



臨床試驗缺失值審查要點考量

吳雅琪¹

一般而言，臨床試驗收集資料都會要求資料的完整性，盡可能不要有缺失值。但實際上，有很多情況下無法達到此要求。例如受試者死亡或是拒絕繼續參與試驗，其理由如治療無效導致提早退出試驗或由於治癒而未再參與訪視、或因為藥物不良反應致無法繼續或因受試者搬遷無法再追蹤等。其中有些原因可能與藥物有關，但有些原因則與藥物是完全無關的。然而如果不探討缺失資料產生的原因與選擇適當的處理來方式，可能會造成母體參數估計上的偏差(bias)。

針對臨床試驗缺失值的問題，相關的法規指引有 1998 年 ICH-E9 臨床試驗統計指導原則 (Statistical Principles for Clinical Trials)。歐盟先於 1999 年公布缺失值的審查要點 (Points to consider on missing data)，2010 年發佈確認性臨床試驗中缺失值的處理準則 (Guideline on missing data in confirmatory clinical trials)。美國 FDA 雖然尚未發布正式的指引，但已請美國國家科學院探討缺失值處理相關議題，亦於 2010 年出版 *The prevention and treatment of missing data in clinical trial* 一書。這些相關的指導原則，可作為新藥研究的重要參考資料。

臨床試驗因主要以 ITT (intent to treat) 群體來分析，所以較易有缺失值的問題。若遇缺失值不處理，則違背了 ITT 原則 (ITT principle)。所謂 ITT 原則是指不論受試者實際接受的是何種治療、不論受試者的順從性及有無依計畫書執行，皆應列入分析，且依隨機分派時的組別來分析。因此在療效確認性試驗，ITT 通常定義為所有接受隨機分派者，主要目的為維持隨機分派最重要的特質：即各組間所有可測量、或不可測量的變項會具可比性 (comparability)。

在第三期療效確認性試驗以 ITT 為主要分析群體，為不違反 ITT 原則，通常要求對缺失值進行插補。目前文獻中已有許多方法用於處理缺失值，但是尚無哪一種方法被認為是適用於所有情況的黃金標準 (gold standards)。以下介紹幾種常見的缺失值處理方式並探討其優缺點：

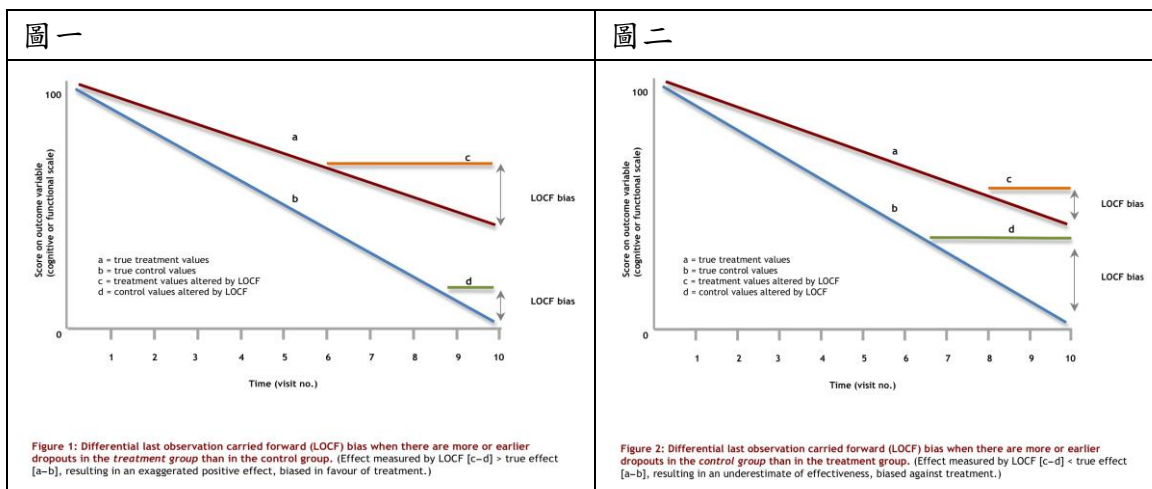
第一種是完整資料分析法 (complete case analysis) 或是可觀察群體分析 (available case analysis)，前者是將包含缺失值的受試者資料完全排除；後者是當分析時涉及到該缺失值，該受試者將不列入分析。這兩種方法的優點是簡單且不須插補資料，但缺點是

¹ 新藥科技組



會因為任意排除受試者，而破壞各組間的可比性，導致評估偏差。此外，僅採用完整資料去進行統計分析，對試驗結果的解釋易產生其他的問題，如：因資料不完整，導致分析中可用的有效病例數量減少，從而降低了試驗的檢定力 (power)，也就是說缺失值越多，統計檢定力越低。另外，不完整的資料有可能是一些極端值，如治療無效導致提早停止使用試驗藥物或由於治癒而未再訪視。少掉這部分缺失的資訊會導致變異性被低估，反而使得療效差異之信賴區間錯誤地變窄，而導致評估偏差。而且，少掉這部分病患後之病人群體，有時已無法代表整個目標病人群 (targeted patient population)。因此，我們通常不會將完整資料分析法或可觀察群體分析法作為療效確認性試驗的主要分析方法，而是將此兩種方法視為敏感度分析，以探討缺失值對主要統計分析結果之影響。

第二種是單一數值插補法 (Simple imputation method)，包括以最終觀察值(LOCF; last observation carried forward)、基礎值(BOCF; baseline observation carried forward)或是平均值(mean imputation)來取代缺失值。其中 LOCF 是臨床試驗中最常用的單一數值插補法，其優點是操作簡單，也較符合 ITT 原則：即所有接受隨機分派者只要於隨機分派後至少有一個觀察值者皆可列入分析。但值得注意的是，LOCF 是假設受試者離開試驗時與之後的觀察值是維持恆定的。在這個假設下，LOCF 方能產生不偏估計量(unbiased estimate)。倘若這個假設不成立，LOCF 將可能會高估或低估各組之參數值，也可能因為變異數低估而導致型一誤差(Type I error) 擴增。在疾病的適用範圍中，若預期病情會隨時間自動好轉的(例如急性疼痛等的疾病)，採用 LOCF 處理缺失值是會保守地估計(即低估)各組之療效；然而當受試者的病情預期會逐漸惡化 (如阿茲海默症等的疾病)，LOCF 則不會保守地估計(即高估)各組之療效。另外，不少文獻研究(Molnar, 2009)提醒執行此類病情預期會逐漸惡化之臨床試驗應注意：當試驗藥物組與對照組相比，出現較早退出試驗、較高比例退出試驗，或是疾病惡化較快速，則 LOCF 不是保守的方式用以處理兩組間的比較 (圖一，採自參考資料 4)。反之，若是對照組比試驗組較早退出試驗、較高比例退出試驗，或是疾病惡化較快速，則 LOCF 可算是保守的方式處理兩組間的比較(圖二，採自參考資料 4)。



BOCF 是以隨機分派前之基礎值來插補缺失值，其優點是符合 ITT 原則，缺點是忽略試驗期間之觀察值之變異，容易造成變異數低估，導致型一誤差擴增的問題。這種插補常見於慢性疼痛試驗(chronic pain trial)，但須確認符合受試者退出試驗後疼痛狀況會返回到基礎值的前提之下，方可考慮使用。

平均數插補法 (mean imputation) 是指變數的缺失值以該樣本中已觀察到案例之平均值來插補。其優點亦是符合 ITT 原則，其缺點是未考慮受試者間之差異，容易造成變異數低估，導致型一誤差擴增的問題。

第三種插補方式是多重插補法 (multiple imputation, MI): 多重插補法是指利用隨機重複地由多個機率模式中抽取由兩個以上的值來取代缺失值的插補方式。其優點為可將有缺失值的資料形成完整資料進行分析且這種方法可反應抽樣變異並評估不同模式間的敏感性 (sensitivity)。其缺點為此方法產生、儲存和分析上都比單一數值插補法複雜。

第四種是多變量常態線性模型 (Multivariate Normal Linear Model)，這主要是 Mallinckrodt 等於 2001 提出 restricted maximum likelihood-based mixed model repeated measures (MMRM) 統計分析法。這種方法不需插補資料，而是將同一受試者試驗期間所有觀察到的資料皆列入分析，並考慮同一受試者量測值間之變異。Mallinckrodt (2001) 也以模擬的方式比較 MMRM 與 ANOVA LOCF 兩種處理方式，其結果顯示因 MMRM 較不會低估變異數，且其犯型一誤差的機率較 LOCF 低。另，Siddiqui et al 亦於 2009 年比較 25 個申請新藥查驗登記的案例，包括 48 個第三期雙盲臨床試驗，108 個試驗組與對照組 (安慰劑或是有效對照要) 之比較。於此 108 個組間比較中，若同時採用 MMRM 與 ANCOVA LOCF 兩種方式處理缺失值，兩種處理方式結果很相似，也就是說有 94% (101/108) 的結果是一致的，剩下 7 組不一致結果其中 5 組 LOCF 顯示顯著的結果，而 MMRM 的結果則是不顯著。



第五種是模式混和模型(Pattern Mixture Model)，這是將缺失模式引入隨機效應模型(random effect model)，不但允許各種缺失值模式下的截距(intercept)不同，其他參數之間也可以不同。

至於應該選擇哪一種方式作為主要療效指標之缺失值處理方式，建議宜先瞭解缺失值的產生是屬於以下哪一種類型：

- 第一是完全隨機發生的缺失值(missing completely at random, MCAR): 資料發生缺失值是完全隨機的，與受試者已觀察到的資料及未觀察到(缺失)的資料完全無關。例如缺失值的產生與藥物副作用無關，與試驗組別無關，亦與疾病嚴重度與症狀均無關。常見的例子如受試者因為搬家的原因無法繼續參與試驗。
- 第二是隨機發生的缺失值(missing at random, MAR): 資料發生缺失值與受試者已觀察到的資料有關，而與未觀察到的資料無關。例如一個糖尿病試驗，計畫書事先訂定退出試驗條件是: 受試者若試驗期間空腹血糖值超過 270 mg/dl 則需停止試驗用藥並退出試驗。
- 第三是非隨機發生的缺失值(missing not at random, MNAR): 資料的缺失和未觀察到(遺失)的資料有關。當資料缺失值發生的原因不是 MCAR 或 MAR，即可歸為是 MNAR。

基本上，若缺失資料是屬於完全隨機缺失(MCAR)，文中所提到的五種方式均可得到不偏估計量。只是採用 complete case 來分析，其缺點是檢定力較低；若以 LOCF 來插補也需確認觀測值在之後的時間點仍可保持恆定的假設是否成立。若缺失資料是屬於隨機缺失(MAR)，使用 complete case，LOCF 等單一數值插補方法是會導致評估偏差的，建議可採用 multiple imputation (MI) 或 MMRM 分析方法。若缺失資料是屬於非隨機缺失(MNAR)，以上這四種方式(complete case, LOCF, MI 與 MMRM) 均會導致評估偏差，建議採用模式混和模型(Pattern mixture models) (Little, 2002) 以避免 MNAR 資料所產生的偏差。

然而，在多數情況下，幾乎很難或不可能去驗證缺失值和未觀察到的變數間是否有相關。有鑒於缺失值可導致潛在偏差，在試驗的設計和執行階段，應盡所有努力儘量避免缺失值的出現。但是，即使有了這些努力，通常還是會有缺失值出現；而不同的缺失值處理方法，會影響到試驗結論。採用保守的插補方法是必要的。另外，建議試驗宜執行敏感度分析，以比較不同處理方法對試驗結果的影響，亦作為主要分析的支持性結果。敏感度的分析是利用不同的缺失處理方法對研究結果影響的分析。如果敏感性分析的結果是一致的，並得到對療效評估得到相當近似的結果，那麼這就說明了所損失的資訊對整個研究結論不造成嚴重影響。在這種情況下，證實了結論的穩定性。相反地，如果敏感性分析出現結果不一致，那麼對試驗結果的影響就必須進行討論。



由於臨床試驗幾乎是無法避免缺失值的發生，因此法規單位通常會要求在試驗計畫書中應事先載明所選用處理缺失數據的主要方法，廠商於樣本數計算時，亦應考量可能發生缺失值的比例，以便最終分析時具有足夠的檢定力。由於敏感度分析有助於對所選用的方法的適當性提供依據，計畫書亦應載明所採用之敏感度分析方法。如果敏感度的分析結果一致，且療效估計相似，則其缺失值較不會影響試驗結果的可靠性；反之，如果敏感度的分析結果不一致，則其試驗結果的可靠性會受影響。

參考文獻:

1. Points to consider on missing data in clinical trials (CPMP/EWP/1776/99)
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003641.pdf
2. Committee for Medicinal Products for Human Use. *Guideline on Missing Data in Confirmatory Clinical Trials*. European Medicines Agency: London, 2010.
3. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press: Washington DC, 2010.
4. Molnar et al (2009) *Have last-observation-carried-forward analyses caused us to favor more toxic dementia therapies over less toxic alternatives? A systematic review*. Open Medicine, 3, e31-e50.
5. Mallinckrodt et al (2001) *Type I error rates from mixed effects model repeated measures versus fixed effects ANOVA with missing values imputed via last observation carried forward*. Drug Information Journal, Vol 35, pp1215-1225.
6. Siddiqui et al (2009) *MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets*. Journal of Biopharmaceutical Statistics 19, 227-246.
7. Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data* 2nd edition, John Wiley & Sons, INC., Publication