



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

## 癌症第二期試驗採用 Simon 兩階段設計之新思維

吳雅琪<sup>1</sup>

### 前言

基於倫理考量，研發中癌症新藥的早期臨床試驗，希望儘早得知其具潛在的療效可能性，以決定可否進一步地發展。Simon 的兩階段設計<sup>[1,2]</sup>即是針對此項需求的試驗設計。此設計通常根據歷史資料為試驗中新藥設定一個無療效門檻(例如：無效反應率或歷史反應率  $p_0 = 0.1$ )，再設定一個欲達到的有效反應率(例如：有效反應率  $p_1 = 0.3$ )。

Simon 的兩階段設計，第一階段先招募  $n_1$  位受試者，若此等受試者呈有效反應的人數，低於預設標準  $r_1$  時，則停止試驗，並認定試驗藥無須進一步發展；反之則進入第二階段，並繼續招募  $n_2$  位受試者，共收納  $n = n_1 + n_2$  位受試者。當兩個階段之具療效反應總人數，低於預設標準  $r$  時，則認定試驗藥無潛在療效；反之，則認定試驗藥可進一步發展。上述四個參數( $r_1$ 、 $n_1$ 、 $r$  及  $n$ )的決定，可根據 optimal 準則或 minimax 準則，於給定的顯著水準 (significant level,  $\alpha$ ) 與檢定力 (power,  $1-\beta$ ) 下，運用二項式分配推導而得<sup>[1,2]</sup>。目前已有免費的計算平台或 R 軟體<sup>[3]</sup>可協助計算。

Simon 兩階段設計的優點，除避免太多受試者接受無效藥的治療，亦可達到提早淘汰無效藥品或無效劑量的目的；值得注意的是，Simon 兩階段的設計不會因療效太好而提早停止試驗。若能進入第二階段，最終的分析將會合併兩階段的試驗資料，以判定新藥的反應率是否優於歷史反應率( $p_0$ )。

由上述可知，Simon 兩階段設計中的決策準則，基本上仍是根據假說檢定，亦即控制偽陽性(false positive)及偽陰性(false negative)的犯錯機率( $\alpha$  及  $\beta$ )。惟試驗的統計分析，例如有效反應率的估計(包括點估計及信賴區間估計)，以及檢定的  $p$  值；研究者常以樣本比率(sample proportion)及其信賴區間來估計真實反應率(true response rate)。由於估計量忽略了兩階段設計可能提前終止試驗，估計值往往有所偏差(bias)<sup>[4-6]</sup>。因此，本文將介紹如何得到真實反應率的無偏估計<sup>[4-6]</sup>。

此外，Simon 兩階段設計的試驗常見的另一個問題，是執行後最終樣本數可能與預

<sup>1</sup> 財團法人醫藥品查驗中心新藥科技組



先設計的樣本數不符；無論是超收或是少收，預設的療效臨界值 ( $r_1$  與  $r$ ) 恐不再適用。如何在樣本數改變的情況下，得到有效反應率的無偏估計以及尋找新的臨界值<sup>[8]</sup>，以正確地判斷試驗新藥是否具有潛在的療效，本文亦將針對此項議題有所討論。

同時，本文也將介紹，如何讓 Simon 兩階段設計推展至同時收納多種不同腫瘤適應症病患的籃型試驗(basket trials)<sup>[9]</sup>、及第二期晚期試驗常用的隨機分派設計。

### Simon 兩階段設計真實反應率之無偏估計量

於 Simon 兩階段試驗中，若以  $m(=1 \text{ or } 2)$  代表試驗停止的階段， $s$  代表於該階段停止前所累積具療效反應者的人數，目前常用的估計量為合併成一個階段所計算的樣本百分比(sample proportion)或稱為最大概似估計量(maximum likelihood estimate, MLE)，亦即

$$\hat{p}(m, s) = \begin{cases} \frac{s}{n_1} & \text{if } m = 1 \\ \frac{s}{n_1 + n_2} & \text{if } m = 2 \end{cases}$$

然而，Jung 等人的研究<sup>[4]</sup>指出，MLE 估計量未考慮到提前終止試驗的可能性，證實會有偏差。而一項試驗的歷史反應率( $p_0$ ) 若是經由 Simon 兩階段設計以 MLE 估計，則偽陽率會增加；反之，若經由 Simon 兩階段設計以 MLE 估計該試驗之療效反應率，則會降低該試驗的檢定力，提早扼殺一個可能具有療效的藥品。可見在藥物研發的早期試驗，得到一個無偏估計量的重要性。

無偏估計量的計算，主要來自 Jung 等人的研究<sup>[4]</sup> 藉由 Rao-Blackwell 定理推演出均勻最小變異無偏估計量(uniformly minimum variance unbiased estimator, [UMVUE])，完整細節可參考 Jung 等人著作<sup>[4]</sup>附錄 A 與 B。值得注意的是，若試驗提早終止 ( $m=1$ )，則此無偏估計量與 MLE 相同；若試驗進入第二階段 ( $m=2$ )，其估計量就須要修正，其公式如下：

$$\tilde{p}(m, s) = \begin{cases} \frac{s}{n_1} & m = 1 \\ \frac{\sum_{x_1=\max(r_1, s-n_2)}^{\min(s, n_1)} \binom{n_1-1}{x_1-1} \binom{n_2}{s-x_1}}{\sum_{x_1=\max(r_1, s-n_2)}^{\min(s, n_1)} \binom{n_1}{x_1} \binom{n_2}{s-x_1}} & m = 2 \end{cases}$$



有關  $m=2$  時信賴區間的無偏估計，亦請參考該文獻本文與附錄<sup>[4]</sup>。

## 案例介紹

接下來，我們引用 Razak 等人文章的案例<sup>[7]</sup>，來說明 Simon 兩階段設計試驗其反應率如何檢定及估計：在給定  $(p_0, p_1, \alpha, 1-\beta) = (0.05, 0.15, 0.05, 0.8)$  之下，Simon optimal 準則決定的一、二階段臨界值與樣本數  $(r_1/n_1, r/n)$  分別為  $(2/23, 6/56)$ 。

該案例試驗在第一階段結束後，於前 23 位受試者中觀察到 4 位有療效反應(完全反應或部分反應)，依試驗設計，可進入第二階段。而於合併一、二階段收錄的前 56 位受試者中，觀察到 7 位具療效反應，大於臨界值的 6 位，故宣稱試驗達到預期的目標。此為 Simon 兩階段設計的檢定過程。

至於反應率的估計，傳統通常以  $7/56 = 0.125$  為點估計的估計值，再以二項式精確方法(binomial exact method)計算雙尾 90%信賴區間，得到  $(0.060, 0.222)$ 。若以 UMVUE 無偏估計量套用上述公式，當觀察到  $(m, s) = (2, 7)$ ，則點估計值， $\tilde{p}(2, 7)$ ，為 0.138，高於 MLE 的 0.125。

$$\tilde{p}(2, 7) = \frac{\sum_{x_1=2}^7 \binom{23-1}{x_1-1} \binom{33}{7-x_1}}{\sum_{x_1=2}^7 \binom{23}{x_1} \binom{33}{7-x_1}} = \frac{\sum_{x_1=2}^7 \binom{23-1}{x_1-1} \binom{33}{7-x_1}}{\sum_{x_1=2}^7 \binom{23}{x_1} \binom{33}{7-x_1}} = \frac{27,882,107}{202,171,288} = 0.138 > 0.125$$

Jung 等人<sup>[4]</sup>也證明，當試驗未在第一階段提早停止時，UMVUE 反應率  $\tilde{p}(m, s)$  的估計量，會隨著階段  $(m)$  與反應人數  $(s)$  排序的遞增而遞增，亦即

$$\begin{aligned} \tilde{p}(1, 0) &< \tilde{p}(1, 1) < \dots < \tilde{p}(1, r_1 - 1) \\ &< \tilde{p}(2, r_1) < \tilde{p}(2, r_1 + 1) < \dots < \tilde{p}(2, n_1 + n_2) \end{aligned}$$

反之，MLE 的估計量  $\hat{p}(m, s)$ ，僅在各階段內依反應人數排序遞增而遞增，而兩階段之間並無依順序遞增的特性(詳見表一)。此外，當  $r_1 \leq s < r$  時，兩個估計量差異最大，也就是說，當試驗進入第二階段，若總反應人數未達臨界值  $r$  時，MLE 會明顯低估真實反應率。但當  $s \geq r$  時，兩個估計量會隨反應人數  $(s)$  愈多而差異愈小。



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

表一、第一、二階段下 UMVUE 與 MLE 的估計值及其差異

| m                             | s                 | UMVUE | MLE          | Difference (UMVUE – MLE) |
|-------------------------------|-------------------|-------|--------------|--------------------------|
| Stage 1 (n <sub>1</sub> = 23) |                   |       |              |                          |
| 1                             | 0                 | 0     | 0.00 (0/23)  | 0.00                     |
| 1                             | 1                 | 0.043 | 0.043 (1/23) | 0.00                     |
| Stage 2 (n = 56)              |                   |       |              |                          |
| 2                             | r <sub>1</sub> =2 | 0.087 | 0.036 (2/56) | 0.051                    |
| 2                             | 3                 | 0.095 | 0.054 (3/56) | 0.041                    |
| 2                             | 4                 | 0.103 | 0.071 (4/56) | 0.032                    |
| 2                             | 5                 | 0.114 | 0.089 (5/56) | 0.025                    |
| 2                             | r = 6             | 0.125 | 0.107 (6/56) | 0.018                    |
| 2                             | 7                 | 0.138 | 0.125 (7/56) | 0.013                    |
| 2                             | 8                 | 0.152 | 0.143 (8/56) | 0.009                    |
| 2                             | 9                 | 0.167 | 0.161 (9/56) | 0.006                    |
| ⋮                             |                   |       |              |                          |
| 2                             | 56                | 1.00  | 1.00         | 0.0                      |

另外，亦可以利用  $\tilde{p}(m, s)$  隨階段與反應人數  $(m, s)$  的增加而遞增的特性，得到 UMVUE 估計量的精確雙尾 90% 信賴區間以及 p-value。

經計算，該案例的 UMVUE 雙尾 90% 信賴區間為  $(0.062, 0.214)$ ，信賴區間的下界大於  $p_0$  (即 0.05)。而基於 UMVUE 檢定虛無假說  $(H_0: p \leq p_0)$ ，當觀察到  $(m, s) = (2, 7)$  時，p 值為  $\Pr(\tilde{p}(m, s) \geq 0.138 \mid p = p_0) = 0.0188$ ，亦小於顯著水準之 0.05，本試驗達到預期的目標。有關 UMVUE 的點估計、信賴區間及對應的 p 值，已有 R 程式<sup>[3]</sup>支援相關的計算。

## 樣本數改變情況下的估計與檢定

試驗執行後，最終總樣本數常會比原先設計超收或少收，原因不外乎不合格或無法評估的受試者被排除；或多中心試驗，在各中心同時收案情況下，很難保證最終樣本數與原先設計的完全相同；或由於收案困難，以至於受試者人數比原設計為少。無論是超收或是少收，事先設定的臨界值  $(r_1$  與  $r)$  將不再適用。

所幸前述所介紹的 UMVUE 估計量，已證實在樣本數改變情況下，仍是適當的無偏估計量<sup>[4,5]</sup>，因此可藉用此估計量之信賴區間或 p 值，判定試驗是否達到預期的目標。

讓我們再回到 Razak 的案例<sup>[7]</sup>，此試驗第二階段實際上是收錄了 40 位受試者(多於



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

原先設計的 33 位)·總樣本數達 63 位(超過原先設計的 56 位)。於 63 位受試者中·共觀察到 8 位有反應者。因樣本數有變動·事先設定的臨界值 ( $r = 6$ ) 將不再適用·欲評估試驗是否達到預期的目標·宜採用無偏估計的 UMVUE·得到反應率的點估計為 0.141·亦高於 MLE 估計量 ( $8/63 = 0.127$ )。

$$\tilde{p}(2,8) = \frac{\sum_{x_1=2}^8 \binom{23-x_1}{x_1-1} \binom{40}{7-x_1}}{\sum_{x_1=2}^8 \binom{23}{x_1} \binom{40}{7-x_1}} = 0.141 > 0.127$$

此外·雙尾 90%信賴區間修正為 (0.066, 0.213)·信賴區間下界大於  $p_0$ (即 0.05)。檢定虛無假說  $H_0: p \leq p_0$  之  $p$  值為 0.012·亦小於顯著水準之 0.05·故得到相同的結論·即本試驗達到預期的目標。

## 樣本數改變情況下新臨界值的決定

至於樣本數改變情況下·新的臨界值該如何決定? 假設在第一階段已觀察到  $x_1$  位有效反應者(即  $X_1 = x_1$ )·而第二階段的人數有改變(如  $n_2' > n_2$ )·依目前已發表的文獻·有兩種可行的方法可供參考:一是 Koyama 與 Chen<sup>[6]</sup>(簡稱 KC)提出的新臨界值  $r'(x_1)$ ·可以由新的樣本數及臨界值之型一誤差條件機率(new conditional type I error)小於或等於原先樣本數及臨界值之型一誤差條件機率的不等式來推導·亦即

$$P_{p=p_0} \left( X_2' \geq r'(x_1) - x_1 \mid X_1 = x_1 \right) \leq P_{p=p_0} (X_2 \geq r - x_1 \mid X_1 = x_1)$$

KC 的想法是很直觀的·就是在第一階段每個可能  $x_1$  觀察值(即  $X_1 = x_1$ )條件下·新設定臨界值所產生型一誤差條件機率都小於或等於原設定臨界值的型一誤差·就可保證整體型一誤差機率可控制在  $\alpha$ 。但很明顯的·KC 的方法偏於保守·從而降低試驗的檢定力。針對這一點·Zeng 等人<sup>[8]</sup>(簡稱 Zeng)則提出在整體型一誤差機率控制在  $\alpha$  條件下·以檢定力最大的選為新臨界值  $r''(x_1)$ ·即:

$$\text{Type I error} = \sum_{x_1=r_1}^{n_1} \binom{n_1}{x_1} (p_0)^{x_1} (1-p_0)^{n_1-x_1} \sum_{x_2=r''(x_1)-x_1}^{n_2'} \binom{n_2'}{x_2} (p_0)^{x_2} (1-p_0)^{n_2'-x_2} \leq \alpha$$



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

$$\text{Maximize Power} = \sum_{x_1=r_1}^{n_1} \binom{n_1}{x_1} (p_1)^{x_1} (1-p_1)^{n_1-x_1} \sum_{x_2=r''(x_1)-x_1}^{n_2} \binom{n_2}{x_2} (p_1)^{x_2} (1-p_1)^{n_2-x_2}$$

兩者不同點在於 Zeng 容許某些新臨界值的型一誤差條件機率可大於原設定臨界值，但整體型一誤差是被控制於 $\alpha$ 或以下。Zeng 認為他們的方法比 KC 方法更具檢定力。

以 Razak 超收的案例<sup>[7]</sup>為例，KC 所得到的新臨界值  $r'(x_1 = 4)$  等於 7(表二)，而試驗實際觀察到 8 位有反應者，故結論仍為本試驗達到預期目標的。Zeng 的方法決定的新臨界值  $r''(x_1 = 4)$  為 6(表二)，除了再度認定本試驗達到預期目標；也顯示出 Zeng 方法的門檻比 KC 寬鬆，故比 KC 方法更具檢定力。

表二、第二階段收案人數從 33 人增加到 40 人採用兩種方法計算的新臨界值

| First stage response: $x_1$   | $\leq 1$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|-------------------------------|----------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Critical Value: KC approach   |          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| $r'_2(x_1)$                   | 0        | 5 | 4 | 3 | 2 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| $r'(x_1) = r'_2(x_1) + x_1$   | 0        | 7 | 7 | 7 | 7 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| Critical Value: Zeng approach |          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| $r''_2(x_1)$                  | 0        | 5 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| $r''(x_1) = r''_2(x_1) + x_1$ | 0        | 7 | 7 | 6 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| $r'_2(x_1) = r'(x_1) - x_1$   |          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| $r''_2(x_1) = r''(x_1) - x_1$ |          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

## 籃型試驗的兩階段設計(Two-stage designs for basket trials)

近年來，癌症免疫治療藥物之早期研發階段，試驗常採用籃型設計(basket design)，即同時納入多種不同腫瘤適應症的患者，以評估試驗新藥是否有效。在此種設計下，倘若各腫瘤適應症分別採用 Simon 兩階段設計，可能因未能控制整體型一誤差，導致對腫瘤適應症無效的試驗新藥，成功進入第三期試驗。另一方面，若採用多重檢定策略，適當地控制整體型一誤差，則將需要更多的樣本數。因此，研究者常在有限的受試者人數下，將所有腫瘤適應症受試者收集的數據，整合分析以評估試驗新藥的療效；惟選擇的試驗設計，須注意當試驗藥對大多數腫瘤適應症皆無效時，可能稀釋某些有效適應症的療效。例如，Vemurafenib 對 BRAF-V600 突變型 NSCLC 和黑色素瘤有效，惟對直腸結腸癌無效，若將所有腫瘤患者合併起來分析，則在樣本數較少的情況下，可能過早扼殺一個有效的藥物<sup>[9]</sup>。



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

有鑑於此，Zhou 等人<sup>[9]</sup>提出將 Simon 兩階段設計之 optimal 準則和 minimax 準則推廣至上述多腫瘤適應症的籃型試驗。其想法是在第一階段先找出試驗藥無效的腫瘤適應症，而第二階段則將試驗藥可能有潛在療效的腫瘤適應症合併分析，以評估試驗新藥的整體有效性。Zhou 等人<sup>[9]</sup>所提出的方法可控制整體型一和與型二誤差。

舉例來說，表三顯示針對四種腫瘤適應症( $K = 4$ )，於給定的 $(p_0, p_1, \alpha, 1-\beta) = (0.1, 0.25, 0.05, 0.8)$ ，依 Simon 兩階段設計 optimal 準則，各適應症第一階段須收錄 10 位受試者；若觀察到 3 位或以上有效反應者，則該適應症可進入第二階段，繼續收錄 24 位受試者，最終合併分析之顯著水準控制於 0.024。最終合併分析的樣本數，最少可以為 34 位（一種腫瘤適應症），最多則到 136 位受試者（四種腫瘤適應症）。無論樣本數為何，只要虛無假說檢定的  $p$  值小於 0.024，即可宣稱該試驗藥具有活性。

表三、型一和型二誤差控制於 ( 0.05 , 0.2 ) 之籃型兩階段設計

| $p_0$   | $p_a$ | Optimal |       |            |     | Minimax |       |            |     |
|---------|-------|---------|-------|------------|-----|---------|-------|------------|-----|
|         |       | $r_1$   | $n_1$ | $\alpha^*$ | $n$ | $r_1$   | $n_1$ | $\alpha^*$ | $n$ |
| $K = 4$ |       |         |       |            |     |         |       |            |     |
| 0.01    | 0.1   | 1       | 9     | 0.026      | 25  | 0       | NA    | 0.048      | 20  |
| 0.01    | 0.15  | 1       | 6     | 0.041      | 15  | 0       | NA    | 0.051      | 9   |
| 0.05    | 0.15  | 2       | 19    | 0.024      | 37  | 2       | 31    | 0.014      | 32  |
| 0.05    | 0.20  | 1       | 7     | 0.022      | 22  | 3       | 17    | 0.013      | 18  |
| 0.10    | 0.25  | 2       | 10    | 0.024      | 34  | 3       | 24    | 0.015      | 27  |
| 0.10    | 0.30  | 2       | 7     | 0.046      | 21  | 3       | 13    | 0.022      | 17  |

註：optimal 準則在虛無假說成立下的預期樣本數為最小；minimax 準則為總樣本數為最小<sup>[9]</sup>

## 第二期晚期試驗之隨機分派設計

以上介紹的試驗都為單一組別設計，雖能得到反應率的無偏估計，惟此項設計無法避免選擇偏差(selection bias)及評估偏差，以致估計的反應率常被質疑。因此，單一組別設計建議用於早期探索性試驗；一旦進入第二期晚期試驗，最好還是採用具對照組之隨機分派設計，以減少各項偏差，並達到受試者已知或未知的基礎值在各組間的分佈是平均的。



第二期晚期試驗，在有限樣本數下，若欲控制型一與型二誤差，其試驗設計可考慮：

- 1) 隨機選取設計 (randomized selection design) (或稱選擇獲勝者設計 [pick the winner design])，例如：受試者經隨機分配至平行的兩個或多個劑量組，每個劑量組皆有獨立的決策準則 (例如採用 Simon 兩階段設計)；再選取最有潛力的劑量；及
- 2) 隨機篩選設計 (randomized screening design)，例如：試驗新藥的多個劑量組分別與有效對照藥比較，篩選原則可僅根據主要評估指標的結果，或者當主要評估指標差異很小時，亦可以加入其他考量因素，例如安全性、便利性或生活品質等。因此，此為相當具彈性的試驗設計。

我們先以 Lustberg 等人<sup>[10]</sup> 2010 年所發表文章的案例來示範隨機選取設計。在一項針對無法切除的食道癌和腸胃腺癌病人的臨床試驗中<sup>[10]</sup>，研究者想在試驗藥高低劑量間進行選擇。該試驗採用 Simon 兩階段設計，設定各劑量組的 $\alpha$ 和 $\beta$ 皆為 0.1，潛在療效的認定則依獨立的決策準則。在給定 $(p_0, p_1, \alpha, 1-\beta) = (0.3, 0.5, 0.1, 0.9)$ 條件下，Simon minimax 準則的參數  $(r_1/n_1, r/n)$  為  $(8/28, 16/39)$ 。即每個劑量組最多納入 39 人，最後再針對這兩組的結果進行比較，以選取反應率較高的組別。

其結果詳見於表四，由於兩劑量組最終樣本數都與事先設定之 39 人不同 (高劑量組：40；低劑量組：33)，故可採用 UMVUE 的估計量。結果顯示高劑量組的反應率數值上高於低劑量組 (52.5% vs. 33.6%)。再者，兩劑量組僅高劑量組之雙尾 95% 信賴區間下界高於  $p_0$  之 30%，據此，本試驗可建議以高劑量組作進一步的發展。

表四、Lustberg 案例試驗反應率的估計與檢定<sup>[10]</sup>

| 組別 | 劑量方案  | 反應率 (有反應者/總人數) | 最佳反應率 UMVUE 估計 (雙尾 95% CI) | 檢定 p 值 (虛無假說：反應率 $\leq$ 0.3) |
|----|---|----------------|----------------------------|------------------------------|
| A  | 6 mg/m <sup>2</sup> mitomycin C [MMC] on Day 1 and 125 mg/m <sup>2</sup> irinotecan on Days 2 and 9 | 21/40 (52.5%)  | 0.525<br>(0.39, 0.64)      | 0.0024                       |
| B  | 3 mg/m <sup>2</sup> mitomycin C [MMC] on Day 1 and 125 mg/m <sup>2</sup> irinotecan on Days 2 and 9 | 11/33 (33.3%)  | 0.336<br>(0.20, 0.46)      | 0.396                        |

再以 Johnson 等人<sup>[11]</sup>所發表文章的案例來示範隨機篩選設計。在一項執行於未曾治療的非小細胞肺癌患者的試驗中，比較高低劑量的 bevacizumab 合併 carboplatin 和 paclitaxel 治療與僅以 carboplatin 和 paclitaxel 治療 (有效對照組) 的療效。



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

表五總結 Johnson 案例試驗的結果。顯示 bevacizumab 高低劑量組的反應率數值上都優於對照組；高劑量 bevacizumab 的腫瘤惡化時間(TTP) 顯著長於對照組 ( 中位數 7.4 vs. 4.2 個月；  $p = 0.023$  )，而低劑量 bevacizumab 與對照組的 TTP 則為相近 ( 中位數 4.3 vs. 4.2 個月 )。至於整體存活時間，高劑量 bevacizumab 組數值上優於對照組 ( 中位數 17.7 vs. 14.9 個月；  $p = 0.63$  )，惟未達統計顯著差異。因此，該試驗最終建議高劑量 bevacizumab 與 carboplatin 和 paclitaxel 合併使用可進入第三期試驗。而在之後進行的第三期療效確認性試驗中<sup>[12]</sup>，印證三種藥物合併治療之整體存活率顯著優於對照組(carboplatin 和 paclitaxel 合併治療)，確立此新的組合對未曾治療的非小細胞肺癌患者的療效。

表五、Johnson 案例試驗的結果<sup>[11]</sup>

|  | Bevacizumab-Paclitaxel-Carboplatin groups |                               | Paclitaxel - Carboplatin group (N = 32) |
|--|---|-------------------------------|---|
|  | 7.5 mg/kg bevacizumab (N = 32)            | 15 mg/kg bevacizumab (N = 34) |   |
| Response rate %                          | 28.1%                                     | 31.5%                         | 18.8%                                   |
| Time to progression (TTP) Median (range) | 4.3 (0.2, 12.9)                           | 7.4 (0.7, 12.5)               | 4.2 (0, 12.6)                           |
| Overall survival (OS) Median (range)     | 11.6 (0.2, 56.8)                          | 17.7 (0.8, 57.8)              | 14.9 (0.2, 57.0)                        |

## 結語

癌症藥物早期試驗常以腫瘤反應率為評估指標，常用的試驗設計為 Simon 兩階段設計。本文的目的主要提醒研究者，反應率的點估計及對應信賴區間的估計，宜採用無偏估計量的估計方法，以避免低估真實反應率，降低試驗之檢定力。另外，若試驗最終樣本數與試驗設計時不相同，反應率仍可以所建議之無偏估計量來估計，並以適當的統計方法找尋新的臨界點，判定試驗是否達到預期的目標。癌症免疫治療為近年癌症研究與治療的熱門議題，本文也介紹 Simon 兩階段設計如何推展至同時收錄不同腫瘤適應症病患的籃型設計試驗，以評估試驗新藥是否具潛在的療效。而單一組別的設計往往有選擇偏差及評估偏差之疑慮，因此建議在第二期晚期試驗，可考慮採用具對照組之隨機分派設計，運用不同的統計篩選決策準則，挑選較具潛力的藥物或劑量，作為進一步的發展。

## 參考文獻



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

1. Richard Simon: Optimal Two-Stage Designs for Phase II Clinical Trials. *Controlled Clinical Trials*, 1989; 10:1-10.
2. 王蕙雯: 簡介 Simon 二階段試驗設計 · 當代醫藥法規 · 2012: 15: 11-13.
3. Meinhard Kieser, Marius Wirths, Stefan Englert, et al.: OneArmPhaseTwoStudy: An R Package for Planning, Conducting, and Analyzing Single-Arm Phase II Studies. *Journal of Statistical Software*. 2017; 81: 1-28.
4. Sin-Ho Jung, Kyung Mann Kim: On the Estimation of the Binomial Probability in Multistage Clinical Trials. *Statistics in Medicine*. 2004; 23: 881-896.
5. Sin-Ho Jung: Statistical Issues for Design and Analysis of Single-Arm Multi-Stage Phase II Cancer Clinical Trials. *Contemporary Clinical Trials*. 2015; 42: 9-17.
6. Tatsuki Koyama, Heidi Chen: Proper Inference from Simon's Two-Stage Designs. *Statistics in Medicine*. 2007; 27(16): 3145 – 3154.
7. A. R. Abdul Razak, D. Soulieres, S. A. Laurie, et al.: A Phase II Trial of Dacomitinib, An Oral Pan-Human EGF Receptor (HER) Inhibitor, As First-Line Treatment in Recurrent and/or Metastatic Squamous-Cell Carcinoma of the Head and Neck. *Annual of Oncology*. 2013; 24: 761-769.
8. Donglin Zeng, Fei Gao, Kuolung Hu, et al.: Hypothesis Testing for Two-Stage Designs With Over or Under Enrollment. *Statistics in Medicine*. 2015; 34(16): 2417 – 2426.
9. Heng Zhou, Fang Liu, Cai Wu, et al.: Optimal Two-Stage Designs for Exploratory Basket Trials. *Contemporary Clinical Trials*. 2019; 85: 1-7.
10. Maryam B. Lustberg, Tanios Bekaii-Saab, Donn Young, et al.: Phase II Randomized Study of Two Regimens of Sequentially Administered Mitomycin C and Irinotecan in Patients with Unresectable Esophageal and Gastroesophageal Adenocarcinoma. *Journal of Thoracic Oncology*. 2010; 5: 713-718.
11. David H. Johnson, Louis Fehrenbacher, William F. Novotny, et al.: Randomized Phase II Trial Comparing Bevacizumab Plus Carboplatin and



Paclitaxel With Carboplatin and Paclitaxel Alone in Previously Untreated Locally Advanced or Metastatic Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*. 2004; 22: 2184-2191.

12. Alan Sandler, Robert Gray, Michael C. Perry, et al.: Paclitaxel-Carboplatin Alone or With Bevacizumab for Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*. 2006; 355: 2542-2550.