檢測異質性，當運用統計檢定時，必須注意到在為了檢定主效果而設計的試驗中，檢定交互作用項的檢定力是很低的。

如果發現療效有異質性，解釋時須非常謹慎，應力圖從試驗的管理或受試者的特徵等其他方面找到原因，這樣的原因通常會建議進一步適當的分析和闡述。如果無法作出解釋，例如，治療效果的異質性由顯著的定量交互作用(見詞彙表)所證實，則意味著需要用不同中心予以不同權重之其他治療效果的估計方法，使治療效果的估計具有穩健性。尤為重要的是，須瞭解任何用定性的交互作用(見詞彙表)所顯示的任何異質性，若找不到合理的解釋，

則可能需要執行更多的臨床試驗，直到治療效果被可信地估計出來。

到目前所討論的多中心試驗都是基於固定效應模型(fixed effect model)。而混合模型(mixed model)也可用來探索治療效果的異質性，這類模型是把中心(center)和治療-中心交互作用(treatment-by-center)看作是隨機的，這在中心數目大的時候是適當的。

3.3. 比較的類型

3.3.1. 較優性試驗

從科學上來講，安慰劑對照試驗(顯示優於安慰劑)、活性對照試驗(顯示優於活性對照組)，或在劑量反應關係所證實的療效證據是最可信的。這類試驗都稱為較優性試驗(見詞彙表)，除非有特殊說明，一般情況下，本指導原則所述的均指較優性試驗。

對於嚴重疾病來說，如果已經存在由較優性試驗證實有效的治療，採用安慰劑對照試驗就有違倫理，此時考慮採用活性對照試驗是合乎科學性的。採用安慰劑對照還是活性對照應根據各個試驗而不同。

3.3.2. 相(對)等性或不劣性試驗

在有些情況下，研究用藥與參考治療相比不是為了證實較優性的目標，這類試驗依據研究目標可分為兩大類：一是相等性試驗(見詞彙表)，二是不劣性試驗(見詞彙表)。

生物相等性試驗屬於前一類。有時，進行臨床相等性試驗是為了其他法規上的理由，例如，學名藥產品與市售產品相比，當成分不被吸收，血液中無法檢出時，則需要執行臨床相等性試驗來驗證其相等性。

很多活性對照試驗是用來說明一個研究用藥不比活性對照差，這屬於第二類。另一種可能是試驗藥品的幾個不同劑量與標準用藥的單一或數個建議劑量相比，這種試驗目的是同時顯示試驗用藥的劑量反應關係，及試驗用藥與活性對照的比較。

活性對照組的相等性(或不劣性)試驗，可能也會包含安慰劑組，因而滿足一個試驗的多重目標。例如，優於安慰劑可以驗證試驗的有效性，並同時與活性對照藥比較有效性與安全性的相似程度。在活性對照的相等性(或不劣性)試驗中，有很多眾所皆知的困難導致其無法納入安慰劑組或新藥的多個劑量組，導致缺乏內在效度的測量(與較優性試驗相比)，因此外在效度的評估是必要的。相等性(或不劣性)試驗本質上是不保守的，因此，試驗設計或執行中的許多缺失，可能使之偏向得出相等性的結論。基於這些理由，設計這種試驗需要特別注意，執行的過程也需特別小心。例如，要儘量減少違反納入條件、順從性不良、退出、失訪、資料缺失及其他違反計畫書事件，也儘量減少這些事件對後續分析中的影響。

活性對照藥物要謹慎選擇，一個適當的活性對照藥，應是已廣泛使用的治療藥物，且其相關適應症的療效已由設計良好、紀錄完備的較優性試驗所確認並量化，因此，相似的療效可期待在目前執行的活性對照試驗中顯現。為此，新的試驗設計必須與前述較優性試驗具有相同的設計重點(如主要評估指標、活性對照藥的劑量、納入排除條件等)，因活性對照藥物已在前述較優性試驗中清楚地顯示其療效，因此可將臨床與統計實務上進一步的考量納入新試驗中。

於驗證相等性或不劣性的試驗計畫書中，語意清楚明確地表述是極其重要的。計畫書必須訂定相等性臨界值，這個臨界值是臨床上所能接受的最大差異，並且應小於活性對照藥於較優性試驗所觀察到的差異。活性對照的相等性試驗須指定上界和下界，活性對照的不劣性試驗則只需要定下界。臨界值的選擇須考量其臨床上之合理性。

統計分析常採用信賴區間方法(見 5.5 節)。相等性試驗需要使用雙尾信賴區間，當信賴區間完全落在相等性臨界值之內時，則可推斷具有相等性。在運作上，這等於同時進行兩個單尾檢定，檢定的虛無假說為治療差異落在相等性臨界值之外，對立假說為治療差異落在相等性臨界值之內，因為這兩個虛無假說是無交集的，型一誤差可以被適當的控制。不劣性試驗則使用單尾信賴區間，這信賴區間方法是有一個單尾檢定，檢定的虛無假說是治療差異(研究用藥減對照組)等於或小於不劣性臨界值，而對立假說是治療差異大於不劣性臨界值。型一誤差大小的選擇需依單尾或雙尾分開考量，樣本數計算也是需要依據這些考量(見 3.5 節)。

虛無假說是假定研究用藥與活性對照藥沒有差異，若觀察到兩者差異不顯著而宣稱相等或不劣性是不適當的。

分析群體的選擇也是特別的議題，無論是治療組或是對照組的受試者失訪、或退出試驗通常與效果不佳有關。因而，全分析群體(見詞彙表)的分析結果將產生趨向顯示相等性的偏差(見 5.2.3 節)。

3.3.3. 劑量反應關係試驗

新的研究產品的劑量反應關係可在各時期(all phases)用多種方法(參見 ICH E4)得到答案。劑量反應關係的試驗有多重目標，其中特別重要的有：確認療效、建立劑量反應曲線的形狀和位置、估計適當的起始劑量、決定個別劑量調整的最佳決策、確定最大劑量，超過這個劑量不會有更多好處。為了這些目標，需要收集研究用藥不同劑量組的資料，如果可以的話，也包括安慰劑(零劑量)組的資料。為此，估計劑量與反應關係的方法如信賴區間的建構、圖形法的使用都與統計檢定同樣重要。而所用的假說檢定方法必須順應劑量的自然順序，或劑量反應曲線形狀(如單調性)的一些特殊問題。擬選用的統計方法需在計畫書中詳細說明。

3.4. 群集逐次設計 (Group Sequential Designs)

群集逐次設計是可用於期間分析的方法(見 4.5 節與詞彙表)。儘管群集逐次設計不是唯一可用於期間分析的設計，但它是被最常被使用的，因為它在試驗期間對某時段的受試者資料進行評估，比連續性的對每個受試者資料進行評估更可行。統計分析方法必須在獲得治療結果的訊息及受試者所分派的組別前(如解盲，見 4.5 節)在計畫書中完整的訂定。獨立數據監測委員會(見詞彙表)將負責群集逐次設計試驗之期間分析的執行或評估(見 4.6 節)。群集逐次設計已被廣泛地、成功地應用於大型、長期的試驗來觀察死亡率或非致命結果的療效指標，它在其他方面的應用也越來越廣泛。特別的是，由於已認識到在所有試驗中都必須監視安全性，故因安全性理由而提早終止試驗等正式程序也都需考慮到。

3.5. 樣本數

臨床試驗中的受試者數目必須足夠多，以確保能對所提出的問題給予可靠的答案。樣本數大小通常依試驗的主要目標來決定，如果是依其他基礎來決定，則須清楚說明並合理解釋。例如，一個試驗樣本數依據安全性問題、安全性要求或重要的次要目標來計算，則所需要的樣本數會大於依主要療效問題所計算的樣本數(見 ICH E1A)。

使用一般的方法來決定適當的樣本數需要先確定以下各項：主要指標、統計檢定量、虛無假說、所選劑量下的對立假說(包括對所選劑量和所選病人族群下要檢出或拒絕的治療差異的考量)、錯誤地拒絕虛無假說之機率(型一誤差)、錯誤地不拒絕虛無假說之機率(型二誤差)，以及對失訪、違反計畫書的處理方法。在有些情況下，事件率是估計檢定力的主要參數，因此需作一些假設來使所需要的事件數外推至試驗最終所需的樣本數。

計算樣本數的方法及計算時所需的統計估計值(如變異數、平均數、反應率、事件率、治療差異值)須明定於試驗計畫書中，亦須提供這些估計值的依據。探討樣本數對現有假設之各種偏離程度的敏感度是很重要的，這可以由假定一個合理的偏離區間而得到一個樣本數的範圍。在確認性試驗中，樣本數計算的假設一般需依據已刊載的文獻資料或早期試驗的結果來估算，治療差異值可基於臨床治療這類病人中，可接受的最小的療效或新治療方法預期的效果，看何者為大。按照慣例，型一誤差機率設在 5% 或更小，或者由多重比較考慮所需要的調整來決定，精確的選擇可能會受到先前合理的假說檢定與試驗所預期的結果之影響。型二誤差機率傳統上都定在 10% ~ 20%，廠商會希望在可行的範圍內越低越好，特別是當試驗很難或不可能再次執行時。在某些情況下，相較於傳統型一誤差、型二誤差設定的數值之外的設定，或許也可接受，甚至更好。

樣本數的計算須參照主要分析所需的受試者數目。如果是“全分析群體”，則估計的療效大小要比“依計畫書群體”(見詞彙表)還小，這是因為納入的受試者資料有失訪或順從性差的狀況，會稀釋治療效果，關於變異的假設亦需加以檢視。

相等性或不劣性試驗(見 3.3.2 節)的樣本數估計通常基於治療差異的信賴區間，此一信賴區間顯示治療差異最多只能到臨床上可接受的差異。當相等性試驗的檢定力是以真實差異為 0 時設定，如果真正的差異不為 0 的話，則欲達到此檢定力所需的樣本數將會被低估。當不劣性試驗的檢定力是以真實差異為 0 時設定，如果研究用藥的療效低於活性對照組時，則欲達到此檢定力所需的樣本數將會被低估。選擇一個”臨床上可接受”的差異需要解釋其對未來使用病人的意義，且可能小於前述較優性試驗中所建立的”臨床上有意義”差值。

群集逐次試驗中的樣本數在試驗前無法確定，其值須依所選定的試驗的停止條款與真實治療差異合併考量，試驗的停止條款須考量樣本數的範圍，通常這會落在期望和最大的樣本數之間。

當事件發生率低於預期或變異大到超乎預期，可藉由適當的統計方法，在不解盲或不對治療組間進行比較下重新估算樣本數(見 4.4 節)。

3.6. 數據收集與處理

數據的收集和傳送，從研究者到廠商可透過多種媒介，包括紙本病例記錄表、遠端監視系統、醫學電腦系統和電子傳輸器。無論採用何種儀器收集數據，數據的形式和內容必須與計畫書完全一致，且在試驗執行前訂定。應著重於收集預計執行分析所需的數據，包括確認計畫書順從性與嚴重違反計畫書定義的前後相關資訊(如服藥的時間)。缺失值(missing values)需與”0 值”和本質為無(characteristic absent)相區別。

數據收集的過程到資料庫的完成必須依照藥品優良臨床試驗規範(ICH E6，第 5 節)執行。尤其要以及時可靠的方式記錄數據、更正錯誤和增補遺漏是建立高品質資料庫、透過預計執行的分析達到試驗目標所必需的。

4. 試驗執行之考量

4.1. 試驗監測與期間分析

依照試驗計畫書謹慎地執行臨床試驗，對試驗結果的可信度有著重要的影響(見 ICH E6)。謹慎地監督試驗進行能及早發現問題，並使問題的發生或再發生達到最小。

由廠商發起的確認性臨床試驗有兩種不同的監測方法，一種是監測整個試驗的品質，另一種是以解盲方式進行治療效果的比較(如期間分析)。這兩種類型的監測方法，除人員職責不同外，所用的數據類型和資訊也不同，因此用於控制可能的統計學和操作上的偏差原則也不同。

為了瞭解試驗的品質，對試驗管理的監測應包括試驗是否按計畫書進行、收集到的數據質

量如何、是否達到預期收案的目標、設計的假設是否適當、是否成功地把病人留在試驗中等等(見 4.2 ~ 4.4 節)。這類的監測既不需要比較治療效果也不需要將資料解盲，所以對型一誤差沒有影響。為達到此一目的而對一個試驗進行監視是廠商的職責(見 ICH E6)。它可以由廠商或廠商指定的獨立單位進行，這種監測一般從研究地點的選定開始，到最後一位受試者數據的收集和清理結束。

另一類的監測(期間分析)涉及到治療結果的比較。期間分析需要對治療組別的分派及各治療組的摘要資訊進行解盲。因此，計畫書(或第一次分析前的計畫書變更)需包含期間分析採用的統計方法以防止某些偏差。這將在第 4.5 和 4.6 節討論。

4.2. 變更納入及排除條件

納入排除標準在試驗對象選擇的全過程中應按計畫書的定義保持不變，但有時對其做些變更也是適當的。例如在長期的臨床試驗中，無論是從外部或從本試驗期間分析中不斷獲得的醫學知識就有可能建議對納入條件進行變更。此外，標準的改變也可能來自監測人員的發現，如他們在監測中發現，常常不能按納入條件選擇受試者或由於太嚴格的納入條件導致收納速度非常慢。變更納入排除條件不能解盲且需於計畫書變更中寫明，內容包括對統計分析後續的影響，如不同的事件發生率導致須調整樣本數，或修改統計分析方法，如按變更的納入排除條件進行分層分析等。

4.3. 受試者收納速度

在受試者收納期間長的試驗中，必須對收納速度進行監視，如果收納速度明顯低於計畫書預先設定，則須查明理由並採取補救措施，以維持試驗的檢定力。在多中心試驗中，這些考量須應用到每一個中心。

4.4. 樣本數調整

在時間較長的臨床試驗中，常會對原設計及樣本數計算所基於的假設進行審視。如果試驗計畫是較為初步的或建立在不確定的資訊上，這種調整尤為重要。在不解盲的情況下，對數據進行期中檢查，可能發現所得總反應變異數、事件發生率或存活狀況與預期不符，則應適當地修訂假設條件，重新計算樣本數，並需清楚說明緣由，且於計畫書變更與試驗報告中載明。若曾為控制整體型一誤差及信賴區間寬度而採取某些措施以維持試驗的盲性及其後續結果，也須加以闡述。試驗有重新估算樣本數的潛在需要時，宜儘可能於計畫書中預先設定(見 3.5 節)。

4.5. 期間分析與試驗提早終止

任何在臨床試驗正式結束之前，為了比較治療組間有效性或安全性的分析稱為期間分析。

由於期間分析的次數、方法及結果將對試驗結論的闡釋產生影響，因此所有的期間分析必須預先計畫，並在試驗計畫書中寫明。有些特殊情況下，可能會需要執行未事先計畫的期間分析。在這種情況下，應在解盲分析前變更試驗計畫書增加期間分析的章節。如果一個期間分析是為決定是否終止試驗而設計，這通常採用以統計監督計畫為指導原則的群集逐次設計(group sequential design 見 3.4 節)。試驗因期間分析而提前終止的原因如下，研究治療組的較優性已被清楚地建立，或有意義的療效差異已經不可能被建立，或出現了無法耐受的不良反應。一般而言，為監測療效而設定的界限比監測安全性而設定的界限需要更多的證據才能讓試驗提早終止(即療效監測更保守)。當試驗計畫或監測目標包含了多個評估指標，則需有多重檢定的考量。

試驗計畫書應載明期間分析的日程，或執行期間分析的相關考量，例如採用彈性的 α 消耗函數(alpha spending function)，須於計畫書或第一次期間分析前之計畫書變更中寫明。試驗終止的規則及其特性須在試驗計畫書或變更中詳述，試驗提前終止而影響其他重要指標的分析也需要考量。試驗如果有數據監測委員會，則這些內容需由數據監測委員會撰寫或核准(見 4.6 節)。偏離預先計畫的程序常有使試驗結果變成無效的可能。如果試驗計畫書需要變更，則任何相應的統計方法便需儘早在計畫書變更中寫明，應特別就變更而對任何分析或推論所產生的影響進行討論。所選定的方法必須確保整體型一誤差機率得到控制。

期間分析的執行應該是完全保密的過程，因為它可能包含了解盲的資料及結果。所有參與試驗的人員必須對這類分析的結果保持盲性，以避免他們對試驗的態度可能改變，進而影響新納入受試者特徵的改變或對治療組別的比較產生偏差。這一原則可應用於除了直接參與執行期間分析的人員之外的所有研究人員和廠商聘雇人員。研究者僅僅被告知是繼續試驗、暫停試驗，或是對試驗程序進行調整。

大部分用以支持研究用藥的療效和安全性的臨床試驗，須完成預計收納的樣本數。只有在遇到倫理方面的原因或檢定力大小不再是可以接受時，方可終止試驗。儘管如此，大家已認識到，藥物研發計畫由於各種理由，廠商有查閱治療組間比較數據的需要，比如要設計其他的試驗。大家也公認到有一部分的試驗會採用嚴重威脅生命的指標或死亡率，基於倫理考量，這些都需要對不斷積累的比較性療效結果進行持續的監測。在以上這些情況中，統計相關的期間分析內容必須在解盲前於試驗計畫書或計畫書變更中載明，以避免可能產生的統計上和操作上的偏差。

對許多新藥臨床試驗，特別是與公眾健康關係重大的新藥，必須另外指定一個獨立的小組負責監督有效性和安全性結果的比較，並明確規範其職責。這個小組通常被稱為獨立數據監測委員會(Independent Data Monitoring Committee, IDMC)，或數據與安全性監測委員會(Data and Safety Monitoring Board)，或數據監測委員會(Data Monitoring Committee)。

當廠商負起監測有效性或安全性比較的職責，可從中獲得解盲的結果時，須特別注意保持試驗的完整性，並適當地管理和限制資訊分享。廠商必須保證並書面記錄內部監測委員遵

守了書面的標準操作程序，並保證且保留該委員會包括期間分析結果記錄在內的決策會議記錄。

任何設計不良的期間分析(不管是否提早終止試驗)都可能導致偏差的產生，所得的結論也缺乏可信度，因此，應避免這樣的分析。如果執行了計畫外的期間分析，必須在試驗報告中解釋其必要性與解盲的程度，並評估可能導致偏差的嚴重程度以及對結果解讀的影響。

4.6.獨立數據監測委員會(IDMC)之角色

獨立數據監測委員會可由廠商成立，它的主要任務是對臨床試驗的進展、安全性資料和主要療效指標持續地進行評估，建議廠商這個試驗是可以繼續進行、須調整設計或是終止試驗。IDMC 應當有書面的操作程序書，並保留每次的會議記錄，包括期間分析結果，在試驗結束時可供查閱。IDMC 的獨立性旨在它既能控制重要的比較試驗數據的共享，又能防止因接觸試驗資料而可能對整個試驗的完整性所產生的影響。IDMC 是與人體試驗委員會(Institutional Review Board, IRB)或獨立倫理委員會(Independent Ethics Committee, IEC)相獨立的機構，它的成員須有通曉包括統計學等有關相應學科專業知識的臨床試驗科學家。

當有廠商代表參與組成獨立數據監測委員會時，這些代表的權責應在委員會的操作程序書中加以明確規定(比如，在關鍵議題上是否具有投票權)。由於這些廠商代表能夠獲得解盲的資訊，因此在委員會的操作程序書中還須說明如何控制期間分析結果在廠商內部散佈。

5. 數據分析之考量

5.1.預先設定分析方法

在設計臨床試驗時，最終的主要統計分析方法須撰寫在計畫書中的統計章節。這個章節包括主要指標的確認性分析方法之所有主要特徵，及預期的分析問題之處理方式。對於探索性試驗，這一章節可描述更一般的原則及方向。

統計分析計畫(見詞彙表)可以在完成試驗計畫書後單獨成文，包含計畫書中主要特徵更多技術性的詳細闡述(見 7.1 節)。這個計畫可包括主要指標、次要指標及其他資料的統計分析方法之詳細執行步驟。在對試驗資料進行盲性評估(見 7.1 的定義)分析後，應對統計分析計畫進行再檢視和必要的變更，並應在解盲之前完成。統計分析計畫完成的時間點與各解盲的時間點都應保留正式的記錄。

如果盲性評估的結果是建議更改計畫書的主要設計，須變更試驗計畫書。否則，根據盲性評估的結果對統計分析計畫做出修改即可。但唯有按照試驗計畫書(包括計畫書變更)所設定的統計分析來執行，其結果才是可被認可的。

在臨床試驗報告的統計分析章節，必須詳細描述統計方法，包括試驗進展到甚麼時候才決

定分析方法的(見 ICH E3)。

5.2. 分析群體

要用於主要分析之受試者資料需在計畫書中的統計章節明確定義。另外，記錄所有受試者進入試驗(例如，納入期間)的時間是有用的，該記錄的內容須取決於試驗本身的特性，但至少人口學及病情的基礎值應盡可能地收集。

如果所有接受隨機分派進入臨床試驗的受試者均符合納入條件、完美的遵循試驗程序且無失訪，也提供了完整的資料記錄，很明顯地，這樣的受試者應都包括在分析中。試驗的設計與執行應盡可能地達到這一目標，但實際上要完全做到是不容易的。因此，試驗計畫書中統計章節應寫明預期會發生的問題，並說明這些對用於分析的受試者及數據有何影響。試驗計畫書中還必須定出一些做法，以減少試驗執行過程中任何預期的違規行為發生而損害分析的完善。這些行為包括各種類型的違反計畫書、退出治療及缺失值。試驗計畫書應考量如何減少這類問題發生的頻率，以及分析資料時出現這類問題的處理方式。在盲性評估時，應在可能有的計畫書變更中，寫明分析時對違反計畫書事件的處理方法。最好確認任何違反計畫書事件發生的時間、原因及對試驗結果的影響。違反計畫書事件的頻率、類型、缺失值，以及其他問題均須記錄於試驗報告中，它們對試驗結果可能產生的影響亦須論述(見 ICH E3)。

關於分析群體的決定需遵循以下兩個原則：(1)使偏差達到最小；(2)控制型一誤差的擴增。

5.2.1. 全分析群體

意圖治療(intention-to-treat，見詞彙表)的基本原則是：主要分析應包括所有接受隨機分派的受試者。遵守這一原則需要完整地追蹤所有接受隨機分派受試者的試驗結果。但實際上這一理想難以達到，理由後述。因此，在此文件中，以“全分析群體”來描述分析群體，是指盡可能完整並且盡可能接近包括所有接受隨機分派受試者的意圖治療理想。在分析中保留初始隨機分派在防止偏差並提供統計檢定安全的基礎是重要的。在很多臨床試驗中，使用“全分析群體”是保守的策略，在許多情況下它提供的治療效果估計更能反映爾後實際上的情況。

有少數情況可能導致從“全分析群體”中排除已接受隨機分派的受試者，包括不符合主要納入條件(違反納入資格)、沒有接受一次試驗用藥以及在隨機分派後沒有任何數據，排除這類受試者都需說明其合理性。不符合納入條件的受試者可以從分析中排除而不會造成偏差的只有以下一些狀況：

- (1) 在隨機分派之前已判定納入條件。
- (2) 可以完全客觀地找出相關的資格不符。
- (3) 所有受試者接受相同的資格違反檢查(這在開放性試驗中很難達到，即使在雙盲試驗

若資料在檢查之前解盲也是，此處強調盲性評估的重要性)。

(4) 所有違反特定條件的受試者已被排除。

在某些情況下，從所有接受隨機分派受試者中，排除從未用過試驗藥物的受試者可能是合理的。儘管排除這些受試者仍有可能維持意圖治療的原則，例如，是否開始治療的決定，不會受到因瞭解病人所接受的治療組別而受影響時。在其他情況下可能有必要從所有接受隨機分派受試者中排除任何隨機分派後沒有資料的受試者。除非這些排除，或任何其他原因引起的可能偏差都已考量，否則沒有一個分析是完整的。

當使用全分析群體時，在隨機分派後的違反計畫書可能對資料和結論造成影響，尤其是當它的發生與治療組別有關時。從多方面考慮，將這些病人的資料納入在分析中是恰當的、與意圖治療原則一致。特殊疑慮在於有些受試者接受一次或多次劑量後退出，而在此後沒有任何資料，以及有些因其他原因而退出試驗之受試者，因為在全分析群體中不包括這些受試者可能嚴重地破壞這個方法的基礎。因此，當主要指標是在病人基於任何理由失訪的情況下測定的，或是隨後根據計畫書預計的日程收集到的都是有價值的；當主要指標是死亡率或嚴重的發病率時，後續收集的資料特別重要，以這種方法意圖性收集資料應在計畫書中載明。插補的技巧，從最後觀察值遞延插補(carrying forward)，到應用複雜的數學模型的方法，都可以用來插補缺失值。其他用於可對全分析群體中每一個受試者的主要指標進行量測值的方法，可能需要對受試者的結果做一些假設，或者對結果做較簡單的選擇(如成功/失敗)。使用這些策略應在計畫書的統計章節描述並說明其選取之恰當性，所採用的數學模型其假設也應清楚說明。提供相關分析結果的穩健性也是很重要的，尤其是當採用的策略可能導致治療效果的估計有偏差。

由於某些問題的不可預測性，有時針對試驗期間所出現的無規律性情況，推遲到試驗結束對數據做盲性評估後，才做出詳細考量會更好，若打算這樣做，應在試驗計畫書中清楚說明。

5.2.2. 依計畫書群體

依計畫書群體的受試者有時又稱為有效病例、療效樣本或可評估受試者樣本，它是全分析群體的一個子集，這些受試者對計畫書更具順從性，並符合以下特徵：

- (i) 完成預定的治療療程的最小規定部分。
- (ii) 具有主要指標的量測值。
- (iii) 沒有任何重大的違反計畫書事件，包括違反納入條件。

將受試者排除在依計畫書群體外的理由應清楚說明，並應以適合這特定試驗情況的方式，在解盲前詳盡寫明。

採用依計畫書群體可能使新的治療在分析中展現附加效果的機會最大化，並且更密切地反

映了計畫書下的科學模型。然而，相應的假說檢定和治療效果的估計，依據試驗的不同可能是保守的或是不保守的。依計畫書群體分析所產生的偏差有時極為嚴重，主要導因於遵循計畫書可能與接受的治療以及治療效果有關。

排除受試者於依計畫書群體之外的理由，以及其他的違反計畫書事件，應當清楚記載並加以總結。有關的違反計畫書事件可能包括治療分派錯誤、使用不許使用的藥物、順從性不佳、失訪和資料缺失。評估各治療組間關於這些事件發生的頻率和發生的時間之模式是良好的做法。

5.2.3. 不同分析群體的角色

一般說來，顯示出不同分析群體對試驗主要分析結果的不敏感性是很好的。在確認性試驗中，同時對全分析群體及依計畫書群體進行分析一般來說是恰當的，由此可以對他們之間的任何差異進行明確的討論和解釋。在有些情況下，最好能選擇不同的分析群體進行對結論的敏感性探索。當全分析群體和依計畫書群體得出實質上是相同的結論時，試驗結果的可信度會增加。然需銘記在心的是，依計畫書群體若排除大量的受試者則會對試驗的整體有效性產生疑慮。

在較優性試驗(為了想要顯示研究產品的效果更好)和相等或不劣性試驗(為了想要顯示研究產品具有可比性，見 3.3.2 節)，全分析群體和依計畫書群體中扮演不同的角色。在較優性試驗中，全分析群體用於主要分析(除了特殊情況)，因為它傾向避免來自於依計畫書群體療效估計值的過度樂觀，這是因為全分析群體包括了順從性不佳的受試者，這類受試者一般會沖淡療效的估計。然而，在一個相等性或不劣性試驗中，採用全分析群體一般而言並不保守，而其角色應當非常仔細地考慮。

5.3. 缺失值及異常值

缺失值是臨床試驗中會產生偏差的一個原因，因此，必須盡一切努力完成試驗計畫書中所有關於搜集資料和數據管理的各項要求。然而，事實上任何試驗幾乎不可避免地總有缺失值。不過，試驗處理缺失值的方法若合理且在計畫書中已經預先定義，仍可認為是有效的試驗。在盲性評估時，統計分析計畫可更新這方面的內容使其更完善。遺憾的是，尚無一個普遍適用的缺失值處理方法可供推薦。需探討不同缺失值處理方法對分析結果的敏感性，特別是有大量缺失值的時候。

類似的方法也應用於探討異常值的影響。統計學上對異常值的定義在某種程度上來講是主觀認定的。從醫學上和統計學上共同清楚地判斷某一個特定值是異常值是最可信的，而醫學上的認定的異常值常會採取適當的後續行動。在計畫書或統計分析計畫中訂定異常值的處理方式不應當對任何一個治療組別有偏袒。同樣，這方面的分析計畫也常在資料的盲性評估時進行有用的更新。如果在試驗計畫書中未事先訂定處理異常值的方法，則用實際資

料分析所得結果，與排除或減弱異常值影響後的至少一個分析結果均須提供，並對結果不一致之處加以討論。

5.4. 數據轉換

對重要的指標是否要進行轉換，最好根據先前類似試驗的資料，並在試驗設計時就決定好。擬採用的轉換(如平方根、對數)及其原理需在試驗計畫書中說明，特別是針對主要指標。一般轉換原則是為了確保資料能滿足統計方法的假設，這在標準教科書上均能找到。一些特定指標常用的轉換方法已在某些特定的臨床領域中使用。一個指標是否轉換，以及如何轉換，常受到臨床解釋方便性的影響。

類似的考量也須應用到衍生的指標(derived variables)，比如從基礎值的改變量、從基礎值的改變百分比、重複測量曲線下的面積或兩個不同指標的比值。惟須仔細考量後續在臨床上的解讀。衍生的指標推導的方式需在試驗計畫書中清楚說明。與此密切相關的重點已在 2.2.2 節討論過。

5.5. 參數估計、信賴區間及假說檢定

試驗計畫書中的統計章節應說明為滿足試驗主要目標，所要檢定的試驗假說和所要估計的治療效果。為完成這個任務，主要指標(最好也有次要指標)的統計分析方法，以及所採用的統計方法或模式應清楚闡述；如可能，治療效果的估計值應盡可能附上信賴區間，並說明其計算方法。若想採用基礎值以提高估計精準度，或對基礎值差異估計值進行調整，例如使用共變數分析，亦需在試驗計畫書中寫明。

明確說明所採用的假說檢定是單尾還是雙尾是非常重要的，尤其是要採用單尾檢定更需事先說明。如果認為假說檢定不適當，則須提出其他可以得到統計結論的方法。關於統計推論用單尾還是雙尾是有爭議的，在統計文獻中可見到不同的觀點。單尾檢定的型一誤差設定為雙尾檢定型一誤差的一半是醫藥品查驗中心比較建議的，這與通常用於估計兩種治療組間差異大小的雙尾信賴區間相一致。

所選用的統計模型應當要反映目前醫學和統計學關於所分析的指標以及試驗的統計設計的知識。所有在分析(例如在變異數分析模型)中納入的影響因子應完整地說明，同樣地，若初步分析後影響因子有變更，也須完整解釋。相同的考量也需應用在共變數分析中的共變數設定(見 5.7 節)。在選擇統計方法時，應注意到主要和次要指標的統計分布，在選擇時(例如選擇母數還是無母數方法)，應記得提供治療效果的大小及其信賴區間(除了提供統計顯著檢定之外)。

主要指標的主要分析應該與主要或次要指標的支持性分析區分。在計畫書的統計章節或統

計分析計畫中，也須概述除了主要和次要指標之外的資料總結和報告的方法，亦應包括為達到跨試驗比較的一致性所採用方法的文獻，比如安全性資料的比較。

以建立模型的方式納入已知的藥理學參數、受試者對計畫書的順從性或其他生物學上的數據，可以對實際上的可能的效果，特別是對於治療效果的估計，提供有價值的見解。這類模型所基於的假設應清楚地加以說明，而任何結論的侷限性也應仔細地描述。

5.6. 顯著水準的調整

當出現多重檢定時，頻率學派常用的方法會是調整型一誤差。多重檢定的問題可能由以下情況而產生，例如多個主要指標(見 2.2.2 節)、治療組間的多重比較、隨著時間的重複量測、期間分析(見 4.6 節)等。如果可以的話，最好採用避免或減少多重檢定問題的方法，比如指定一個重要的主要指標(多重變量)、選擇關鍵的治療組比較(多重比較問題)、運用綜合指標如”曲線下面積”(重複量測問題)。在確認性分析中，作了這一類的處理後若仍有多重檢定的問題，則需在試驗計畫書中說明。分析計畫必須寫明調整型一誤差的詳細步驟，若不需調整也要說明理由。

5.7. 次群體、交互作用與共變數

主要指標通常系統性地受到治療以外的其他因素影響，例如，與共變量有關的(如年齡、性別)，或有差異的次群體(如多中心試驗，不同中心所治療的受試者)。在有些情況下，對共變數的影響或次群體效應的調整是分析計畫中不可或缺的一部分，故需在試驗計畫書中陳述。試驗前須審慎識別可能對主要指標有重要影響的共變數和因子，並且應考慮如何將其納入分析以提高估計的精準度，以及補償治療組間不平衡所產生的影響。如果設計了一個或多個分層因子，則在分析中應納入這些因子。當調整因子數值遭質疑時，建議將未經調整的分析結果作為主要結果，而調整後的分析結果作為支持性證據。特別要注意試驗中心的效果及基礎值在主要指標的角色。不建議主要分析對隨機分派後測量的共變數值進行調整，因為它們可能受治療的影響。

治療效果的大小可能會因次群體或共變數的不同而不同，例如，治療效果可能會隨年齡的增加而減少，或對某一類受試者較大。在有些情況下，這種交互作用可被預期或特別感興趣(如老年醫學)。因此，次群體分析或統計模型中包含交互作用項都是屬於確認性分析的一部分。然而，在大多數情況下，次群體分析和交互作用分析都是探索性的，且應當清楚地如是定義，它們是用來探索整體治療效果在不同情況下都是一樣的。一般而言，這類分析首先應在研究的統計模型中增加交互作用項，再以相關的次群體分析或共變數定義的層內分析作為附加的探索性分析加以補充。進行探索性分析時，這些分析的解讀都必須非常小心，任何僅基於次群體探索性分析的療效(或無效)結論或安全性結論，都不太會被接受。

5.8. 數據完整性與電腦軟體的效度

數據分析結果的可信度取決於用於數據處理(包括輸入、儲存、查核、更正和檢索)和用於處理數據的方法以及軟體(內部和外部的)的品質和有效性。因此，數據處理必須遵循完善、有效的標準作業程序。用於數據處理和統計分析的電腦軟體必須是可靠的，並提供適當的軟體檢驗過程的文件。

6. 安全性與耐受性之評估

6.1. 評估範圍

在所有的臨床試驗中，安全性及耐受性(見詞彙表)的評估相當重要。在早期的試驗，這種評估多半是探索性質，且只對明顯表現的毒性敏感。而在晚期試驗，藥物的安全性和耐受性可由較多的受試者經驗更全面地瞭解。晚期的對照組試驗，代表了一個重要的、以無偏的方式探索任何新的潛在藥物不良反應的方法，儘管這類試驗一般在這方面都缺乏檢定力。

某些試驗設計是為了宣稱試驗藥物在安全性與耐受性方面與其他藥物或該藥物的其他劑量相比具有較優性或相等性。這類宣稱需要確認性試驗的相關證據來支持，這跟宣稱療效的要求是一樣的。

6.2. 指標選擇與數據收集

在任何一個臨床試驗中，用於評估一種藥物的安全性和耐受性的方法及量測準則依賴以下一些因素，包括與之相近的藥物之不良反應知識、非臨床試驗或早期臨床試驗的資訊、特殊藥物可能的藥效學及藥動學特性、服藥方式、所研究的受試者類型及試驗的期間等。實驗室檢驗包括生化學和血液學、生命徵象及臨床不良事件(疾病、徵象和症狀)，這些通常構成了安全性與耐受性資料的主體部分。嚴重不良事件及導致終止試驗治療的不良事件在法規審查上特別重要(見 ICH E2A 及 ICH E3)。

此外，建議在整個臨床試驗規劃中，採用一致的資料收集與評估方法，方便日後可以合併不同試驗的資料。採用一個共通的不良事件分類詞典是很重要的，該詞典的編排應將不良事件數據整理為三種不同層級的分類，即系統-器官分類、選用詞與納入詞(見詞彙表)。在描述不良事件的整合性資料分析時，通常先以選用詞來統一彙整所記錄的不良事件，然後在數據的描述性表達時，將屬於同一系統-器官級的選用詞進行合併(見 ICH M1)。

6.3. 評估群體與數據的呈現

評估整體安全性與耐受性時，用於總結的分析群體常定義為至少接受了一劑研究藥物。從這些受試者中收集的安全性及耐受性指標應盡可能地周詳，包括不良事件類型、嚴重程度、發生時間及持續時間(見 ICH E2B)。某些特定的次族群可能需要額外的安全性及耐受性的

評估，如女性、老年人(見 ICH E7)、重病患者，或接受常見併用藥物的受試者，這些評估可能需要說明具體的議題(見 ICH E3)。

所有安全性及耐受性指標在評估時都須十分重視，所用的分析方法需在計畫書中載明。所有的不良事件均須報告，無論其是否被認為與治療有關。試驗群體中所有可用的資料均需納入評估。實驗室指標的測量單位及參考值範圍必須認真制訂，如果在同一試驗出現不同單位及不同的參考值範圍(如多個實驗室參與研究)，則須進行適當的標準化以便於統一評估。毒性分級尺度也必須事先訂定並說明。

某不良事件的發生率通常是以出現不良事件的病人數與暴露病人數之比來表示。然而，如何評估發生率並不總是十分清楚，例如，根據不同情況，可考慮用暴露病例數或暴露程度(用人年表示)作為分母。無論目的是用來估計風險還是進行治療組間的比較，在試驗計畫書中定義計算的方式是很重要的。尤其是長期的治療，估計會有較大的退出治療的比例及死亡比例時，對這類情況，需考慮採用存活分析，並計算累積不良事件比率，以避免低估的風險。

當徵象和症狀易受背景雜訊干擾時(如精神疾病的臨床試驗)，需考慮採用可涵蓋這些影響的方法來估計不同不良事件的風險。其中一種方法是運用”治療後發生的事件”(見詞彙表)的概念，當該不良事件相對於治療前基礎值為新發生、或惡化時才會被記錄。

其他可以消除背景雜訊的方法也是適當的，比如忽略程度輕微的不良事件，或在重複訪視時才觀察到的不良事件方可列入分子計算。這些方法需在試驗計畫書中解釋並說明其合理性。

6.4. 統計評估

安全性與耐受性的研究是一個多方面的問題。儘管某些藥物的不良反應通常可被預期而進行監測，但藥物的不良反應可能範圍很廣，而新的、未預期的不良反應總是有可能發生。更甚者，違反試驗計畫書後(比如使用禁用藥物)才發生的不良事件也可能導致偏差，此種情況造成統計在分析與評估藥物安全性及耐受性上的困難。這意味著由確認性臨床試驗得到有關安全性的結論只是一種例外而不是通例。

在大多數試驗中，最好用描述性統計來評估藥物安全性與耐受性，並輔以信賴區間以利解讀。用圖形來表示治療組內及受試者間的不良事件型態亦有價值。

計算 p 值有時也有幫助，既有利於評估感興趣的差異，對於大量安全性與耐受性指標又可作為一種標記工具以突出值得進一步注意的差異，這對實驗室數據特別有用，因為很難有其他適當的彙整方法。建議實驗室數據要作定量分析，如估計治療效果的平均數，也要作定性分析，計算高於或低於某一閾值的受試者個數。

如使用假說檢定，則在進行多重比較時，須進行統計上的修正以控制型一誤差，但通常也留意型二誤差的大小。如未對多重檢定問題進行修正，在解讀是否具統計上有意義的結果時需特別小心。

在大多數臨床試驗中，研究者希望確定，與活性對照藥或安慰劑相比時，安全性及耐受性未出現臨床上不可接受的差異。就如同於不劣性或相等性療效評估，此種情況採用信賴區間較假說檢定來得適當。如此，因發生頻率較低而造成較大的不精確性可被清楚地呈現。

6.5. 綜合總結

在藥物的開發過程，特別是進行上市申請時，藥物的安全性與耐受性通常是跨試驗連續整合出來的，然而，上述整合需基於適當、良好對照組、高品質資料的各別試驗。

一個藥物的整體評估即風險與利益的平衡；通常是在完成整個臨床研發計畫時進行風險效益評估，但對於單一試驗亦可進行風險效益評估（見 7.2.2 節）。

有關安全性與耐受性報告中所需要之更詳細的內容請見 ICH E3 的第 12 章。

7. 報告撰寫

7.1. 評估與報告

正如緒論所述，臨床試驗報告的格式與內容是 ICH E3 的內容，該指導原則涵蓋了統計工作的報告撰寫，亦適當地整合臨床及其他內容，本節只作簡單的討論。

如第 5 章所述，在試驗設計階段，統計分析方法的主要特點應在計畫書中寫明。當試驗試驗結束、資料已收集完整可作初步檢視時，如第 5 章所描述的，對數據按計畫好的分析進行盲性評估是有價值的。這種對治療的組別保持盲性的預先分析，應當包括關於以下一些問題的決定，例如，從分析群體中排除受試者或數據、檢視可能的指標轉換方法、異常值的定義、將其他最新研究中確定的重要共變量增加到統計模型中、重新考慮用母數方法或無母數方法。此時所做的決定須寫入試驗報告中，並與統計人員在知道隨機編碼後，所做的決定相區別；因為在盲性下所做的決定一般導入偏差的可能性較小，參與解盲性期間分析的統計學家或其他人員，不應參加盲性評估或修改統計分析計畫。當治療所誘發的效應可能在數據中顯示出來進而威脅到盲性時，盲性評估需要特別小心。

許多更詳細的報告和列表方式應當在盲性評估時完成最終版，以便在實際分析資料時，整個計畫的各方面都已確立，包括受試者的篩選、資料的篩選與修正、資料的整合與列表、參數估計及假說檢定。一旦資料評估完成，則分析即可按照預定的計畫進行，愈能依循分析計畫執行，所得結論的可信度就愈大。當實際分析有別於試驗計畫書、計畫書變更、及

對資料進行盲性評估時所確定的統計分析計畫時，需特別注意。偏離計畫的統計分析必須提出詳盡的解釋。

所有進入臨床試驗的受試者均須在研究報告中呈現，不管其是否納入統計分析。所有被排除在分析之外的理由也均須寫明。任何一個包含在全分析群體但不包含在依計畫書群體的受試者，亦須寫明其被排除在依計畫書群體的理由。同樣的，所有納入分析集的受試者，其重要指標的所有測量值均須記錄其量測時間點。

所有病患或數據的缺失、退出治療及重大違反計畫書等情況對主要指標分析的影響必須仔細評估。受試者的失訪、退出治療或嚴重違反計畫書必須一一指出，並對其進行描述性統計，包括退出的理由與其治療及結果的關連。

描述性統計是報告中不可或缺的部分。應用適當的表格或圖形清晰地呈現主要指標、次要指標、重要預後因子及人口學指標的重要特徵。與試驗目標相關的主要分析結果應當是報告中需要特別仔細描述的部分。在撰寫統計顯著性檢定結果時，應寫出精確的 p 值(如 $p = 0.034$)，而不是只依臨界值列出較優的結論。

儘管臨床試驗分析的主要目的是回答主要目標提出的問題，但在解盲性分析時基於觀察到的資料又會出現一些新問題，這時就需要採用其他或更複雜的統計分析方法來處理。在報告中，這些附加的內容必須與計畫書預定的分析內容嚴格區分開來。

由於機會的作用，可能導致治療組間的基礎值出現未預期的不平衡，而在原計畫的分析中沒有預定為共變數，但它對預後還是有一定的重要意義。處理這種不平衡的最佳方法是一種附加的統計分析，說明在考量這些不平衡因素後可以得出與原計畫的分析一致的結論。如果經過如上處理不能得出一致的結論，則須討論這種不平衡對結果的影響。

一般而言，非計畫中的分析應儘量少用。如果認為治療效果可能隨其他某個或某些因素的改變而有不同時，常進行這種非計畫中的分析。這時可能是企圖找出治療效果特別好的次族群，然而過度解讀計畫外的次族群分析結果的潛在危險是眾所皆知的(見 5.7 節)，應設法小心地避免。儘管類似的解讀問題也出現在治療無效或次族群有副作用時，但我們應對其可能性做出適當的評估並加以報告。

最後，統計學上的判斷對臨床試驗結果的分析、解讀及呈現有關。為此，試驗統計人員應該是臨床試驗報告負責團隊的成員之一，並核准臨床試驗報告。

7.2. 臨床試驗資料摘列

藥品申請上市時，需要提供所有臨床試驗關於安全性和療效證據的總結和綜合結論(如歐盟的專家報告、美國的整合報告和日本的概要)，適當時也會附上統計合併的分析結果。

在總結中，應包括如下專門的統計學分析內容：參與臨床試驗治療的受試者之人口學特徵和臨床表徵之描述；根據有關的試驗結果(一般是有對照組的試驗)提出療效的重要問題，且著重說明其一致和不一致的程度；總結所有為上市申請準備的臨床試驗安全性資訊，並找出可能的安全性議題。在臨床試驗設計規劃時，必須注重指標的定義及量測值收集之一致性，這將有利於後續試驗結果的解讀，特別是進行跨試驗合併分析時。必須選用一個共同的詞典來記錄用藥詳情、病史及不良事件，對主要指標與次要指標採用共同的定義往往是有利的，而且是後續統合分析的必要條件。測量重要療效指標的方法、評估進行隨機分派/進入試驗的時間、處理違反或偏離試驗計畫書的方法以及可能的預後因子定義，都必須保持一致，除非有充分理由不這麼做。

任何用於跨試驗合併資料分析的統計方法均需詳細描述。對於因試驗的選擇而可能導致的偏差、對於試驗結果的同質性以及對於不同的變異來源建立恰當的模型都必須十分注意。應當探索在不同假設及不同選擇下分析結果的敏感度。

7.2.1. 療效資料

每一個臨床試驗的樣本數都必須足夠多，以確保達到預期的目標。額外有價值的資訊可能來自於整合一系列臨床試驗的結果，這些臨床試驗本質上皆探討相同的重要療效問題，這一系列試驗的主要結果應當以統一的格式呈現以便於比較，一般是用表格或圖形的方式呈現估計值和信賴區間。使用統合分析(meta-analysis)的技術來合併這些估計值就是一個很好的方法，因為該方法為整體治療效果提供一個更精準的估計，也提供一個完整而簡潔的總結。在某些特殊情況下，統合分析也可能是最合適或唯一的方法來提供足夠的整體療效證據進行整體假說檢定，若是為這樣的目的而使用統合分析時，需有其專屬的計畫書。

7.2.2. 安全性資料

在總結安全性資料時，徹底檢查安全性資料庫中任何藥物潛在的毒性，及追蹤尋找相關的支持性觀察模式是非常重要的。將所有暴露於新藥的受試者的安全性資料合併起來可提供重要的資訊來源，因為大樣本為較罕見的不良事件提供最佳的檢出機會，也許還可估計約略的發生率。由於缺乏對照組，很難對這一類資料庫得到的發生率進行評估，而對照組試驗的資料在克服這方面的困難就顯得特別有價值。針對採用共同對照組(安慰劑或特定的活性對照組)的試驗結果應該合併，並對每一個對照組提供充分的數據資料。

探索資料時發現的任何藥物潛在的毒性均須報告。在評估這些潛在不良反應時，若有多個比較，也需考量多重檢定問題。評估時也可適當地運用存活分析方法，探索不良事件的發生率與服藥時間及隨訪時間(follow-up)之間的潛在關係。對於已辨識出的不良反應，其相關連的風險必需適當地量化，以便於進行風險效益評估。

詞彙表

貝氏方法(Bayesian Approaches)：一種資料分析的方法，針對某些參數(如治療效果)，由觀察資料以及驗前機率分布所推得出的驗後機率分佈作為統計推論之基礎。

偏差(統計上和操作上的)(Bias [Statistical and Operational])：與臨床試驗設計、執行、分析和結果評估相關的，會造成系統性地使估計值偏離真實數值的任何因素。由執行不正確所造成的偏差稱為操作上的偏差，其他原因造成的偏差則稱為統計上的偏差。

盲性評估(Blind Review)：

在試驗完成(最後一例病人的最後一次觀察)與解盲之間對資料進行查核和評估，目的是為了完成分析計畫。

內容效度(Content Validity)

一個指標(如評分尺度指標)測量其應該測量的大小之程度。

雙虛擬(Double-Dummy)

在臨床試驗中當兩組治療不能做到一樣時，使提供的藥物仍能維持盲性的一種技術。所準備的藥物必須為活性藥物 A 及其相同無法分辨之安慰劑、活性藥物 B 及其相同無法分辨之安慰劑，受試者將接受 A、B 兩種藥物：活性藥物 A 及安慰劑 B，或是安慰劑 A 和活性藥物 B。

退出(Dropout)

臨床試驗中的受試者，由於任何原因不能繼續進行試驗到依計畫所規定的最後一次回診。

相(對)等性試驗(Equivalence Trial)

一個試驗的主要目的是要顯示兩種或多種治療的差異是臨床上可接受的。這通常經由計算真實的治療組間差異落在相(對)等性臨界值之間來推論。

頻率學派方法(Frequentist Methods)

統計方法，如顯著性檢定和信賴區間，可解讀為相同之試驗環境若重覆實現時，某些結果發生的頻率。

全分析群體(Full Analysis Set)

盡可能接近按意圖治療原則的理想病例集。由所有接受隨機分派的受試者，以最少且合理的方法排除病例得出。

廣義性(Generalisability, Generalisation)

一個臨床試驗的結果可以被可信地由參加試驗的受試者族群，外推到更廣大的病人族群和

更廣泛的臨床環境的程度。

整體評估指標(Global Assessment Variable)

單一指標，通常是把客觀的指標和研究者對受試者的狀況或者狀態的改變情形整合起來的順序分類等級尺度。

獨立數據監測委員會(資料和安全性監測小組、監測委員會、數據監測委員會)(Independent Data Monitoring Committee [IDMC], Data and Safety Monitoring Board, Monitoring Committee, Data Monitoring Committee)

一個獨立的數據監測委員會可以是由廠商成立，定期評估試驗進度、安全性數據與重要的療效指標，並可以向廠商提出建議是否繼續、修正或停止試驗。

意圖治療原則(Intention To Treat Principle)

一種依安排好的治療方案而不是依據實際接受的治療為基礎之評估方式，這是可以對治療效果作出最好評估的原則。其結果是分到任一組別的受試者即應作為該組的成員接受訪視、評估與分析，而不論他們是否遵從計畫中的治療。

交互作用(定性和定量)(Interaction, Qualitative and Quantitative)

一種治療差異(如研究產品與對照組的差異)會由另一個因素(如試驗中心)而決定的情況。定量交互作用指的是治療差異的大小在因素的不同水平時(如不同試驗中心)會有不同，而定性交互作用指的是治療差異的方向性至少在因素的某一水平上(如某一試驗中心)不同。

評估者間信度(Inter-Rater Reliability)

不同評估者在不同情況下產生相同結果的性能。

評估者內信度(Intra-Rater Reliability)

同一評估者在不同情況下產生相同結果的性能。

期間分析(Interim Analysis)

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。

統合分析(Meta Analysis)

對同一個問題的兩個或更多試驗的定量證據進行正式評估。這常常是從各試驗的摘要統計量進行統計上的合併分析，但此名詞有時也用於原始數據的合併分析。

多中心試驗(Multicenter trial)

按照同一個計畫書在多個試驗中心所進行的臨床試驗。因而，由多個試驗主持人執行。

不劣性試驗(Non-inferiority trial)

主要目的為顯示研究產品的反應在臨床上不劣於對照組(活性或安慰劑對照)的試驗。

選用詞與納入詞(Preferred and Included Terms)

在一個多層次的醫學名詞分類法中，例如 MedDRA，納入詞(included terms)是最低階的醫學名詞，直接以研究者對不良事件的描述詞進行編碼。選用詞(preferred terms)是對納入詞進行分類與統整，用於通報事件發生的頻率。例如，研究者寫的是“左臂疼痛”，納入詞編碼為“關節疼痛(joint pain)”，在選用詞可通報為“關節痛(arthralgia)”。

依計畫書群體(有效病例，有效樣本，可評估受試者樣本)(Per Protocol Set [Valid Cases, Efficacy Sample, Evaluable Subjects Sample])

由能充分遵從計畫書以確保這些數據會按所基於的科學模型而表現出治療效果之病例子集所產生的數據集。遵從性包括以下的考量，如使用過治療藥物、有可用的測量值，以及沒有重大違反計畫書事件。

安全性和耐受性(Safety and Tolerability)

藥物的安全性關係到病患醫療風險時，試驗中通常以實驗室檢查(包括生化與血液學)、生命徵象、不良事件(疾病、徵象和症狀)，以及其他專門的安全性檢查(例如心電圖、眼科檢查)等來評估。藥物的耐受性代表了病患所能忍受明顯不良反應的程度。

統計分析計畫(Statistical analysis plan)

統計分析計畫是包括比計畫書所描述的統計分析的主要特徵，更多技術性和更多詳細細節的文件，並且包括了對主要、次要指標及其他數據進行統計分析的詳細過程。

較優性試驗(Superiority Trial)

主要目的是顯示研究產品的反應優於對照組(活性或安慰劑對照)的試驗。

替代指標(Surrogate Variable)

當直接測量臨床效果是不可能或不實際時，所提供的間接測量的指標。

治療效果(Treatment Effect)

在臨床試驗中歸因於治療所產生的效果。在大多數臨床試驗中感興趣的治療效果是兩個或多個治療組間的比較。

治療後發生的事件(Treatment Emergent)

用藥前沒有，但在試驗用藥期間出現的事件；或是用藥前已有，但用藥期間更加惡化的事件，都稱為治療後發生的事件。

試驗統計人員(Trial statistician)

經過教育、培訓並有足夠經驗以貫徹本指導原則，並負責試驗統計方面的統計學家。