

分子檢驗的驗證方法

曾嶽元^{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

「實驗室自行開發的檢驗 (laboratory-developed test; LDT)」在用於病人之前，以及「體外診斷醫療器材 (in vitro diagnostic device; IVD)」在送查驗登記審查前，都必須經過驗證 (validation) 的程序。驗證所需的資料至少包括：精密度 (precision)、準確度 (accuracy)、報告區間 (reportable range)、參考值範圍 (reference range)、分析敏感性 (analytic sensitivity) 和分析特異性 (analytic specificity)。雖然不同的專業團體因定性或定量檢驗而對這些參數有稍微不同的定義，但在基本認知上是相當一致的。LDT和IVD研發者應真正地了解這些定義，才能正確地完成驗證程序。如此，病人才能受益於生醫科技。(生醫2013;6(1):30-35)

關鍵字：驗證 (validation)、精密度 (precision)、準確度 (accuracy)、報告區間 (reportable range)、參考值範圍 (reference range)、分析敏感性 (analytic sensitivity)、分析特異性 (analytic specificity)

前言

自從1978年以基因檢測來診斷產前鐮刀型貧血後¹，各種分子檢驗項目如雨後春筍般冒出，用途也從遺傳疾病擴展到傳染病、癌症、個人化醫療、體質分析等範圍。現在我們已經有超過二千項的分子檢驗可供臨床使用了。將「驗證過的檢驗方法 (validated laboratory test)」用於品質良好的檢體，如果能再加上良好的品管監控和實施矯正預防措

施，那麼病人即可充份得到醫學進步帶來的好處。

然而，分子檢驗管理的現況並不理想，和藥物的管理比起來，真是差太遠了。每個人都知道未經證實的藥物是不可用於臨床上的，現今也沒有醫療單位敢將自行開發的藥物給病人用，因為病人不是白老鼠。若以同樣的標準來看，目前通過美國食品藥物管理局 (Food and Drug Administration; FDA) 或台灣衛生署食品藥物管理局審查的分子檢驗套組 (test kit) 就非常少。換句話說，絕大多數的分子檢驗項

通訊作者：曾嶽元 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

目，沒有合法上市可用的「體外診斷醫療器材（in vitro diagnostic device; IVD）」，而必須使用實驗室自行開發的（home brews）方法。這些實驗室自行開發的檢驗（laboratory-developed test; LDT）在用於病人之前並未受到政府的監督。雖然LDT品質不如IVD，但如果LDT經過驗證（validation）後，也可謹慎地用於臨床上。此驗證過程雖不若IVD的驗證程序複雜，但也頗為費勁，因為需要提供的資料至少包括：精密度（precision）、準確度（accuracy）、報告區間（reportable range）、參考值範圍（reference range）、分析敏感性（analytic sensitivity）和分析特異性（analytic specificity）。以上六項參數簡稱為「PARRAsAs」。本文即對PARRAsAs作簡單介紹，協助各實驗室執行驗證，以提高分子檢驗的品質。

精密度

精密度指的是在一定的條件下，檢驗結果互相接近的程度²。因此，精密度評估的是「隨機誤差（random error）」。進一步說，精密度包括重複性（repeatability）與再現性（reproducibility）。前者指的是在相同條件下同一次檢驗中所得結果的一致性程度；後者指的是在不相同條件下（如不同操作者、不同次檢驗或不同實驗室間）所得結果的一致性程度。在精密度研究方面，包括實驗室內部精密度研究及再現性研究。用來測試的樣本數常為10至50個，對每個樣本檢測的訊號常常以平均值及變異部份來表示，其中「變異部份」包括標準差（standard deviation，簡稱SD）和變異係數（coefficient of variation，簡稱CV）。此外，也應該將高於和低於閾值的測量值之比例算出。

就定性的檢驗而言，高精密度指的是在操作者之間、使用不同批號的耗材、在不同的工作天都可到相近的結果。譬如，3位操作者以2台儀器和3個不同批號的試劑，在6星期中檢測8個樣本15次，即可得到大量的數據來評估定性檢驗的精密度。

就定量檢驗而言，所測試的分析物必須在報告區間（見下所述）內。在諸多精密度評估項目中，最重要的就是器材的精密度（within-device precision）了。精密度的數據要以標準差和變異係數百分比表示。通常至少要有20個工作天的數據才有辦法得到有意義的評估。進行分析時，每天要做兩次測試，每次使用二個樣本，每個樣本使用二種不同的濃度。此外，每次測試都需同時測試一個品管樣本。如果可能的話，應使用二種極不相同濃度的樣本，其中一個濃度最好接近「臨床決策值（medical decision level）」。將每日間（between-day）、每次間（between-run）、同次間（within-run）之標準差平方後相加再開平方，即可得到器材精密度的標準差。將此數值除以測試的濃度即可算出其變異係數百分比。

以「人類乳突病毒（Human papillomavirus; HPV）」檢驗的精密度為例，首先需要建立一組精密度系列（precision panel），內含10至20個樣本，每個樣本含有一定量的分析物及特定的HPV基因型。此系列是用來測試臨床決策點（medical decision point）及內含中等量分析物之檢體。

如果擬分析某特定基因型的HPV，但臨床案例少見，那麼可將HPV感染的細胞株和無HPV感染的「背

景」細胞株（如Jurkat細胞）混合，製造出仿造的臨床樣本。細胞株的使用，對液相細胞學檢體而言頗為重要，因為此法可協助我們了解從混雜的懸浮細胞中採樣以及在處理過程中所造成的改變。如果連這種細胞株也沒有，那麼可用HPV DNA質粒（plasmid）或RNA轉錄物（transcript）來製造出所用的測試系列。當使用人為製造的樣本時，需在測試中加入1例陰性的臨床樣本以及至少4例陽性的臨床樣本，這些臨床樣本應含有適當之分析物濃度以便測試臨床閾值。由於完整的分析過程需要大量的檢體，所以在分析前要先確保有足夠量的臨床檢體可應付測試，以免因臨床檢體無法隨意獲得，而造成測試的失敗。必要時可以把所有的臨床檢體倒在一起湊成足夠的檢體量。

仿造的精密度系列（simulated precision panel）必須包含至少2種不同基因型的HPV，配製3種不同病毒含量，因此至少有6種樣本。其中「高陰性（high negative）」樣本之分析物濃度低於臨床建立的閾值，而且在重覆受檢多次中，約95%的機會為陰性，5%的機會為陽性，此即「C5濃度」（例如在即時PCR檢測中，樣本之分析物濃度不低於臨床閾值的十分之一）。而「低陽性（low positive）」樣本所含的分析物濃度只是略高於臨床閾值，因而在反覆的檢測中，95%的機會可得到陽性結果，此即「C95濃度」。至於介乎其間的「中等陽性（moderate positive）」樣本所含的濃度可讓該樣本在反覆的檢測中，約100%可得到陽性結果（例如該樣本之濃度為臨床閾值的3倍）。

機構內的實驗室在做內部精密度研究時，並不要求高陰性、低陽性分別與C5及C95完全一致。如果

高陰性及低陽性樣本與閾值相當接近，以致標準差或變異係數在閾值附近約略恆定，那麼C5及C95就可用下列方法從實驗室內部精密度研究中計算出來。

$$C95 = C50 + 1.645 \times SD \text{ 或 } C50 \div (1 - 1.645 \times CV);$$
$$C5 = C50 - 1.645 \times SD \text{ 或 } C50 \div (1 + 1.645 \times CV)。$$

在測試實驗室（包括儀器及其他相關醫材）的內部精密度（within-laboratory precision）測試中，需要同時評估變異性（variability）的來源（如操作者、工作日、儀器、檢測本身等等），必須要監測至少12個工作天（可以不必連續日子），每天做2次檢測，每個檢體做雙重複檢驗。在精密度評估中，應使用3台不同的儀器以及3個不同批號的試劑來檢測。此外，在評估再現性時，可考慮（1）在3個不同的場所（如2個外面的實驗室，一個自己內部的實驗室）測試；（2）執行為期5天的評估，這包括一天至少做2次檢驗（除非在設計時，此檢驗套組已不允許一天做多次檢測），而且每次檢測使用重複樣本；（3）每天至少有2名操作者在每個指定的地方執行檢測；（4）使用前節所述的樣本系列（sample panel），包括C5及C95樣本；（5）測試時應從檢驗的第一個步驟（即核酸之萃取）開始。

準確度

對某一檢驗而言，準確度指的是其檢驗數值與真值（trueness）的接近程度^{3,4}；但就比對研究（comparative study）而言，一般認為準確度要達90%以上，才能被認定是足夠接近的。準確度所評估的是「整體分析誤差（total analytic error）」。而分

析研究而言，所謂「真值」就是某物真正的屬性（陽性、陰性）或其度量衡之數值。因此真值須以公認的標準方法和國際參考物（或已知的黃金標準）來定義。就某些定性的醫學檢驗而言，真值的定義有時還需要藉助影像學、病理學和臨床的追蹤才能獲得；甚至還得參考文獻的統計數字。不過這類所謂的真值並非真正的真值，而是在當時的時空環境下認為的真值。譬如說，在許多歐美的文獻中都報告A基因的突變率為5至10%，因此我們就接受此為真值；但A基因的突變率在台灣不見得就是5至10%，那怎麼辦？因此，在判定醫學檢驗之準確度時必須了解時空背景。的確，某些醫學參數的真值不像物理及化學的真值那麼清楚地可定義。

參考值範圍

精參考區間（reference interval）又謂「參考值範圍（reference range）」或「正常值（normal range）」，指的是在一特定的族群中某檢驗值的變動範圍⁵⁻⁷。若對該族群中「正常人」所取得的一系列樣本加以檢驗，即可建立參考值範圍。在此所謂的「正常人」即是根據既有的方法認為「無某特定疾病的人」，而落在95%區間之樣本檢測結果即是參考值範圍。

對於定性的檢驗而言，參考值範圍常是「陰性」、「正常」、「無此單株」等等字眼。例如癌症突變檢驗的參考值範圍即為「陰性」或「野生型」，因為正常人（無某種癌瘤者）不會有此突變。但要注意的是PCR的靈敏度可被推高到超出想像的地步，譬如慢性白血病人的融合基因「*bcr-abl*」，若以極為靈

敏的方法，則可在正常人測出*bcr-abl*的存在。因此，有時癌症突變檢驗的參考值範圍會以「無可偵測出突變」或「未驗出突變」等保守字眼表示。然而值得注意的是，某些檢驗是沒有參考值範圍的。譬如C型肝炎的基因型鑑定即無參考值範圍，因為此檢驗中所謂的「正常人」是指無C型肝炎感染的人。有時定性檢驗是建立在定量測量上，但只要檢驗報告為描述性的，該項檢驗的參考值範圍就不該以數字表達。

報告區間

簡單地說，醫學檢驗之報告區間（reportable range）即是在報告書上有可能會呈現的方式，不管它是定量檢驗的數據或定性檢驗的描述。不過要注意的是報告區間的所有誤差都必須在可接受的範圍。亦即，不管此誤差是源自於非線性（nonlinearity）、不精密度（imprecision）或其他原因，產生報告區間內檢測值之誤差（含技術與儀器）是可被接受的。

定性檢驗的報告區間顯然比定量檢驗簡單得多，不外是「陰性或陽性」、「正常或異常」、「可測出或未測出」、「同型合子或異型合子」、「A型、B型」等等，甚至於還可包括「未定」等字眼。有時會把定量測量改為定性檢驗，目的是讓閱讀報告書的人明瞭該檢驗的臨床實用性。這時該項檢驗的報告區間就不應以數字表達。

分析敏感性

一般而言，偵測分析物(analyte)之能力謂之「分析敏感性」，亦即分析物之偵測值的下限（lower

limit)。然而，不同專業團體對分析敏感性的定義不盡相同，因此，分析敏感性的定義有時變得有點模糊。美國醫學遺傳學院（American College of Medical Genetics; ACMG）和國際標準化組織（International Organization for Standardization; ISO）⁸對此名詞的定義是「檢驗為陽性的生物樣本中有多少比例為正確的」。就比對研究而言，指的就是「陽性一致性（positive agreement）」；對於癌症突變分析的檢驗而言（如大腸癌的*KRAS*），指的是檢驗為陽性的檢體中癌組織的最低比例；對於病理檢驗的「免疫組織化學染色法（immunohistochemistry）」而言，指的是偵測微量抗原的能力；而對於病原體之偵測而言，指的則是偵測微量微生物的能力。

以上這些定義適用於定性檢驗，但對大多數的定量檢驗就不適用了。就定量檢驗而言，分析敏感性指的是，分析物可被偵測到的最低濃度，因此和「偵測極限（limit of detection; LoD）」的意思相同。所謂「LoD」即分析物可被測出的最低濃度，又稱檢測下限（lower limit of detection）。亦即，分析物在某種濃度時可使95%的檢驗都可得到陽性結果，該濃度即為LoD。美國FDA雖然沒有對分析敏感性下定義，但根據審核通過的案件來看，不管是定性檢驗或定量檢驗，FDA所採用的定義是LoD。

首先，我們應先決定訊號的閾值，亦即「背景值上限（limit of blank; LoB）」。因此，當病人的檢體在檢測時出現訊號，就表示分析物已被偵測到。此外，也應定出LoD，以確認分析物在何種濃度下可導致95%的檢測都可得到陽性結果。

如果命中率（即分析物能被偵測到的百分比）是在可驗出之濃度中（0%至100%）測試，那麼LoD也須要用命中率來估計。在LoD研究中，我們可使用一系列的稀釋濃度來定出醫材的LoD，稀釋時應以樣本收集緩衝液來逐步稀釋分析物。如果實驗室宣稱可檢驗某分析物之各種亞型，那麼就應為每一種亞型及樣本收集液定出LoD。統計時，可使用「概率單位（probit）」分析。若要確定LoD值，應準備至少20個重覆檢體，濃度為LoD，然後證明95%的檢體可驗出有分析物。在訂定LoD時，應使用2至3個不同批號的醫材，在3至5天中檢測3到5個檢體，以評估變異性之來源。

要注意的是，LoB是定義分析物是否存在，而臨床閾值則定義分析物檢測為陽性或陰性，因此閾值可以高於LoB。如果把LoB當作臨床閾值，那麼C95濃度就和LoD一樣；但如果某一檢測之臨床閾值高於LoB，那麼C95就可能與LoD不同了。而如果把LoB視為臨床閾值，那麼C95濃度即和LoD同義，而濃度為0（即無分析物之檢體）時則為C5。

以愛滋病檢驗為例，若某IVD以核酸檢驗法（nucleic acid testing; NAT）檢測人類免疫缺乏病毒（Human Immunodeficiency Virus; HIV），篩檢不同捐贈者聚集之血液及血漿是否帶有HIV-1病毒，而該研發者宣稱其方法可偵測HIV-2，那麼驗證時還須使用包括HIV-2的樣本。這些試驗必須證明核酸檢驗法在聚集的檢體中可偵測100個RNA拷貝/毫升或在原始的捐贈血液中可偵測5000個拷貝/毫升，且有95%的偵測率。

分析特異性

就定性檢驗或半定量檢驗而言，美國「臨床與實驗室標準協會（Clinical and Laboratory Standards Institute; CLSI）」認為「分析特異性」指的是，和參考方法（reference method）比起來該檢驗方法偵測陰性結果的能力。亦即比對研究中的「陰性一致性（negative agreement）」。這和「美國醫學遺傳學院」的看法相似，因為該機構對分析特異性的定義是「檢驗為陰性的生物樣本中有多少比例為正確的」。對於病理檢驗的免疫組織化學而言，分析特異性指的是對特定的組織或細胞出現預期的陰性染色。就定量檢驗而言，CLSI對此名詞的定義則是「在干擾物的存在下，某檢驗方法可測出分析物之能力」。此含意著重在評估交叉反應的可能性。美國FDA所採用的定義較接近CLSI對「定量檢驗之分析特異性」所下的定義。

結語

我們若使用合法上市的IVD檢驗套組，那麼就不必再驗證，因為IVD在查驗登記時，食品藥物管理局已經對驗證把關了。可惜目前只有少數項目的分子檢驗套組通過美國FDA和台灣衛生署食品藥物管理局的審查。由於許多臨床上需要使用的分子檢驗項目並無合法的IVD可用，因此分子醫學實驗室常常需要使用LDT，以應付臨床上的需求。然而由於分子檢驗的專業人員大多來自做實驗的研究者，這些研究人員通常對檢驗的品管不熟悉，因此不太可能會投入時間和精力來獲得PARRAsAs資料。可想而知，目前大多數分子檢驗的LDT並未經過驗證。由於驗證的成效之一是

減少檢驗錯誤的機率，因此使用未驗證過的LTD必然會出現較高的錯誤率。但願本文能為致力於分子檢驗之驗證的實驗室，提供簡單的介紹。

參考文獻

1. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc Natl Acad Sci U S A. 1978;75:5631-5635.
2. International Organization for Standardization. International Vocabulary of Basic and General Terms in Metrology (VIM). Geneva, Switzerland: International Organization for Standardization. 2004.
3. Isler JA, Vesterqvist OE, Burczynski ME. Analytical validation of genotyping assays in the biomarker laboratory. Pharmacogenomics 2007;8:353-368.
4. Naito HK. The need for accurate total cholesterol measurement. Recommended analytical goals, current state of reliability, and guidelines for better determinations. Clin Lab Med 1989;9:37-60.
5. Gräsbeck R. The evolution of the reference value concept. Clin Chem Lab Med 2004;42:692-697.
6. Sunderman FW, Jr. Current concepts of "normal values," "reference values," and "discrimination values," in clinical chemistry. Clin Chem 1975;21:1873-1877.
7. Tadano J, Niwa M, Saito H, Sato T. The current concepts of "normal values" and "clinical reference values" in the clinical laboratory (trials for determining clinical reference values of the Tokai University Hospital). Tokai J Exp Clin Med 1980;5:251-262.
8. International Organization for Standardization. ISO/IEC Guide 2: Standardization and Related Activities-General Vocabulary. Geneva, Switzerland: International Organization for Standardization. 2004.