National Translational Medicine and Clinical Trial Resource Center
國家轉譯醫學與臨床試驗資源中心

第二期臨床試驗設計

報告者：蔡靜慧

參考文獻：

Randomized Phase II Designs

Larry Rubinstein, John Crowley, Percy Ivy, Michael LeBlanc, and Dan Sargent

Clin Cancer Res 2009;15:1883-1890.

摘要

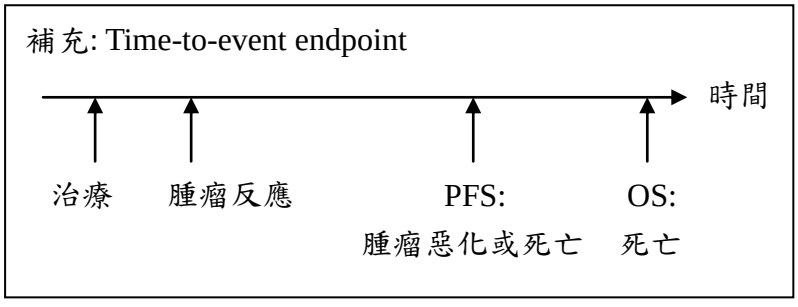
文章簡介第二期臨床試驗中常見的名詞，並比較多種第二期臨床試驗的研究設計

大綱

- ✧ 第二期臨床試驗常用名詞
- ✧ 臨床試驗設計的優缺點及適用時機
- ✧ 第二期臨床試驗終點之比較

前言

第二期臨床試驗主要目的是瞭解藥物的療效，傳統是進行單臂的兩階段設計(single arm two-stage design)，比較腫瘤的反應率(tumor response)，型一誤差(type I error)設為 5~10%，型二誤差(type II error)設為 10~20%。然而目前已有需多標靶藥物研發，此類分子層次的藥物可讓腫瘤停止生長，但不一定會使腫瘤縮小，因此會以總存活時間(overall survival; OS)或無疾病進展的存活時間(progression-free survival; PFS)來當作試驗終點(endpoint)。試驗設計可以為單臂研究，與歷史對照組(historical controls)比較 OS 或 PFS 的中位數，或者使用隨機性(randomized)試驗。



表一、第二期臨床試驗常用名詞

名詞	說明
歷史對照組 (Historical controls)	如果使用單臂設計時，通常會找過去曾接受過標準治療 (或未接受過治療)的病人當作對照組，藉以評估目前實驗的藥物是否較具療效。
型一誤差 (Type I error，α)	實際上，實驗組的治療效果沒有優於對照組，但卻將實驗組誤判為有治療效果的機率，又稱「偽陽性率(false-positive rate)」。
型二誤差 (Type II error，β)	實際上，實驗組的療效優於對照組，但卻將具有治療效果的實驗組誤判為沒有效果的機率，又稱「偽陰性率 (false-negative rate)」。統計檢定力(power)= 1 –β。
Z_α 和 Z_β	指α或β的標準常態分佈值(standard normal distribution value)。例如$Z_{0.05}$ = 1.645。
二項比例檢定 (Binomial proportion test)	在隨機試驗中，檢定實驗組的腫瘤反應率(tumor response rate)是否與對照組不同。
風險比 (Hazard ratio)	風險(hazard)是指瞬時事件發生率(instantaneous failure rate)。風險比是指兩組不同治療的風險相比值，也等於兩組發生某事件的中位時間(median time to failure)的比值。
對數等級檢定(Logrank test)	用來分析兩組間存活是否有差異的統計方法，並且考慮設限資料(censored data)。設限資料為未發生事件(event)者，事件依研究者設定，通常為死亡或疾病復發，因此設限資料包含失去追蹤者(loss-to-follow-up)或試驗結束時仍未發生事件者(end-of-study)。

表二、臨床試驗設計的優缺點及適用時機

臨床試驗設計	優缺點及適用時機
歷史對照組 (Historic controls)	適用於以腫瘤反應(tumor response)為終點的試驗。所需要的樣本數最少，但會存在一些偏差。如使用PFS為終點時，實驗組與歷史控制組在重要的預後因子、照護品質或追蹤時間有差異時，就會出現偏差。但若可以從兩組病人中獲得相關資訊時，統計分析可校正重要

	的共變項。
參考組 (Reference arm)	隨機分配一個小型的參考對照組，以此確認歷史對照組的適當性。但實際上很難比較參考對照組和歷史對照組間的差異，因此並不建議使用此方法。
第二/三期臨床試驗 (Phase II/III trial)	為了有效利用收取的病患，將第二期試驗的成果合併至第三期臨床試驗，當第二期試驗發現實驗組結果低於對照組時，第三期試驗才會終止。若第二期試驗樣本數不夠大時，很難設定α或β值。適用於研究者有很大的理由可證明實驗方案足以進入第三期試驗。一般來說，不建議使用於第二期的篩選(phase II screening)。
選擇設計；選出最勝者的設計 (Selection; pick-the-winner design)	為一種有效率方法，可同時比較兩組或兩組以上的試驗組，從中找出最佳結果的組別。但由於一開始並沒有假設實驗組會優於標準治療，當多組比較時，可能會選出沒有優於標準治療的實驗方案。因此，實際執行時會讓實驗組直接跟歷史對照組比較。所需樣本數較少，每組僅需 29~37 人即可達到 90%的檢力(兩組腫瘤反應率差異達 15%)。不適用於討論試驗組比標準方案多出的額外作用。
篩選設計 (Screening design)	以校正型一誤差、型二誤差和目標差異值，計算在隨機性第二期試驗中，所需要觀察到的事件發生數目，藉此算出所需要的樣本數。適用於討論實驗組比標準方案(通常當作對照組)多出的額外作用，並適用於以 PFS 為終點的評估。 Table 1. 為以 PFS 為終點時，各種不同條件下事件發生的數目(failures)。假設α=β= 10%，風險比為 1.75，則需要觀察到 84 位事件發生者，如果研究的癌症疾病死亡率為 87.5%，則所需總樣本數為 96 位(84/0.875=96)。 Table 2. 為治療追蹤一段時間後以PFS率為終點，各種不同情況下所需要的總樣本數。與table 1相比，可以看出二項比例檢定力不如對數等級檢定。
隨機終止設計 (Randomized discontinuation design)	以實驗方案初步治療所有的病人，針對達到疾病穩定狀態(stable disease)的次群病人，再給予隨機分派至實驗組或安慰劑組(或標準治療)。優點是可以減少接受安慰劑治療的病人，兩組的預後因子分配會比較平均，因此試驗用藥若是有效果，可以較容易觀察到效果。但缺點是一開始所有病人均為實驗組，若此實驗無

	效，會有太多病人接受無效實驗。

Table 1. Approximate required numbers of observed (total) treatment failures for screening trials with PFS endpoints, using the logrank test

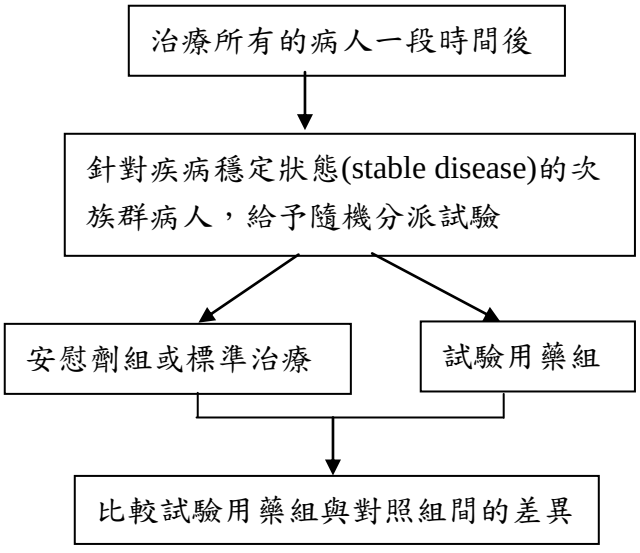
Error rates	Hazard Ratios (Δ)			
	Δ = 1.3	Δ = 1.4	Δ = 1.5	Δ = 1.75
(α,β) = (10%,10%)	382	232	160	84
(α,β) = (10%,20%) or (20%,10%)	262	159	110	58
(α,β) = (20%,20%)	165	100	69	36

NOTE: Calculations were carried out using nQueryAdvisor 5.0 software (Statistical Solutions) based on methods given in Collett (26) with one-sided α.

Table 2. Approximate required numbers of total patients for screening trials with PFS rate (at a specified time) endpoints, using the binomial proportion test

Error rates	PFS Rates at a given time point (with equivalent hazard ratios Δ)			
	20% vs. 35% (1.53)	20% vs. 40% (1.76)	40% vs. 55% (1.53)	40% vs. 60% (1.79)
(α,β) = (10%,10%)	256	156	316	182
(α,β) = (10%,20%) or (20%,10%)	184	112	224	132
(α,β) = (20%,20%)	126	78	150	90

NOTE: Calculations were carried out using nQueryAdvisor 5.0 software (Statistical Solutions) based on methods given in Fleiss et al. (27) with one-sided α.



圖：隨機終止設計(Randomized discontinuation design)

隨機性第二期臨床試驗終點之比較

在第二期臨床試驗中，使用 PFS 為終點明顯的優於 OS，因為相較於治療到死亡的時間(time-to-death)，治療到腫瘤惡化的時間(time-to-progression)較短，事件發生(failure)較多，HR 也較大，更適用 logrank test。但缺點是 PFS 測量不易，會受研究者的治療誤差或追蹤時間所影響。盡可能做到盲目(blinding)試驗。