



## 以無惡化存活時間(PFS)評估抗癌藥物療效之考量

王玫<sup>1</sup>

### 前言

自 1975 至 2009 年，刊登於臨床癌症期刊(Journal of Clinical Oncology)，針對乳癌、大腸直腸癌、及非小細胞肺癌治療所執行的隨機分派、具控制組的試驗，以「無惡化存活時間(Progression Free Survival, PFS)」為主要療效指標的試驗有逐年增加的趨勢；1975-1984 年為 0%，1985-1994 年為 2%，1995-2004 年為 7%，近年來 (2005-2009 年) 則急速攀升至 26%<sup>1</sup>。根據統計，2005 至 2007 年美國食品藥物管理局所核准的抗癌藥物新適應症，依據的臨床試驗，23% 以 PFS 或「自基礎點至發生惡化時間(Time to Progression, TTP)」為主要療效指標<sup>2</sup>，並且自 2000 年至 2010 年獲核准的抗實體腫瘤(solid tumor)藥物中，其樞紐性試驗 29% 以 PFS/TTP 為主要療效指標<sup>3</sup>。儘管如此，多項研究卻認為，僅有治療晚期大腸直腸癌及晚期卵巢癌藥物，且限於細胞毒性藥物(cytotoxic agents)，似乎才有足夠證據顯示 PFS 可以作為「整體存活(Overall Survival, OS)」的替代性療效指標，且截至目前為止，發表的文獻多未能支持治療晚期乳癌、前列腺癌、及非小細胞肺癌藥物之療效，可以 PFS 作為 OS 的替代指標。主要原因為 PFS 特有的測量評估偏差，常易導致錯誤的療效結論的緣故。本篇文章首先就 PFS 療效指標可能造成的評估測量偏差提出討論，為避免此等測量評估偏差產生錯誤的療效結論，亦提出建議執行的敏感度分析(sensitivity analysis)<sup>4</sup>。另外，美國食品藥物管理局於 2007 年頒佈的關於抗癌藥物臨床試驗適用的療效指標規範<sup>5</sup>，其中關於 PFS 的相關規定亦列入本文內容。而由美國 PhRMA 所組成的 PFS 工作小組以及其他學術研究單位針對 PFS 是否一定由「盲性中央獨立評估委員會(Blinded Independent Central Review, BICR)」來評估，其研究結果、會議的結論、以及所提出的建議<sup>6</sup>亦一併列入本文重點。

### 一、PFS 的療效評估偏差及建議的敏感度分析

#### (一) 評估時間點導致的評估偏差

因前一個治療療程相關的不良反應事件，導致下一個療程訪診時間的延後，以致於偵測記錄到發生腫瘤惡化的時間向後延移，造成 PFS 時間的延長；或由於臨床症狀顯示情況變差，受試者在預先設定的訪診時間之外，頻繁地增加額外訪診的次數，因而被偵測到腫瘤發生惡化的機率增加，造成 PFS 時間的縮短。以上兩種情況若治療組及對照組皆為隨機發生，且發生的頻率相當，則不致於造

<sup>1</sup> 財團法人醫藥品查驗中心新藥科技組



成 PFS 評估的偏差；但若加上人為因素，則易產生療效評估的偏差。譬如開放設計的臨床試驗，若試驗主持人每當對照組受試者有療效不佳的跡象時，皆將其訪診的時間提早，而試驗組受試者如果臨床狀況看似穩定則將其訪診的時間延後，諸如此類以人為的方式，主觀地調整訪診的時間，則易造成療效評估的偏差。研究顯示，非排定的訪診，若對照組的頻率明顯較試驗組為多，試驗結果雖然呈現試驗組相較於對照組，療效具有明顯的改善，但通常並非真實的情況<sup>7</sup>。另一種情形則是，兩個治療組別療程的時間長短不同，若無法調整到相同的評估時間，亦可能由於評估時間的不同，導致療效評估的偏差。

一般而言，若試驗 A 結果觀察到兩組的評估時間間隔或頻率呈現不平均的情況，為避免療效評估的偏差，建議執行的敏感度分析包括(1) 模擬研究 (simulation studies)」，及(2) 時間回推 (backdating)。

**(1) 模擬研究：**由電腦隨機產生兩組 PFS 時間分佈完全相同的試驗組及對照組的模擬數據樣品；另外，就試驗 A 結果觀察到的兩組評估時間間隔不平均的情況，於模擬的數據樣本中，亦模擬相同的訪診時間間隔的差距(例如對照組分別在第 42、84、126、168、210、及第 252 天訪診，試驗組則分別在第 44、86、128、170、212、及第 254 天訪診)，如此重覆模擬數千次甚至數萬次，如果大部分模擬樣本分析結果皆顯示錯誤的療效推論(兩組 PFS 時間具顯著差異)，則對試驗 A 結果的療效證據將有所質疑。

**(2) 回推時間：**此項敏感度分析係將發生腫瘤惡化的時間，回推至原先排定的訪診的時間，再重新分析，若結果呈現一致性的療效結果，則試驗 A 的療效結論可以接受，反之則不然。

## **(二) 非放射線診斷腫瘤惡化導致的評估偏差**

某些試驗將非放射線診斷亦納入腫瘤惡化的認定，此項認定無標準的定義，僅由試驗主持人根據臨床症狀主觀地來判定，也很難由病例報告書(Case Report Form, CRF)的紀錄得到驗證。尤其試驗主持人在非盲性作業下，若明顯地將其中一個治療組此項認定的時間提早，更容易產生療效評估偏差。

為避免可能的療效評估偏差，建議執行的敏感度分析包括(1) 將非經放射線診斷的惡化，不列為腫瘤惡化事件，再重新分析；(2) 僅考慮客觀認定且具文件支持的惡化事件或發生死亡才列為 PFS 事件，且若此等客觀認定的事件係在一次排定的訪診時間缺席，或未能作完整的評估之後，則 PFS 時間回推至那一次排定的訪診時間，再重新分析。



### (三) 缺失資料(missing data)導致的評估偏差

一般以 PFS 為主要療效指標的試驗，執行分析時，中途退出試驗者或無法追蹤評估者等缺失資料，皆視為設限資料(censored data)，且設限在最後評估為無惡化的時間點。然而如果設限資料所佔的比例太高，或者各治療組所佔的比例相差很大時，則受試者發生設限與病情惡化無關(亦即設限資料為隨機)的主要療效分析之關鍵性假設可能無法成立，宜視為與療效有關的設限資料，此等「具信息的設限資料(informative censoring data)」，亦可能造成療效評估的偏差。

為避免錯誤的療效推論，針對此項缺失資料可能造成的評估偏差，建議執行的敏感度分析包括(1) 時間回推；(2) 試驗組以較保守方法來插補缺失資料。

(1)時間回推：此項敏感度分析針對一次或一次以上的訪診缺席之後，所發生的惡化事件，將惡化時間回推至第一次訪診缺席的時間。

(2)保守方法插補缺失資料：此項敏感度分析，對於因訪診缺席所導致的缺失資料，在試驗組以較保守的方法，在對照組則以較寬鬆的方法來插補認定(例如在試驗組認定為發生 PFS 事件，在對照組則認定為設限資料)。

Dodd 等人在 2008 年發表的文章<sup>8</sup>裡提出另一類的「具信息的設限資料」所導致的療效評估的偏差。發生的原因主要由於試驗依據「獨立評估委員會(Independent Review Committee, IRC)」評估認定的 PFS 事件列為主要療效分析，惟受試者往往一經各試驗中心的試驗主持人認定為發生惡化(亦即 PFS 事件)後，即離開試驗開始接受其他治療；以致於雖然之後經 IRC 評估為未發生惡化，然而離開試驗則視為缺失資料，亦即設限於離開前的最後一個評估時間點，如此同樣會造成療效評估的偏差。另外此類同時具有 IRC 及試驗主持人評估結果的試驗，法規單位在審核仿單時往往有應根據那一個評估結果的困擾。

避免此項具信息的設限資料造成的療效評估偏差，建議的敏感度分析包括(1) 將試驗主持人評估為發生惡化而 IRC 評估為未發生且為設限資料的受試者，以試驗主持人評估的 PFS 時間取代 IRC 評估的 PFS 時間再重新分析；(2) 最壞情況方案：對於試驗組的受試者，以 IRC 及試驗主持人評估的 PFS 時間中最短者認定為該受試者的 PFS 時間；而對於對照組的受試者，則以 IRC 及試驗主持人評估的 PFS 時間中最長者認定為該受試者的 PFS 時間，再重新分析。



## 二、美國食品藥物管理局(US FDA)抗癌藥物臨床試驗適用的療效指標

美國食品藥物管理局抗癌藥物臨床試驗適用的療效指標規範中，列出以 PFS 作為主要療效指標以支持抗癌藥物的核准之優缺點如表一。規範中亦指出針對不同的癌症，PFS 在支持新藥核准上市所扮演的角色應有所不同，而 PFS 時間的延長是否可直接代表病人的臨床效益，或僅可視為替代性的臨床效益，端視試驗結果新藥物治療效果的強度，以及與現有治療方法風險效益(risk-benefit)比較的結果。而新藥查驗登記申請若依據腫瘤測量相關指標為唯一的療效證據，則通常須要第二個試驗來確認療效。較特別的是，該局認為試驗若採雙盲設計，由中央獨立評估委員會(IRC)來評估腫瘤相關測量指標並非絕對必要，除非某些不良反應造成相當程度的解盲，而大型、隨機臨床試驗，若治療效果深具強度，且敏感度分析顯示並未有觀察者(試驗主持人)評估偏差，亦可不須增加 IRC 的評估。

表一、以 PFS 為療效指標以支持抗癌藥物的核准的優缺點

優點	缺點
<ul style="list-style-type: none"> <li>● 與「整體存活時間 (Overall Survival, OS)」相比，試驗須要較少的樣本數，試驗追蹤的時間亦較短</li> <li>● 病情穩定(stable disease, SD)亦可視為治療有效</li> <li>● 療效評估不會受到病人因惡化轉換的後續治療的影響</li> <li>● 療效指標一般根據數值性量化的測量，評估相對客觀</li> </ul>	<ul style="list-style-type: none"> <li>● 多未經統計方法驗證可以作為 OS 的替代指標</li> <li>● 無法精確地測量；尤其是開放性試驗設計，容易有評估偏差</li> <li>● 不同試驗，PFS 療效指標的定義有所差別</li> <li>● 通常根據放射線影像或其他評估</li> <li>● 不同的治療組別療程，通常要求相同的評估時間點</li> </ul>

對於此等以 PFS 為主要療效指標的試驗設計，規範中要求將 PFS 的評估、測量、及分析方法，於試驗計畫書及「統計分析計畫(Statistical Analysis Plan, SAP)」中詳細地載明。計畫書對於發生腫瘤惡化事件的認定的標準亦應小心地定義，另外試驗組與對照組的訪診與放射線影像評估時間亦要對稱(symmetrical)相當，以避免系統性的評估偏差。若病人或試驗主持人亦參與腫瘤惡化事件的評估認定，則盲性作業格外重要。

對於試驗的分析，規範亦指出，訪診的時間點要適當，宜在所有排定的腫瘤評估完成之後。統計分析計畫宜預先設定主要療效分析，及包含一個或多個的敏感度分析，以評估試驗療效結果的堅實性(robustness)。另外，試驗亦須評估那些



長時間失去追蹤，之後發生死亡的病人對療效評估的影響，宜將此等病人視為發生 PFS 事件(原主要療效分析認定為設限資料)，執行敏感度分析。主要原因為此等病人若在兩組分佈不平均，且試驗組較對照組為多，將高估試驗組的 PFS 時間，導致療效評估偏差的緣故。至於認定的 PFS 時間，亦應與排定的訪診時間相差在短時間間隔範圍內；若病人經認定為發生惡化，惡化的時間點宜為該病人任何被評估為發生惡化的最早時間點，惟該評估之前不可有缺失資料；若病人經認定為設限資料，則設限的時間點宜為最後一次經放射線影像評估確定為未發生惡化的時間點。

### 三、PhRMA 及其他學術研究單位有關 PFS 的研究與建議

為避免 PFS 療效指標可能的評估偏差，一般建議由「盲性中央獨立評估委員會(Blinded Independent Central Review, BICR)」來執行 PFS 評估；然而如本文第二節所描述，Dodd 等人研究發現，由於臨床上病人經試驗主持人診斷為發生惡化後，通常會離開試驗接受其他處置，即便 BICR 評估認定未發生腫瘤惡化，惟後續的訪診皆為缺席，造成之後 BICR 評估的缺失資料，此為具信息的設限資料，同樣會造成療效評估的偏差。因此對於試驗設計採用 BICR 評估的必要性，PhRMA 工作小組以及其他學術研究單位紛紛參與研究與討論。2008 年布魯金斯癌症研究會議決議認為若試驗真正能達到雙盲，則試驗不須要採用 BICR 評估設計；惟須有一小部份子檔資料(subset) 送交 BICR 評估，以稽核各中心試驗主持人評估的正確性。

PhRMA 工作小組收集了 23 篇文獻以 BICR 評估腫瘤惡化的癌症臨床研究，整理分析後初步發現，在此等研究中 BICR 與試驗主持人，以風險比(Hazard Ratio, HR)估計的療效大小，在數值上皆相當一致。另外，分析結果亦顯示，將試驗組與對照組合併來看，整體 BICR 與試驗主持人評估的不一致，並無關試驗主持人或 BICR 評估療效的可信度；惟分別來看，若不一致的比率在試驗組與對照組有相當的差別，則將造成二者療效估計的不一致。針對以上研究的結果，該工作小組同樣亦建議並非每一試驗皆須採用 BICR 評估設計，惟試驗應設置一個中央資源資料庫，將所有受試者所有影像檢查結果皆掃描儲存於該資料庫，之後從該資料庫，以隨機抽樣的方式，抽取一小部份樣本來執行 BICR 評估。美國國家癌症研究所(National Cancer Institute, NCI)進一步建議：若試驗結果呈現很強的療效(風險比很小)，則只須要很小的樣本來執行 BICR 稽核(audit)；若療效強度為中等，則須要較大的樣本數來執行 BICR 稽核；而若試驗結果雖具統計顯著療效，惟療效很小，則全部試驗樣本皆須執行 BICR 評估，以確認療效。



### 結語：

總結而言，以「無惡化存活時間」為主要療效指標，以支持抗腫瘤藥物上市核准的晚期臨床試驗，近年有快速增加的趨勢。考量該指標特有的評估測量上的限制，廠商宜有嚴謹的試驗設計，並執行適當的敏感度分析，俾提供法規單位以快速、有效率的方法決定試驗結果是否具有潛在的評估偏差，並儘速驗證其可信度及臨床效益，將可提升抗癌新治療方法的快速發展。

### 參考文獻：

1. Kay A, Higgins J, Day AG, et al: Randomized controlled trials in the era of molecular oncology: Methodology, biomarkers, and endpoints. *Ann Oncol* [epub ahead of print on November 2, 2011]
2. Sridhara R, Johnson JR, Justice R, et al: Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *J Natl Cancer Inst* 102: 230-243, 2010
3. Amir E, Seruga B, Martinez-Lopez J, et al: Oncogenic targets, magnitude of benefit, and market pricing of antineoplastic drugs. *J Clin Oncol* 29: 2543-2549, 2011
4. Suman Bhattacharya, Gwen Fyfe, et al: Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trial. *J Clin Oncol* 27(35): 5958-5964, 2009.
5. US FDA. Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics.  
[http://www.fda.gov/cder/drug/cancer\\_endpoints/default.htm](http://www.fda.gov/cder/drug/cancer_endpoints/default.htm)
6. Christopher M. Booth and Elizabeth A. Eisenhauer. Progression-Free Survival: Meaningful or Simply Measurable? *J Clin Oncol* 30 (10): 1030-1033, 2012
7. Freidlin B, Korn EL, et al: Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol* 25: 2122-2126, 2007
8. Dodd LE, Korn EL, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 26: 2791-3796, 2008