生物醫學 2013年第6卷第1期：21-24

體外診斷醫療器材和實驗室研發之檢驗法的比對研究

曾嶔元[1,2]

[1]國泰綜合醫院病理暨檢驗醫學部，台北，台灣

[2]輔仁大學醫學系，台北，台灣

摘要

新的「體外診斷醫療器材（in vitro diagnostic device; IVD）」和新的「實驗室研發的檢驗法（laboratory developed test; LTD）」在運用於臨床之前，都必須先透過驗證（validate）或查證（verify）來證明其有效性（efficacy）。在這方面，「比對研究（comparative study）」是常用的方法之一。此方法看來很簡單，但充滿了陷阱。本文即是介紹「比對研究」的概念和方法，並指出在分析比對結果時，應如何評估整體一致性（overall agreement; O_A）、陽性一致性（positive agreement; P_A）和陰性一致性（negative agreement; N_A）。（生醫 2013;6(1):21-24）

關鍵字：比對研究（comparative study）、整體一致性（overall agreement; O_A）、陽性一致性（positive agreement; P_A）、陰性一致性（negative agreement; N_A）

前言

隨著科技的進步，體外診斷醫療器材（in vitro diagnostic device; IVD）不斷地推陳出新，以達到提高診斷正確性、簡化操作程序或降低成本等目的。同樣地，「實驗室研發的檢驗法（laboratory developed test; LTD）」也不斷地研究出新方法以改良現有的檢驗法。

當推出新的IVD時，研發者必須面對的挑戰是：「新的IVD和現有的IVD比起來如何？」類似的情況也會發生在檢驗方法的改變上。例如，當實驗室主管考慮是否採用較便宜或較簡單的檢驗法以取代現有的檢驗法時，一定會被問到的是：「新的方法和現行的方法比起來如何？」會有這類的質疑是因為，如果新的IVD與現有的IVD在性能上相當時，則新的IVD可考慮准予上市；如果新方法與舊方法在性能上相當，則

通訊作者：曾嶔元 教授
電話：886-2-2690-7965 ext 2518
傳真：886-2-2691-9800
地址：106 台北市仁愛路四段280號 病理暨檢驗醫學部
電子郵件：jeffbucknell@gmail.com

2013年1月5日來稿；2013年1月13日修改；2013年1月19日同意刊登

我們可為了得到新方法之其他優點（如便利性或低成本），而考慮汰舊換新。回答這類質疑的最佳方法當然就是以標準品來測試，看新、舊兩者何者較優。然而常碰到的困難是，沒有足夠數量的標準品來測試，甚至無從界定樣本是否合乎「黃金標準（gold standard）」。退而求其次，常用的替代方法就是「比對研究（comparative study）」了。簡單地說，比對研究就是用一群特定的樣本來比較新的「測試方法（test method）」和舊的「比對方法（comparative method）」，看看兩者在性能上是否相當。

比對的方法

進行比對時，要儘量避免蒐集標本所造成的偏差。用於測試的檢體必須與實際臨床使用者相似，譬如臨床上所用之檢體是血漬（blood spot），那麼測試時就不要用新鮮的血液檢體。此外，每件樣本所蒐集到的檢體量必須足夠，以應付未來可能發生需要複檢的情況。檢體在測試結束前都必須要穩定地貯存，以避免前後檢驗之間的偏差。執行比較時，需同時（或幾乎同時）執行兩種比對的方法，以避免檢體老化之差異所造成的誤差。若可能的話，將樣本分批測試，於10至20天內完成，每天進行比對。測試的剩餘樣本需保留下來，以便有爭議時能進一步探討原因。

不管是由新研發的「測試方法」或是由既有的「比對方法」所得到的結果，所有的數據都需記錄下來並立即加以分析，如此可及早偵測出系統或人為的錯誤。如果某些非一致的結果是因為如此而造成的，那麼這些數據可排除而不列入最後的分析。不過所有原因都需紀錄下來，否則非一致的結果都需保留在原始數據檔中。如果定性檢驗是由定量檢驗轉變而來，那麼可進一步探究定性結果之差異是否是因為所測之值接近閾值所造成的。

一致性之評估

當我們以新的「測試方法」和舊的「比對方法」檢驗一群樣本時，我們可得到如表一之結果。顯然，當B及C都等於零的時候，新、舊方法「看起來」是一致的。不過兩者一致並不表示兩者皆正確，因為當兩者都同時錯時，也是有一致性（agreement）的。當B及C不都是零的時候，則表示新、舊方法有不一致的地方。我們可由整體一致性（overall agreement; O_A）來評估兩者的一致性有多高，亦即兩者皆為陽性或陰性之樣本數佔全部樣本數的比例。根據表一，可得知O_A＝（A+D）÷（A+B+C+D）。

表一、新和舊方法之比對研究計算

	比對方法 得到陽性結果	比對方法 得到陰性結果	共計
測試方法 得到陽性結果	A	B	L
測試方法 得到陰性結果	C	D	M
共計	R	S	N

在評估一致性時，必須計算陽性一致性（positive agreement; P_A），亦即在比對方法得到陽性結果的樣本群中，有多少比例會在測試方法中出現陽性。我們可根據表一來計算，則可知 P_A＝A÷（A+C）。另外，我們還須同時評估陰性一致性

（negative agreement; N_A），亦即在比對方法得到陰性結果的樣本群中，有多少比例會在測試方法中出現陰性。根據表一來計算，則N_A＝D ÷（B+D）。由數學的觀點來看，O_A值顯然介於P_A值和N_A值之間，或者三者相等。兩種檢驗方法的陽性一致性高，不見得表示其陰性一致性也高；反之亦然。當測試的樣本數中R遠大於S時，O_A會因P_A高而跟著變高。所以蒐集樣本時宜避免這種偏差。

由以上所述我們可知，在評估時若可再計算陽性或陰性一致性，則可避免因部份樣本數太少而導致的誤判。我們舉一個例子來看，Virchows Arch第460期146頁中列出作者比較新的「IVD套組」與舊的「DNA定序法」，檢驗的項目是*KRAS*密碼子61突變。他們測試的陽性樣本數有7個，陰性樣本數有181個。結果「看起來」不錯，因為B=C=0，整體一致性的95%信賴區間為98%至100%。然而，當深入觀察陽性一致性後，發現其95%信賴區間為64.6%至100%，顯然不高。當B=C=1、A=6、D=180時，整體一致性的95%信賴區間為96.2%至99.7%，「看起來」也不錯，但是陽性一致性之95%信賴區間竟然只有48.7%至97.4%。由這個實例可知，樣本數太少會使O_A、P_A或N_A值達不到預定之標準。

一個常見的問題是，到底要測試多少樣本數才足夠評估呢？我們由表一來看，當表一的B=C=0的時候，由O_A的95%信賴區間來看，只要測試陽性樣本及陰性樣本數各31個，O_A就可達到85%了。如果採用O_A≥88%為標準，那麼測試陽性樣本及陰性樣本數則各需50個才能達到標準。由此看來，臨床與實驗室標準協會（Clinical and Laboratory Standards Institutes; CLSI）建議陽性樣本及陰性樣本數至少各需50個以上，似乎不是沒有道理。雖然從統計觀點來看，樣本數愈多愈好，但是樣本數太多會提高研發成本，甚至某些少見的疾病或亞型，幾乎根本不可能蒐集到50例。筆者建議，這時還是要蒐集至少29例陽性樣本，以免測試完畢還是無法證明兩種方法的一致性。當然，碰到困難決定的時候，最好還是先諮詢統計專家或評審委員。

對於IVD與新方法的研發者或評估者而言，最關心的是到底O_A值要多高，兩種檢驗方法才會被認為「相當一致（concordant）」？這是很值得探討的問題。有些IVD研發者自行認為 O_A ≥85%即可，不過正確的方法是計算（O_AN^2-U）÷（N^2-U）值，而其中U = RL + SM。此值若大於80%（有專家採用75%），則可認為兩種方法相當一致。

不一致時之解決方法

除非既有之比對方法是用於建立「黃金標準」的參考方法（reference standard），否則我們無從得知兩者不一致時，到底那個正確。甚至當兩者一致時，也不一定表示新的測試方法是正確的。所以，在作比對研究時，不應該由測試的數據來計算新方法或IVD的靈敏度與特異性。此外，由於比對方法（無論是舊的LDT或已上市的IVD）通常都是「非參考標準（non-reference standard）」，因此當新、舊兩者之結果不一致時，並不表示新的檢驗方法不如既有之比對方法。這時我們可針對這些不一致的個案作「分歧解決（discrepant resolution）」，亦即用參考方法或另一種「非參考標準」之方法，來分析看看哪一個方法

「對」或「比較對」。不過，「分歧解決」並非「比對研究」的必要步驟，除非是因「測試方法」比「比對方法」好太多或兩者各具不同優點而造成了顯著的不一致性 。但無論如何，我們都不可憑此來計算「測試方法」的靈敏度與特異性。

結語

由於科技的進步，檢驗方法不斷地推陳出新以減低成本或增加效能。若新、舊方法兩者各有千秋的話，那麼檢驗者就多了一項新的利器，因為兩者互補可提高檢驗的靈敏度與特異性。譬如，「新研發的測試方法」和「既有的測試方法」之靈敏度與特異性都是50%的話，那麼合併兩者後可使檢驗之靈敏度與特異性提高到75%。然而，不論是要證明新的IVD產品之有效性，還是要評估新的LDT是否值得採用，我們都必須合理地評估新舊兩者。評估一致性的時候不能只看整體一致性，同時也要查看陽性和陰性一致性，而且還要注意O_A、P_A和N_A值的95%信賴區間。比對時儘量增加陽性樣本數，以避免徒勞無功。總之，研發者和審核者都必須深入了解「比對研究」的含義，才不致誤判。