

**ICH E9 (R1)：臨床試驗之估計目標與敏感
度分析**

Addendum on Estimands and Sensitivity

Analysis in Clinical Trials

**To the Guideline on Statistical Principles for
Clinical Trials**

(草案)

衛生福利部食品藥物管理署

中華民國 115 年 0 月

目錄

A.1. 目的與範圍	1
A.2. 一個將規劃、設計、執行、分析與詮釋連結一致的框架	3
A.3. 估計目標	3
A.3.1. 由臨床關注的問題來反映並發事件	4
A.3.2. 定義臨床關注問題時用以解決並發事件的策略	5
A.3.3. 估計目標的屬性	7
A.3.4. 建立估計目標的考量	8
A.4. 對試驗設計與執行的影響	10
A.5. 對試驗分析的影響	12
A.5.1. 主要估計	12
A.5.2. 敏感度分析	13
A.5.2.1. 敏感度分析的角色	13
A.5.2.2. 敏感度分析的選擇	14
A.5.3. 補充分析	14
A.6. 紀錄估計目標和敏感度分析	15
辭彙表	16

ICH E9 (R1)：臨床試驗之估計目標 (estimand)與敏感度分析 (sensitivity analysis)

A.1. 目的與範圍

為了使製藥公司、法規單位、病人、醫師及其他利益相關者能做出適當的決策，應清楚說明治療特定疾病之藥品的效益與風險；若未能清楚地闡明，所報告的「療效 (treatment effect)」將有被誤解的疑慮。本補遺提出一個結構化的框架，使參與制定臨床試驗目標、設計、執行、分析與詮釋的各學科間，以及廠商與法規單位間，就臨床試驗感興趣的療效，加強彼此的對話。

欲精確地描述試驗感興趣的療效，須建構能反映臨床關注問題的「估計目標(estimand)」(參見辭彙表，A.3.章節)。而估計目標的清楚性，須對「並發事件(intercurrent events)」(參見辭彙表，A.3.1 章節) 有周全的設想。並發事件包括終止指定的治療、使用額外或替代治療或發生終端事件 (例如死亡) 等。估計目標的描述須反映與這些並發事件相關的臨床問題，本補遺介紹多個策略以反映可能被提出不同的臨床關注的問題。策略的選擇亦會影響臨床試驗的常規屬性 (attribute)，例如治療處置 (treatment)、群體 (population)、或變數 (variable)/評估指標 (endpoint)，如何用於反映臨床關注的問題。

臨床試驗資料的統計分析應與估計目標一致。本補遺亦闡明「敏感度分析(sensitivity analysis)」(參見辭彙表) 在探討主要分析結論的穩健性 (robustness)上所扮演的角色。

整個補遺，引用原「臨床試驗統計指導原則 (ICH E9)」章節時以 x.y. 表示，引用本補遺章節時則以 A.x.y. 表示。

本補遺闡明並延伸「臨床試驗統計指導原則」以下的幾個議題：

議題一、對於一個隨機分派、具對照組的試驗，「臨床試驗統計指導原則」提出「意圖治療 (Intention-To-Treat, ITT)」原則，此與治療政策 (treatment policy)所估計的療效相關。在此原則下，受試者無論是否遵循原計畫的治療療程，仍須被追蹤、評估、並納入統計分析。此原則保留了隨機分派的好性質，提供統計檢定穩固的基礎。然而，ITT 原則亦被認為帶來多項結果。首先，試驗統計分析應包含所有和研究問題相關的受試者。其次，受試者應以原隨機分派的組別納入分析。第三，直接取自 ITT 原則的定義(參見「臨床試驗統計指導原則」辭彙表)，受試者無論遵循計畫的治療療程與否，都應持續地被追蹤與評估，且此等評估都須納入分析。無庸置疑，隨機分派乃是具對照組試驗的基石，且試驗統計分析應以盡可能利用隨機分派的優點到極致為目標；然而，符合 ITT 原則的療效估計是否總是代表與法規和臨床決策最為相關的療效，仍然值得存疑。本補遺所提出的框架針對不同的療效提供論述的基礎，並提出設計與分析須考量的重點，以便對療效進行可靠的估計，並為決策提供依據。

議題二、本補遺重新檢視資料處理與「缺失資料 (missing data)」(參見辭彙表)的議題，並做了兩個重要的區分。首先，本補遺認為終止隨機分派治療與退出試驗 (study withdrawal) 應有所區別，前者為一個並發事件，須透過估計目標明確地設定試驗的目標來解決；後者則產生了缺失資料，須藉由統計分析來解決。舉例來說，於一項腫瘤試驗，其中一位受試者轉換了治療處置，另一位受試者則因試驗已經結束，未能觀察到療效相關的結果事件 (outcome event)。在這個案例中，前者為一個並發事件，有關此並發事件的臨床關注的問題須清楚地闡明；後者則為行政設限 (administrative censoring)，須於統計分析中以缺失資料的問題來解決。明確的估計目標對於規劃哪些資料應該被收集提供了依據，若該收集而未被收集，所呈現缺失資料的問題，須要在統計分析中解決。反之，解決缺失資料方法的選擇，亦應與估計目標一致。其次，本補遺強調不同的並發事件所造成的結果應有所區分，例如終止治療、轉換治療或使用額外藥物等，即便能收集到後續的變數測量值，也可能得到不相關或難以詮釋的測量結果；而受試者死亡後的測量值則不存在。

議題三、本補遺所提出的框架將分析群體的觀念納入考量。「臨床試驗統計指導原則」第 5.2 章節強烈建議較優性試驗 (superiority trials) 的分析須基於全分析群體 (full analysis set)，即盡可能納入所有經隨機分派的受試者。然而，試驗常會收集同一受試者的重複測量值。排除認為無關緊要或難以詮釋的某些受試者的某些預定收集的測量值，這種做法與從全分析群體中完全排除受試者一樣，會產生類似的結果，即未能完全地維護初始的隨機分派。所造成的後果是有關療效的假說檢定、隨機分派所賦予的理論上的效益、以及平衡基礎點干擾因子實質上的效益，都將減少。此外，在某些情況下，結果變數 (outcome variable) 可能不存在有意義的數值，例如受試者死亡時。「臨床試驗統計指導原則」第 5.2 章節並未直接觸及上述問題，為使這些問題更加清楚，本補遺仔細地定義感興趣的療效，這種方法既確定了要納入療效估計的受試者群體，也確定在並發事件發生時，每位受試者應納入分析的觀察值。本補遺亦重新檢視依計畫書群體 (per protocol set) 分析的意義及角色；尤其當須要探討違反計畫書或試驗執行偏離的影響時，是否有比單純執行依計畫書群體分析，偏差更少且詮釋更好的解決方式。

最後，在敏感度分析的標題下，進一步討論穩健性（參見「臨床試驗統計指導原則」1.2. 章節）的概念。並對統計推論於所選擇分析方法的假設 (assumptions) 之敏感性，與更廣泛地選擇分析方法的敏感性之間作一個區分。通過精確地設定商定的估計目標及分析方法，該分析方法不僅與估計目標一致且預先設定，其詳細程度應足以由第三方精確地複製，這樣，法規單位對於一個特定分析方法審查的重點，可以專注於對假設偏離及資料局限性的敏感性評估。

於療效的估計或假說檢定，無論是否與療效或安全性相關，本補遺所概述的原則都是適用的。本補遺雖然主要側重於隨機分派臨床試驗，所有原則亦適用於單臂試驗 (single arm trials) 及觀察性研究。所提出的框架可應用於任何類型資料，包括縱向 (longitudinal) 研究資料、首發事件時間 (time-to-first event) 資料、以及復發事件 (recurrent event) 資

料。法規單位較有興趣的，則是將概述的原則應用於確認性試驗，或由多個試驗整合所產生的確認性結論。

A.2. 一個將規劃、設計、執行、分析與詮釋連結一致的框架

試驗的規劃應循序進行（圖一）。明確的試驗目標應藉由定義適當的估計目標轉化為主要臨床關注的問題。估計目標定義了特定試驗目標下的估計標的（即「估計什麼」，參見 A.3. 章節）。然後可以選擇一個適當的估計方法（即分析方法，稱為主要「估計法」，參見辭彙表）（參見 A.5.1. 章節）。主要估計法將以某些假設為基礎，為了探討偏離其基本假設時，主要估計法所得推論的穩健性，須執行一項或多項針對同一估計目標的敏感度分析（參見 A.5.2. 章節）。



圖一、給定的試驗目標將估計標的、估計方法、及敏感度分析連結一致

此框架將使得試驗可適當地規劃，從而清楚地區分估計的標的（試驗目標、估計目標）、估計的方法（估計法）、數值結果（「估計值」，參見辭彙表）、及敏感度分析。該框架亦有助於廠商的試驗規劃與法規單位的審查，而在討論臨床試驗設計的合適性及試驗結果的詮釋時，亦可加強雙方的互動。

適當的估計目標（參見 A.3. 章節）的設定，通常為試驗設計、執行（參見 A.4. 章節）及分析（參見 A.5. 章節）的主要決定因素。

A.3. 估計目標

藥物的開發及獲得許可的核心問題在於療效的確立及療效大小的估計：即同一群受試者接受此項治療相較於替代治療（假若他們沒有接受此項治療、或者他們接受不同的治療）的結果如何。估計目標是對於療效的一個精確的描述，以反映一項給定的臨床試驗目標所提出的臨床問題。它在群體層面（population level）上總結同一群受試者在接受不同治

療處置時，可能出現的結果。估計的標的應在臨床試驗之前確定。一旦定義好，就可以設計一個試驗，對於標的之療效提供可信賴的估計。

估計目標的描述包含精確地設定某些屬性 (attributes)，這不僅須基於臨床的考量，也須基於並發事件如何反映臨床關注的問題。補遺 A.3.1. 章節介紹並發事件；A.3.2. 章節介紹各種策略以描述與並發事件相關的臨床關注問題；A.3.3. 章節描述估計目標的屬性；A.3.4. 章節提出建立估計目標的相關考量。理解各策略之間的差異，並精確地闡明用於建立估計目標的策略是至關重要的。

A.3.1. 由臨床關注的問題來反映並發事件

並發事件是指治療開始後發生的事件，會影響臨床關注問題的相關測量，包含其詮釋或存在與否。在描述臨床關注的問題時，必須考量並發事件，俾能精確地定義所欲估計的療效。

由於變數的測量值可能會受到並發事件的影響，而且並發事件的發生可能取決於所接受的治療處置，因此在描述療效時須考量並發事件。例如，兩位病人最初可能接受相同的治療處置，且得到相同的結果測量值；但是假若其中一位病人接受額外藥物的治療，則兩個病人的測量值對於該治療所提供的訊息有所不同。此外，病人是否需要額外藥物治療，以及是否能繼續原先的治療，可能取決於他們所接受的治療處置。不同於缺失資料，並發事件不應被視為臨床試驗中應該避免的缺點。終止治療、使用額外藥物及其他類似的事件，在臨床實務或臨床試驗中都有可能發生，因此定義臨床關注的問題時，須明確地考慮到它們的發生。

影響測量值的詮釋之並發事件範例，包括終止被分派的治療及使用額外或替代治療。使用額外或替代治療可以有多種形式：包括改變背景 (background) 治療或併用 (concomitant) 治療，以及在感興趣的治療處置之間進行替換 (switching)。影響測量值存在的並發事件範例，包括當事件本身並不是變數的一部分時的終端事件，例如死亡或截肢（當評估糖尿病足潰瘍症狀時）。某些臨床事件也可能是並發事件，它們的發生或不發生定義了一個感興趣的主要層 (principal stratum) (參見 A.3.2. 章節)。其範例包括腫瘤試驗，欲評估療效反應時間，以發生腫瘤縮小所定義的目標療效反應 (objective response) 者為主要層；或評估疫苗對於感染嚴重度的療效，則以起初沒有感染，接種疫苗後發生感染者為主要層。

一個並發事件可以僅標識事件本身，例如終止治療，或者可以更詳細具體。例如可以說明事件發生的原因是藥物毒性或療效不足；或事件可能需要一定的大小或程度，例如持續使用超過規定時間或劑量的額外藥物；或者事件發生的時間，可能與變數評估時間之接近程度有關。某些事件會無限地影響結果測量值的詮釋，例如終止治療；而其他事件對於結果的詮釋則僅有暫時的影響，例如短期使用額外藥物。額外或替代治療的確可以為多樣的形式：或作為取代或輔助受試者無法獲得適當效益的治療；或

當受試者無法耐受被分派的治療時作為替代治療；或當作短期急性治療以控制疾病症狀的暫時發作。在臨床試驗中，額外或替代治療常被認為是，例如背景治療、救援藥物、禁用藥物等，以區別它們不同的角色，並允許分開來考量。假如使用不同的策略，須有額外詳細的描述，以標識為不同的並發事件。假若一個並發事件策略的選擇不僅取決於事件本身，例如未能繼續治療，也取決於事件發生的原因、幅度或時機，這些額外的資訊應該在臨床試驗中適當地定義與紀錄。就理論上來講，並發事件的描述可以反映出有關治療與追蹤非常具體的細節，例如慢性病的長期治療中錯過了一劑，或有一劑在某一天錯誤的時間服用，若此等特殊情況預期不會影響變數的詮釋，它們將不會被視為並發事件。

如上所述，在建立估計目標時，考慮並發事件是必要的。而由於估計目標於試驗設計前即已定義好，所以退出試驗或其他原因所產生的缺失資料（例如存活結果試驗之行政設限）本身都不是並發事件，惟退出試驗的受試者於退出試驗前可能已發生並發事件。

A.3.2. 定義臨床關注問題時用以解決並發事件的策略

以下將列出各個策略的描述，針對特定的並發事件，每一策略各自反映不同臨床關注的問題。無論是否習慣使用策略名稱，估計目標一旦建立，選擇的策略必須明確清楚。所有並發事件，無須都採用相同的策略來解決。事實上，對於不同的並發事件，通常會使用不同的策略來反映臨床關注的問題。補遺 A.3.4. 章節給予一些策略選擇的考量，以建立估計目標。

治療政策策略 (Treatment policy strategy)

並發事件的發生被認為與療效的定義無關：依此策略不管並發事件是否發生，變數的測量觀察值都會被採用。例如，當使用額外藥物為並發事件時，病人無論是否服用額外藥物，變數的測量觀察值都被採用。

如果應用於患者是否繼續接受治療、以及患者是否曾改變其他的治療（例如背景治療或併用治療）等並發事件，則此策略將這些並發事件視為組別間用於比較的一部分。在此情況下，正反映了「臨床試驗統計指導原則」辭彙表 ITT 原則中所描述的比較，即治療政策的療效。

一般而言，治療政策策略不能使用於並發事件為終端事件的情況，主要由於並發事件發生之後變數值並不存在。例如發生死亡後的變數無法測量，從而基於此策略的估計目標無法建立。

假想策略 (Hypothetical strategies)

假設並發事件沒有發生的情況：反映臨床關注問題的變數值，定義為假想情況下所得到的變數值。

雖然可能有許多不同的假想情況，但是某些情況相較於其他，可能對臨床或法規單位而言較具意義。例如考慮與試驗條件不同卻可執行的情況下的療效，可能具臨床或法規的重要性。特別是因為倫理的緣故，必須提供額外藥物，而感興趣的療效卻可能是假若沒有額外藥物可用時的結果。一個非常不同的假想情況，是可以假定並發事件沒有發生，或是假定不同的並發事件發生。例如一位受試者發生不良事件並終止治療，所考慮的假想情況，可以是假想該位受試者沒有發生不良事件，或是儘管發生不良事件卻繼續治療。臨床與法規單位對於上述假想情況的興趣是有限的，而且通常取決於，是否清楚瞭解該並發事件為何及如何發生，或者預期該並發事件所造成的結果在臨床實務中與臨床試驗中將有所不同。

如果採用假想策略，所設想的假想情況必須清楚地描述。例如，使用「若患者沒有服用額外藥物」之表述，可能造成混淆：究竟是「因為病人無法獲得該額外用藥」而未服用，還是「該病人不需要使用額外用藥」。

複合變數策略 (Composite variable strategies)

這與感興趣的變數（參見 A.3.3.章節）有關。由於並發事件本身對病人的治療結果提供有意義的訊息，因此被併入變數的定義中。例如病人因藥物毒性而終止治療，可被視為治療沒有成功。假若結果變數原本就是成功或失敗，則此並發事件將被視為另一種模式的治療失敗。複合變數策略無須侷限於二分法結果 (dichotomous outcomes)。例如一項試驗測量受試者的生理功能，可能建立一個連續尺度的結果變數，受試者發生死亡則被賦予一個反映缺乏功能的測量值。在某些案例中，複合變數策略可看作落實意圖治療原則的策略，在這些案例中，變數的原始測量值可能不存在或者不具意義，而並發事件如病人發生死亡，本身卻能有意義地描述病人的治療結果。

終端事件（例如死亡）也許就是最明顯須要採用複合變數策略的例子。假若一項治療能挽救生命，其對存活病人之各項指標的影響可能值得關注，但若感興趣的彙總測量 (summary measure) 僅為存活者某些數值測量的平均值則是不恰當的。感興趣的治療結果宜為存活與否結合數值性的測量值。例如腫瘤試驗無惡化存活時間 (progression-free survival)，所測量的療效即結合腫瘤生長情形與病人存活與否。

治療期間策略 (While on treatment strategies)

就此策略而言，並發事件發生前的治療效果才是感興趣的。此項策略所用的術語將取決於感興趣的並發事件，例如「活著的期間 (while alive)」即認定死亡為並發事件。

如果是一個變數被重複測量，在並發事件發生前的測量值才可能被視為與臨床問題有關；而不是所有受試者在相同的固定時間點的測量值。這同樣適用於二分法結果的發生，亦是觀察直到並發事件發生時為止。例如具有末期疾病的受試者，可能由於死亡而終止單純的症狀治療，然而，治療成效仍可根據死亡前症狀的改善來評估。或者，受試者可能會終止治療，在某些情況下，感興趣的將是評估病人在接受治療期間，發

生藥物不良反應的風險。

與複合變數策略相同，治療期間策略由於觀察時間限於並發事件發生前，也因此被認為會影響變數的定義。如果所比較的治療處置之間並發事件的發生率不相同，則需要特別注意（參見 A.3.3. 章節）。

主要層策略 (Principal stratum strategies)

這與感興趣的群體（參見 A.3.3. 章節）有關。目標群體中會發生並發事件的群體，可被認定為「主要層」（參見辭彙表）。或者，目標群體中不會發生並發事件的群體，亦可被認定為主要層。臨床關注之療效相關的問題僅於主要層內探討。例如欲瞭解疫苗對感染嚴重度的療效，通常期望在接種疫苗後發生感染的主要層內評估。或者，一種藥物的毒性可能阻止某些病人繼續試驗處置的治療，但仍然希望在可以耐受試驗處置毒性的病人群中，去瞭解治療的效果。

重要的是要能區別基於可能發生並發事件（例如受試者若分派到試驗處置將終止治療）之「主要分層 (principal stratification)」（參見辭彙表）與基於實際發生並發事件（受試者終止其被分派的治療）之次群體。於試驗處置中發生並發事件的次群體，不一定與對照處置中發生相同並發事件的次群體具有可比性。若以比較上述兩組次群體之治療結果來定義療效，宜考量治療結果的不同可能導因於受試者特徵不同所造成的干擾。

A.3.3. 估計目標的屬性

以下的屬性將用於建立估計目標，以定義感興趣的療效。

屬性之一為感興趣的**治療處置**以及與之進行比較之替代治療處置（補遺中其餘部分簡稱「治療」）。可以為單獨的介入措施，或同時使用的合併介入措施，例如附加於標準治療、或包含一系列複雜介入措施的整體治療方案（參見 A.3.2. 章節之治療政策策略及假想策略）。

屬性之二為臨床問題所針對的**病人群體**。將由整個試驗群體、或基礎點量測的一項特殊特徵所定義的次群體 (subgroup)、或由一個並發事件的發生（或不發生）所定義的主要層（參見 A.3.2. 章節下的主要層策略）來代表。

屬性之三為**變數**。由每位病人所獲得的變數（或評估指標）用來闡述臨床問題。變數的設定可以包含病人是否經歷並發事件（參見 A.3.2. 章節下的複合變數策略及治療期間策略）。

精確地設定治療、群體及變數三個屬性，可能就可以解決許多廠商和法規單位對所關注的臨床問題討論中所考量的並發事件。而針對**其他並發事件**相關的臨床問題，則常

以治療政策策略、假想策略或治療期間策略來反映，此為屬性之四。

最後一個屬性則為設定變數在**群體水平的總結** (population-level summary)，以提供治療處置間比較的基礎。

當定義一個感興趣的療效時，須確認該定義能標識出療效歸因於治療，而非歸因於其他如觀察時間或病人特徵的差異等潛在的干擾因子。

A.3.4. 建立估計目標的考量

臨床關注的問題及所對應的估計目標，應於臨床試驗計劃的初始階段設定。對於大多數的試驗，精確試驗目標的設定，須能反映終止治療以及使用額外或替代治療，在某些情況也須處理諸如死亡的終端事件。某些試驗目標只能參考臨床事件來描述，例如達到反應的受試者之反應持續時間 (duration of response)。

估計目標的建立，應考慮特定治療於特定疾病領域的臨床意義為何。考慮的因素包括研究的疾病、臨床的情況 (例如是否有替代療法可用)、治療如何施行 (例如一次性的給藥、短期治療或長期給藥)，以及治療的目的 (例如預防疾病、疾病的改善、症狀的控制)。另外，所得療效的估計是否為決策提供可靠的依據也很重要。例如不管在臨床結果發生之前使用過哪些治療，與假設沒有額外治療可用的情形下，關於療效所探究的臨床問題是不同的。端視不同的疾病治療領域而異，任一個都可能代表一個臨床關注的問題。然而，上述的兩種情況，估計療效的臨床試驗設計，通常須包括使用額外藥物 (若醫療需要) 的可能性，前者探究的臨床問題認為使用額外治療後的變數值仍有意義；後者探究的臨床問題則因為結果變數值亦反映了額外藥物的影響，認為使用額外治療後的變數值與療效沒有直接的相關。在估計目標最終決定前，應達成可靠療效估計是可行的共識。對於上述後者的臨床問題，無法用於分析的變數值的插補方法亦須有共識。

在建立估計目標時，須清楚瞭解臨床關注問題上所涉及的治療(參見 A.3.3. 章節)。清楚地設定感興趣的治療可能已經反映了多個相關的並發事件。具體而言，所設定的治療屬性可能已經反映有關改變背景治療、併用治療、使用額外或後線(later-line) 治療、轉換治療以及調理療法 (conditioning regimen) 等臨床關注的問題。舉例來說，若治療屬性設定為介入治療 A 附加於背景治療 B (依需要的劑量)。在這個案例中，改變背景治療 B 的劑量無須視為並發事件；然而，使用額外治療則須視為並發事件。假若使用任何額外藥物也以治療政策策略來反映，則治療屬性可以設定為介入治療 A 附加於背景治療 B (依需要的劑量)，以及額外治療 (依需要附加)。相反地，若治療屬性設定為介入治療 A，則改變背景治療及使用額外治療兩者皆應以並發事件處理。

在建立估計目標的討論中，還須考慮是否應使用**群體及變數**屬性來反映與任何並發事件相關的臨床關注問題，然後可以考慮對其他的並發事件採取的策略。這通常需要一

個迭代的過程，以達到一個對決策具有臨床意義的估計目標，並且可以對此作出可靠的估計。某些估計目標（尤其是測量值與臨床問題相關的估計目標），往往可在較少的假設下即得到穩健的估計。而其他估計目標之分析方法，可能需要較多難以判定其合理性之特定假設，且對於這些假設中可能出現的合理的變化更為敏感（參見 A.5.1. 章節）。如果對一個特定的估計目標，不存在適當的試驗設計或難以獲得可信賴的估計值時，須考慮採用其他的估計目標、試驗設計及分析方法。

迴避或過於簡化討論與建構主要估計目標的過程，會導致試驗目標、試驗設計、資料收集、與分析方法彼此之間的不一致。儘管可能因為無法得到一個可信賴的估計值而排除某些策略的選擇，但重要的是從試驗目標和對臨床關注問題的瞭解開始，而不是從資料收集和分析方法的選擇來決定估計目標。

此外，還須考慮試驗的情況。如果臨床試驗計畫書中對於受試者的處理（例如針對不耐受的劑量調整、療效反應不佳的救援治療、臨床試驗評估的負擔）被證明與臨床實務預期的情況不同，則須於建立估計目標時反映。

估計目標一旦建立，應該能清楚明確地定義估計的標的。試想一個終止治療的並發事件，最重要的是要能區別感興趣的療效是於主要層（若使用試驗處置能持續治療）病人的療效，或是於持續治療期間的療效。更重要的是，此兩種策略皆不應該用來代表假設所有病人都能持續治療的療效。

誠如上述，當使用假想策略時，某些假想情況比起其他情況，法規單位更能接受用於決策。因此所描述的假想情況須能合理地量化成一個可詮釋的療效，提供法規單位決策及臨床實務用藥時有意義的資訊。例如，假想在沒有救援藥物可用的情形下，感興趣的變數是要收集什麼樣的觀察值，就會變得重要。相比之下，受試者由於藥物不良反應，理論上應終止治療，但假想受試者實際上並沒有停止而是持續治療，此時得到的變數值，無論是從臨床或是法規的角度可能都會被認為是不合理的。臨床上感興趣的療效若是基於所有受試者都能持續地接受治療，而認定會持續治療的假想情況卻未經過透徹地討論，如此定義的療效並不明確。更何況無法耐受治療，本身可能就是無法達到有利結果的證據。

使用治療政策策略所定義的估計目標來描述療效，可能會更普遍地被接受用於支持法規的決策；特別是當其他策略所定義的估計目標可能較有臨床意義，惟其主要估計法與其敏感度估計法，卻不確定可提供一個可靠的估計值或穩健的推論。而以治療政策策略所定義的估計目標，或可提供對仍具臨床意義之治療效果的可靠估計。在這種情況下，建議試驗亦包括那些被認為具有更大臨床意義的估計目標，並呈現所得到的估計值，同時就其在試驗設計或統計分析方面的侷限性進行討論。當採用治療政策策略建立估計目標時，針對該策略所處理的並發事件，可定義一個額外的估計目標及分析作為推論的補充；例如，以症狀分數測量的療效與各治療處置間使用額外藥物的受試者比例做一個對比。同樣地，使用治療期間策略定義的估計目標，通常須伴隨自起始

至發生並發事件的時間 (time to intercurrent event) 分布之額外資訊；及主要層策略定義的估計目標伴隨以受試者中主要層病人所占百分比的資訊 (如果有) 將很有用。

用以支持法規單位以不劣性 (non-inferiority) 或相等性 (equivalence) 為目標的決策，建立估計目標時所考慮的因素，可能不同於以較優性為目標的決策。誠如「臨床試驗統計指導原則」所述，法規單位決策時根據不劣性或相等性試驗，相較於根據較優性試驗所面臨的問題不同。指引第 3.3.2. 章節指出，不劣性或相等性試驗在本質上並不保守，並說明將此類試驗之違反與偏離計畫書、不遵循計畫書以及退出試驗的受試者數目減至最低的重要性。在「臨床試驗統計指導原則」第 5.2.1. 章節中，描述了這類試驗依全分析群體分析的結果一般較不保守，其所扮演的角色須要非常慎重地考慮。在建立估計目標時，若以治療政策策略來處理一個或多個的並發事件，則與不劣性或相等性試驗以全分析群體依 ITT 原則分析所呈現的問題是類似的。兩個治療處置的受試者在終止隨機分派的治療或使用其他藥物治療後，可能會呈現較為相似的療效，而這與最初隨機分派時的兩治療處置療效的相似性無關。估計目標的建立，是可以直接去解決那些可能導致治療處置間差異變小的並發事件 (例如終止治療及使用額外藥物)。在策略選擇的過程中，應明確區分兩類試驗：其一為旨在評估含有相同或相似活性成分的治療之間是否存在差異的試驗 (例如：比較生物相似藥與對照處置)；其二則為採用不劣性或相等性假說，以建立並量化療效證據的試驗。為適用於法規單位的決策，估計目標的建立宜以能敏感偵測到治療處置間差異的療效為優先。

A.4. 對試驗設計與執行的影響

試驗設計須與能反映試驗目標的估計目標一致。一個適用於某個估計目標的試驗設計，可能不適用於其他具潛在重要性的估計目標。而對於療效進行量化所根據的估計目標若有清楚地定義，將對試驗設計相關的選擇提供了訊息。這包括設定納入及排除條件以確定標的群體，治療處置包含計畫書允許與禁用的藥物，以及其他有關病人的管理與資料的收集等方面。舉例來說，假若感興趣的是去瞭解不管是否發生某一特定並發事件的療效，則試驗收集所有受試者的變數資料是恰當的；或是，假若法規單位支持其決策所需的估計目標，無須收集一個並發事件發生後的變數資料，則為了其他的估計目標而收集此等資料所得的效益，須與收集導致的複雜性與潛在的缺點權衡。

應該盡力收集與估計有關的所有資料，包括並發事件的表徵，以及是否發生並發事件及發生時間點等資訊。資料不可能總是收集得到，當然也不能違背受試者意願將他們留在試驗中，在一些試驗中，某些受試者的資料缺失是不可避免的，例如結果為存活的試驗之行政設限。相反地，發生如終止治療、轉換治療、或使用額外藥物等並發事件，並非意味著之後的變數無法測量，只是這些測量值可能沒有相關。而終端事件諸如死亡，在並發事件發生之後的變數無法測量，惟此等資料一般不能視為缺失資料。

若未能收齊評估一個估計目標所需的所有資料，將導致後續的統計推論面臨缺失資料的問題。取決於缺失資料的比例，統計分析的有效性，可能會受到無法驗證假設的影響，

從而可能會破壞試驗結果的穩健性 (參見 A.5. 章節)。若能以一個前瞻性的計畫，收集資料缺失的具體理由，將有助於區分發生並發事件或是缺失資料。這樣可以改善分析，也會對敏感度分析的選擇有幫助。例如，「失去追蹤 (lost to follow up)」可以較準確地紀錄為「因缺乏療效而終止治療」。當它被定義為一個並發事件，就能以選擇用來處理並發事件的策略來反映，而不會認為是缺失資料的問題。為了減少缺失資料，可以執行某些措施提升受試者的留存率。然而，利用措施去減少或避免於臨床實務中正常會發生的並發事件，可能會降低試驗的外在效度 (external validity)。例如，選擇試驗的群體，使用劑量調整方案或併用藥物來減輕藥物毒性的影響，假若相同的措施在臨床實務中不會被採用，這些措施可能不合適。

隨機分派和盲性作業仍然是具對照組臨床試驗的基石。「臨床試驗統計指導原則」第 2.3. 章節提及了避免偏差的設計技巧。某些估計目標可能需要、或因採用某些試驗設計而受益，例如導入期 (run-in) 或強化 (enrichment) 設計、隨機退出設計 (randomized withdrawal design)、或劑量調整設計 (titration design) 等。在將受試者隨機分派至試驗處置或對照處置之前，可使用導入期標識出耐受度高之受試者主要層。針對此項設計，於法規單位與廠商的對話中，須要考慮所擬的導入期是否適合標識出標的群體，以及後續試驗設計的選擇 (例如廓清期、隨機分派等) 是否能支持標的療效的估計和相關的推論。這些考量可能限制此等試驗設計與特定策略的使用。

精確地描述感興趣的療效，可對樣本數的估算提供了訊息。惟當引用歷史試驗的數據時，若該等試驗報告隱含或明確地指出，所估計的療效或變異數是根據不同的估計目標，則須特別注意。如果所有受試者都對分析提供了訊息，且用於反映並發事件之策略的影響也納入標的療效大小 (effect size) 與預期變異數的考量時，通常就無須於所估算的樣本數之外，按預期退出試驗的受試者比例進行擴增。

「臨床試驗統計指導原則」第 7.2. 章節論及彙總多個試驗資料的議題，強調納入的試驗，針對感興趣的變數，須有一致的定義，此可延伸至估計目標的建立。因此，若試驗在計畫的階段，即設想須整合多項臨床試驗計畫中的療效證據時，於試驗計畫書中應建構一個合適的估計目標，且該估計目標須能反映所有貢獻試驗之試驗設計的選擇。類似的考量也適用於統合分析 (meta-analysis) 的設計、或使用已完成的試驗所估計的療效大小來決定不劣性的臨界值 (non-inferiority margin)、或使用外部對照組來詮釋單臂試驗的結果。單純地比較不同來源的資料，或整合來自多個試驗的資料，而未考量或規範每個資料呈現或統計分析中所描述的估計目標，可能造成誤導。

更廣泛來說，一項試驗可能有多個試驗目標而轉化成多個估計目標，且各有其對應的統計檢定與估計，所產生的多重檢定的問題亦須要處理。

A.5. 對試驗分析的影響

A.5.1. 主要估計

估計目標探討試驗處置相對於對照處置的治療效果，可由一組接受試驗處置的受試者，另一組接受對照處置的相似受試者，由其結果的比較來估計。一個給定的估計目標，須執行與其相應的分析方法或估計法，以便得到估計值，並能給出可靠的詮釋。此分析方法亦將用於信賴區間的計算及統計顯著性的檢定。而能否得到一個可詮釋的估計值，其重要的考量則是分析中須要做出何種程度的假設。主要的假設應與估計目標連同其附隨的主要估計法及敏感度估計法一起有明確的說明。而假設必須是合理的，避免不切實際的假設。須透過與估計目標一致的敏感度分析，評估在潛在偏離基本假設下所得結果的穩健性（參見 A.5.2. 章節）。而仰賴許多的假設或很強的假設的估計，則需要更全面的敏感度分析。當偏離假設的影響不能經由敏感度分析來全面性地探討時，該特定的估計目標與分析方法的組合，可能無法被接受用於決策。

所有分析方法皆仰賴假設，即便對應同一估計目標，不同的方法亦可能仰賴不同的假設。然而，有一些假設是固有的，存在於所有分析方法中，此等分析方法與使用不同策略的估計目標一致。例如，在假想的狀況下，預測會觀察到結果的方法；或主要層策略中，標識出合適標的群體的方法。下面舉出一些案例，這些案例與用來反映並發事件發生的不同策略相關。所強調的議題，都是廠商與法規單位在估計目標、主要分析及敏感度分析尚未取得共識前，須討論的關鍵部分。

符合治療政策策略以處理特定並發事件的分析，其所依賴的假設強度，將取決於試驗的設計與執行方式。當大部分的受試者即使在各自的並發事件（例如終止治療）之後都能被追蹤觀察到，剩下的缺失資料的問題可能較小；反之，當並發事件之後即終止觀察，採用此項策略顯然並不恰當，因為要假設終止治療受試者（未觀察到）的結果，與繼續治療受試者（觀察到）的結果是相似的，通常是不合理的。若以其他方法處理缺失資料則須要證明其合理性，並預期要執行敏感度分析。

符合假想策略的分析，納入分析與實際觀察到的結果不同；例如，在實際已給予救援藥物的情況下，分析若未給予救援藥物的可能結果。在使用救援藥物之前的觀察值，以及不需要救援藥物的受試者的觀察值可能是有用的，但只有在很強的假設下。

複合變數策略將並發事件的發生視為結果的一部分，可以避免針對並發事件後的資料，所需做的統計假設。潛在的考量較少在於為了估計所做的假設，較多在於所估計療效的詮釋。為了使估計目標可以詮釋，因並發事件的發生而被給予代表治療失敗的分數，須有意義地反映出病人缺少效益的情形（例如反映死亡與反映因不良反應而終止治療可能不同）。

基於治療期間策略所建立的估計目標，只要是在並發事件發生前收集的結果，都能被

納入分析。同樣地，其關鍵的假設在於詮釋的考量。以終止治療為例，受試者於治療期間的病情結果可能會得到改善，惟治療亦可能激發終止治療使其發生的時間縮短，或延遲終止治療使其發生的時間增長，而上述兩種影響都應於臨床效益詮釋或評估時列為考量。

符合主要層策略的分析通常須要很強的假設。例如，某些推斷主要層的方法來自於受試者基礎點的特性，但卻可能難以評估此項推斷的正確性。然而，使用簡化的方法無法避免此困難性，例如，單純地比較試驗處置未發生並發事件的受試者，與對照處置未發生並發事件的受試者，假設並發事件與治療無關，就很難證明。

在定義估計目標時，即使已經以適當的方式解決並發事件，並努力去收集估計所需的資料（參見 A.4. 章節），一些資料仍然會有缺失，例如存活結果試驗之行政設限。未能收集相關的資料不應與「選擇不收集」或「收集了卻不用」（由於並發事件而變成不相關的資料）相混淆。例如，基於治療政策策略的估計目標，於終止試驗用藥後仍試圖收集資料，如果未能收集到，則為缺失資料；然而，同樣的資料對於另一種策略可能不相關，因此，就使用此策略的另一個估計目標，如果未能收集，也不算是缺失資料。當努力去收集資料卻未能成功，就使得統計分析時為了處理缺失資料而必須做出假設。缺失資料的處理應基於臨床合理的假設，並在可能的情況下，以估計目標描述中所採用的策略為指導。採用的方法，可以根據受試者本身和其他相似的受試者所觀察到的共變量 (covariate) 以及基礎點後的資料。標識相似受試者的標準可以包括是否發生並發事件。例如，對於終止治療未能進一步收集資料的受試者，採行的模型可使用來自其他終止治療而持續收集資料之受試者的資料。

A.5.2. 敏感度分析

A.5.2.1. 敏感度分析的角色

在資料有侷限性，以及與主要估計法的統計模型偏離其假設時，基於某估計目標的推論應當是穩健的。這種穩健性常用敏感度分析來評估。所有對法規決策及藥品訊息的仿單是重要的估計目標之主要估計法，都必須規劃敏感度分析，此將成為廠商與法規單位討論及達到共識的一個主題。

用於支持主要估計法的統計假設都須要記錄。針對同一個估計目標，須預先設定一個或多個分析以對假設進行探討，目的在於驗證當主要估計法偏離假設時，其所得的估計值是否穩健。此類分析（例如：臨界點分析 [tipping point analysis]）可藉由假設偏離的程度，以及該偏離對臨床或統計顯著性結果詮釋之影響來加以表徵。

有別於研究者意圖探索偏離假設的穩健性所執行的敏感度分析，為了更全面地探討及瞭解試驗資料，所執行的其他分析可稱為「補充分析 (supplementary analysis)」(參見辭彙表，A.5.3. 章節)。當主要估計目標是經廠商與法規單位同意的，主要估計

法也已預先明確設定，且敏感度分析證明了所得估計值的詮釋是可信的，補充分析在評估中通常應放在較低的優先順序。

A.5.2.2. 敏感度分析的選擇

在規劃及執行敏感度分析時，同時改變主要分析的多個條件，將難以確定那些假設(如果有的話)，與所看到的任何潛在的差異有關。因此，最好採取一種結構化的方法 (structured approach)，指定在替代分析下所改變的假設，而不是單純地比較基於不同假設集 (sets of assumptions) 之不同分析的結果。至於同時改變多個假設的必要性，則須逐案考慮。當評估不同分析結果的詮釋與相關性時，區分可檢定的和不可檢定的假設條件會有幫助的。

敏感度分析對於缺失資料的必要性已被認定，本補遺的框架亦保留了它的重要性。針對一個特定的估計目標，必須定義及考量缺失資料 (參見 A.4. 章節)。針對特定估計目標，須區分「資料缺失」及「資料與該估計無直接相關」兩種情況，因此在進行敏感度分析時，須考慮不同假設集。

A.5.3. 補充分析

如果敏感度分析已證明對應的估計值是穩健的，試驗結果的詮釋應專注於每個商定的估計目標之主要估計法。針對一項估計目標，可以在主要分析及敏感度分析之外執行補充分析，以對療效的理解提供額外的見解。它們通常在詮釋試驗結果時扮演次要的角色。而對每一個試驗都應考量補充分析的必要性及其作用。

「臨床試驗統計指導原則」第 5.2.3. 章節指出，同時規劃 FAS 與 PPS 兩項分析通常是恰當的，因為兩者之間的差異，可做為明確討論與詮釋的主題。如果 FAS 與 PPS 分析獲得一致的結果，表明對試驗結果更有信心。該指引第 5.2.2. 章節亦提及，依據 PPS 的結果可能會產生嚴重的偏差。而就本補遺中所提出的框架，可能無法建立一個與 PPS 分析一致的有意義的估計目標。如上所述，PPS 分析無法達到任何主要層 (例如，能耐受並持續接受試驗處置的受試者) 所意圖估計的療效，因為不同處置間的受試者可能不盡相似。

違反及偏離計畫書，例如在限定的時間窗 (time window) 之外回診，可能使得受試者自 PPS 中排除，而不必發生並發事件。同樣地，受試者可能發生諸如死亡等的並發事件，卻沒有偏離計畫書。儘管違反及偏離計畫書與並發事件不同，但在估計目標的描述中，已考慮可能會影響測量值詮釋與存在的事件。因此，建立估計目標及其對應的分析方法，可以更好地解決通常與 PPS 分析相關的目的。由此，PPS 的分析可能不會增加額外的見解。

A.6. 紀錄估計目標和敏感度分析

試驗計畫書中應定義並明確地設定與試驗主要目標相對應的主要估計目標。計畫書及分析計畫書應預先設定與主要估計目標一致的主要估計法和主要分析，以及適當的敏感度分析，以探討主要分析在偏離假設時的穩健性。針對次要試驗目標（例如與次要變數相關）的估計目標，若可能用於支持法規單位的決策，亦應明確地定義與設定，且每個估計目標都應有對應的主要估計法和適當的敏感度分析。其他的探索性試驗目標可以探索為目的，也可定義其他的估計目標。

主要估計目標的選取通常是試驗設計、執行與分析等方面的主要決定因素。依照慣例，這些應於試驗計畫書中詳細地記載。假若次要估計目標亦為主要的關注點，相關考量亦應視需要而拓展以支持這些次要估計目標，這些也須詳細地記載於計畫書中。除此之外，有關試驗設計、執行與分析的常規考量仍維持不變。

儘管清楚所要估計的內容對廠商是有利的，但法規單位並不要求每一個針對探索性目標的估計目標，皆要記錄於相關文件中。主要分析、敏感度分析及補充分析的結果，應於臨床試驗報告中系統性地彙報，並具體說明各項分析，是在盲性狀態下預先設定的，還是事後 (post hoc) 才執行的。每個治療處置之每一項並發事件的發生人數、發生時間點之彙總，亦應記錄於報告中。

於試驗期間修改估計目標可能會有問題，且可能降低試驗的可信度。於試驗設計階段未能預見，而於試驗執行期間才發現的並發事件之處理，不僅應就分析的選擇，且應對估計目標的影響（亦即對於所估計療效的描述及試驗結果的詮釋）皆須有所討論。而估計目標的修改通常須透過計畫書變更來反映。

辭彙表

估計目標 (Estimand)：

對於療效的一個精確的描述，以反映一項給定的臨床試驗目標所提出的臨床問題。它在群體層面 (population level) 上總結了，若同一群受試者在接受不同治療處置時，可能出現的結果。

估計值 (Estimate)：

由估計法計算出的數值。

估計法 (Estimator)：

利用臨床試驗資料計算出估計值的分析方法。

並發事件 (Intercurrent Events)：

在治療開始後發生的事件，會影響臨床關注問題的相關測量，包含其詮釋或存在與否。在描述臨床關注的問題時，必須考量並發事件，俾能精確地定義所欲估計的療效。

缺失資料 (Missing Data)：

對給定的估計目標的分析有意義，但未收集到的資料。必須把缺失資料與不存在的資料，或因並發事件而不被視為有意義的資料有所區別。

主要分層 (Principal Stratification)：

於所有治療處置，根據並發事件潛在發生的可能，對受試者進行分類。例如兩種治療，一個並發事件，則共有四個主要層：接受任一治療都不會發生並發事件的受試者；接受治療 A 會發生，但治療 B 不會發生並發事件的受試者；接受治療 B 會發生，但治療 A 不會發生並發事件的受試者；以及兩種治療都將會發生並發事件的受試者。本補遺中，主要層是指主要分層所定義的任何一層（或合併數層）。

敏感度分析 (Sensitivity Analysis)：

主要估計法偏離其模型的基本假設與資料有侷限性時，意圖探討其推論的穩健性所進行的一系列分析。

補充分析 (Supplementary Analysis)：

主要分析或敏感度分析之外的其他分析，意圖對治療效果的理解提供額外的見解。