



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



TFDA 藥物  
食品安全週報



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

## 淺談美國 FDA 不劣性試驗指引最終版之修訂

吳雅琪<sup>1</sup>

### 前言

美國食品和藥物管理局(US FDA)於 2010 年完成「以不劣性試驗確立藥品有效性指引草案(Draft Guidance for Industry: Non-inferiority Clinical Trials)」，之後歷經六年聽取各方意見並予修訂，於 2016 年 11 月頒布最終版本<sup>[1]</sup>。本指引同時取代 2010 年所發布之「使用不劣性試驗支持抗菌藥品上市許可指引」<sup>[2]</sup>。

本指引所探討的不劣性(non-inferiority, NI)試驗，主要針對危及生命或具不可逆傷害的疾病，當市面上已有經核准、效益良好的藥品可供使用，且不治療、或以安慰劑或非常低劑量的活性藥品治療並不符合倫理時，所採用的試驗設計。至於其他非因倫理因素之三組別(試驗組、活性對照組與安慰劑組)試驗設計，則非本指引考量的範圍。

不劣性試驗設計最大的問題在於沒有安慰劑作為對照組，以致影響試驗結果的解讀。一般而言，設計嚴謹、執行良好、以安慰劑為對照之較優性(superiority)試驗，試驗藥品若能成功地顯示與安慰劑具有顯著的組間差異，其療效是可以直接確認的。反之不劣性試驗的結果要能解讀，活性對照藥必須具有檢測靈敏度(assay sensitivity)，否則即使試驗組與活性對照組間達到可接受的小差異(差異小於不劣性臨界值)，仍無法支持試驗藥品的療效。檢測靈敏度定義為活性對照藥品在 NI 試驗中達到預期之效果，亦即與安慰劑相比，達到特定大小的組間差異。另外，NI 試驗亦須符合一致性的假設(constancy assumption)，亦即活性對照藥於 NI 試驗中的療效，與其之前試驗(安慰劑對照)之療效大小相同。惟如上所述，NI 試驗通常不包括安慰劑組，因此無法驗證檢測靈敏度與一致性的假設，只能靠事先審慎保守選取不劣性臨界值(non-inferiority margin)，以及嚴格要求試驗的品質，方可認定試驗的結果。

2010 年 NI 試驗指引草案出爐後，引起產、官、學界熱烈地迴響，網路期刊上亦有不少精闢深入的論述。因此本文並不詳細描述指引整體內容，僅就最終版修訂的部分加以介紹。

<sup>1</sup> 財團法人醫藥品查驗中心新藥科技組



致力法規科學  
守護生命健康  
Regulatory Science, Service for Life

## 最終版修訂的部分

相較於指引草案，最終版修訂的部分主要有以下幾點：1)標題由「Non-inferiority Clinical Trials」改為「Non-Inferiority Clinical Trials to Establish Effectiveness」，以強調本指引主要是探討藥品療效部分，並未提供不劣性設計下藥品安全性評估之建議；2)最終版使用較精簡的文字，頁數由原先 66 頁減為 56 頁。內容亦多採條列式，以突顯指引的重點。另外，圖表部分亦增添說明與建議；3)說明選取 M2 為不劣性臨界值之合理性；4)詳細說明兩個 NI 試驗常用的統計分析方法；及 5)原草案案例附錄部分，第 2 個案例(最終版第 4 例)更改為 CABP 指引草案<sup>[4]</sup>之附錄案例。

### 一、增添之圖表說明與建議

圖 1 顯示較優性試驗三種可能的分析結果。第一種結果為試驗組顯著優於安慰劑組，因此可以宣稱療效；第二與第三種結果皆無法宣稱試驗藥品的療效。惟第二種結果試驗組與安慰劑組差異的點估計值為 2，與第一種結果相同，只是前者信賴區間較寬，跨越 0 之虛無假說(null hypothesis)。針對此種試驗結果，最終版建議可鼓勵廠商再執行一個較大型、受試者人數較多的試驗，也許還有成功的機會。

Figure 1. Possible Results of a Placebo-Controlled Superiority Study (Point Estimate and 95% Confidence Interval (CI))

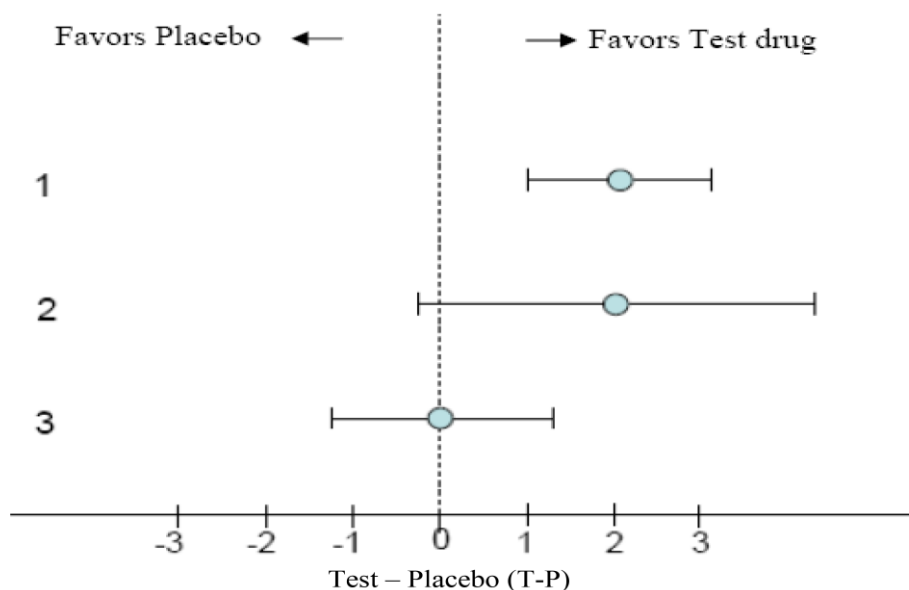


圖 1、安慰劑對照之較優性試驗可能結果<sup>註 1</sup>

<sup>註 1</sup> FDA Non-inferiority Clinical Trials to Establish Effectiveness 2016, p4



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



TFDA 藥物  
食品安全週報



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

## 二、選取 M2 為不劣性臨界值之合理性

不劣性試驗臨界值的選取，有 M1 與 M2 兩種考量，茲分別定義如下：

- M1：活性對照藥品於現今的 NI 試驗中預期會有的療效；
- M2：臨床上可接受的試驗藥品與活性對照藥品療效最大的差異。

M2 臨界值應小於 M1 臨界值。由於 NI 試驗無安慰劑組，僅能從先前活性對照藥品之安慰劑對照試驗中估算其在現今(current)試驗中預期的療效- M1。然而，活性對照藥品在現今 NI 試驗要有 M1 的療效，必須具有檢測靈敏度，亦即現今 NI 試驗之重要設計特性(例如：受試者納入排除條件、病人群體特徵、重要可使用之併用藥物、主要評估指標、對照組劑量與統計分析方法選取等)必須與先前的試驗類似。另外亦須符合一致性的假設，亦即活性對照藥品於現今 NI 試驗中的療效表現須與其之前一貫的療效表現相同。

執行 NI 試驗最關鍵性的考量是試驗品質要好，以排除因為執行上的不嚴謹，造成兩組療效差異變小，因此達到試驗藥品顯著不劣於對照藥品的目標(亦即品質差的試驗，試驗結果對照藥品療效差到像安慰劑一樣，即便試驗藥品不具療效，仍可能達到顯著不劣於對照藥品而宣稱療效)。值得注意的是，品質不良的試驗是無法事後以任何統計方法來調整或補救的。

另外，本指引亦提供一個淺顯易懂的例子說明檢測靈敏度的重要性。例如一項 NI 試驗，其活性對照藥品的 M1 估算為 10，若選取 M1 為臨界值，而在現今 NI 試驗中活性對照藥品的療效卻小於 10，即使試驗組與對照組治療效果的差異小於 10，雖達到顯著不劣性的目標，其結論仍是錯誤的，無法支持試驗藥品的療效。

統計分析常用於推翻「試驗藥品劣於對照藥品」虛無假說的條件為計算(對照組-試驗組)療效差異的雙尾 95%信賴區間，且其上限(upper bound)小於臨界值。NI 試驗若選取 M1 為臨界值，且具有檢測靈敏度，若能達到上述推翻虛無假說的條件，僅能說試驗藥品有較大的可能療效顯著大於 0 (顯著優於安慰劑)；但在某些重大疾病治療，多不能接受試驗藥品之療效差於活性對照藥品太多，因此，就臨床的觀點，為保留活性對照藥品相當的療效，建議選擇較小的 M2 為臨界值。

有鑑於對 M2 的要求並不包含在美國聯邦食品、藥品和化妝品法案中(Federal Food, Drug and Cosmetic Act)，1995 年 4 月美國總統柯林頓和副總統高爾於 Reinventing Regulation of Drugs and Medical Devices, part of the National Performance Review 中有以下的聲明：

『在某些情況，考量新產品是否比現有的療法效果為差可能是重要的，尤其當較差



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



TFDA 藥物  
食品安全週報



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

的療效可能造成患者或公眾的危險時。例如，當以下情況，新療法與已核准上市藥品療效相當，對於公共健康的保護是非常重要的：

1. 治療的疾病危及生命或能引起不可逆的疾病(例如中風或心臟病發作)；
2. 治療的疾病是一種可能對他人的健康造成嚴重後果的傳染性疾病(例如性病)。』

此項聲明已納入 1995 年 8 月 1 日的 Federal Register，以強化法律的效力。

### 三、 NI 試驗常用的統計分析方法

目前有兩種常用的統計方法來分析 NI 試驗的結果。一種為「臨界值固定法(fixed margin method)」或稱為 95%-95%法。另一種則為「合成法(synthesis approach)」，兩種方法都將先前活性對照藥品的試驗與現今 NI 試驗納入考量，只是以不同的方式來處理。

#### (一) 臨界值固定法(95%-95%法)

臨界值固定法亦稱 95%-95%法，第一個 95%係將先前活性對照藥品試驗的結果的摘要統計值(summary statistics)，經過隨機療效整合分析(random-effect meta-analysis)，以活性對照藥與安慰劑差異之 95%信賴區間(雙尾)下限定義為 M1 (附註: 若先前的試驗多於一個，可採用 95%信賴區間下限；若先前的試驗只有一個，由於未能客觀地考量試驗之間的變異，建議採用較保守的 99%信賴區間)，第二個 95%是指現今 NI 試驗之試驗藥品與對照藥品療效差異之 95%信賴區間。

第一個 95%取活性對照藥於先前試驗的平均療效(average effect)之 95%信賴區間下限。2010 年草案認為這是活性對照藥於 NI 試驗的實際療效(actual effect)之下限，但指引最終版則認為 NI 試驗中對照藥品的實際療效仍易受到取樣變異的影響，因此建議採用預測區間(prediction intervals)來估計活性對照藥實際療效的範圍，這範圍會比信賴區間的範圍來得寬。然而卻也認為信賴區間對於 NI 試驗的目的是足夠的，因為由對照藥品先前試驗之平均療效的信賴區間可以得到對照藥品相較於安慰劑的真實療效(true effect)範圍，而由 NI 試驗的信賴區間則可得到試驗藥品相較於對照藥品的真實療效範圍。只要一致性的假設成立，兩項結果可以合併後做出試驗藥品相較於安慰劑是否具有正面療效的推論。

#### (二) 合成法

合成法則是合併先前試驗的數據與現今 NI 試驗的數據，亦即將兩個數據來源的變異皆納入考量。這種處理方式，好像兩個數據皆來自同一個隨機分派的臨床試驗，也就是假設安慰劑在 NI 試驗中是存在的情況。

台灣藥物法規  
資訊網法規公告台灣藥品  
臨床試驗資訊TFDA 藥物  
食品安全週報致力法規科學  
守護生命健康

Regulatory Science, Service for Life

儘管 95%-95%法與合成法不同，但經過數學的推導，其統計檢定的公式還是很類似的，稍作說明如下：

假若先前試驗對照藥品( $C_h$ )相較於安慰劑(P)之療效估計為  $C_h-P$ ，現今 NI 試驗的試驗藥品(T)相較於對照藥品( $C_n$ )之療效估計為  $T - C_n$ ；當一致性的假設成立時，即  $C_h = C_n$ ，合併兩個估計量可以得到： $T - C_n + C_h - P = T - P$ ，即為試驗藥品與安慰劑之差異，亦為兩種方法統計檢定公式的分子部分。因此兩種方法分子部分是相同的，差異僅在統計檢定之分母(標準誤)部分，分別為：

➤ 95%-95%法之標準誤： $SE_H + SE_N$ ，

➤ 合成法之標準誤： $\sqrt{SE_H^2 + SE_N^2}$

其中  $SE_H$  與  $SE_N$  分別為先前試驗對照藥品相較於安慰劑療效差異( $C_h-P$ )之標準誤，及現今 NI 試驗試驗藥品相較於對照藥品療效差異( $T - C_n$ )之標準誤。由於 95%-95%法之標準誤總是比合成法之標準誤為大(分別平方後即可看出)，因此與合成法相比，95%-95%方法是較保守的。雖然兩種方法都要求符合一致性的假設，由於臨界值固定法的標準誤較大，較能容忍輕微偏離一致性的假設，亦即合成法偏離一致性的假設較易導致錯誤的結論。反之，由於臨界值固定法的步驟中須要決定  $M1$  與  $M2$ ，指引亦建議可事先與法規單位討論其合理性。

#### 四、不劣性臨界值之選取 - CABP 案例

本案例取材自 2014 年頒布之 CABP 草案，針對治療細菌引起之社區性肺炎之 NI 試驗，當沒有良好的歷史資料時，建議如何推估不劣性臨界值。由於本案例中，歷史性資料僅限於觀察性研究(observational study)，且所有資料皆已年代久遠 (1960 年代或更早)，缺乏隨機分派、具安慰劑組對照之臨床試驗療效資料，不可能以指引所建議的方法 (95%-95%方法)估計  $M1$ 。

但由表一的結果可以發現，CABP 如果不治療，死亡率顯著高於抗菌藥品治療組。未治療與抗菌藥品治療死亡率的差異，在不同的研究其 95%信賴區間的下限自 15%至 41%，因此本指引認為 NI 試驗臨界值訂為 10%，為適當且臨床可接受的臨界值( $M2$ )。



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

表一、觀察性研究中肺炎球菌肺炎之死亡率

Publication	Population	Mortality (%): Untreated Patients (Study Years)	Mortality (%): Antibacterial- Treated Patients (Study Years)	Treatment Difference Untreated-Treated (95% CI)
Finland (1943)	≥ 12 years old bacteremic and nonbacteremic	N=2,832 (1929-1940)* 41%	N=1,220 (1939-1941) 17% (sulfonamides)	24% (21%, 27%)
Dowling and Lepper (1951)	≥ 10 years old bacteremic and nonbacteremic	N=1,087 (1939, 1940)* 30.5%	N=1,274 (1938-1950) 12.3% (sulfonamides) N=920 (1938-1950) 5.1% (penicillins and tetracyclines)	18.2% (15%, 21%) 25.4% (22%, 28%)
Austrian and Gold (1964)	≥ 12 years old bacteremic	N=17 (1952-1962) 82%	N=437 (1952-1962) 17%	65% (41%, 79%)

\* Historical control patients

## 結語

不劣性試驗雖早已廣泛用於第三期療效確認性試驗，惟因試驗設計沒有安慰劑組的限制，試驗結果要能解讀，以宣稱試驗藥品的療效，必須符合多項條件，包括臨界值的選取須保守且合理、對照藥品須具有檢測靈敏度與符合一致性的假設、以及試驗品質必須良好等。常用的統計分析方法有 95%-95%法與合成法。與合成法相比，95%-95%法須事先決定 M1 與 M2，可於試驗執行前與法規單位討論所選取臨界值之適當性。由於 95%-95%法相較於合成法其標準誤較大，較不受輕微偏離一致性假設的影響，為目前較常使用的分析方法。至於須要執行幾個 NI 試驗方可支持新藥之療效，本指引建議若對於 M1 的選取沒有把握，可考慮執行兩個療效確認性試驗；若有其他相關理論支持其藥理分類與機轉與對照藥品的相似性，也許一個療效確認性試驗即已足夠。

## 參考文獻

1. Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) (2016) Guidance for industry, Non-Inferiority Clinical Trials to Establish Effectiveness.

[www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/)



[UCM202140.pdf](#).

2. Center for Drug Evaluation and Research (CDER) (2010) Guidance for industry, antibacterial drug products: use of non-inferiority trials to support approval.

<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070951.pdf>

3. 吳雅琪. 不劣性試驗統計審查重點. 當代醫藥法規. 2012 年第 22 期

4. Center for Drug Evaluation and Research (CDER) (2014) DRAFT GUIDANCE : Guidance for Industry : Community-Acquired Bacterial Pneumonia: Developing Drugs for Treatment.

[www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm123686.pdf](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm123686.pdf).