

IVD開發流程

USDA指引

概念化

Conceptualization

開發與優化

Development & Optimization

驗證

Verification

批次發布

Serial release

監控

Monitor

OIE手冊

初步考量

Preliminary consideration

開發與優化

Development & Optimization

確效

Validation

監控

Monitor

在開發早期確認問題

ex: 定義預期用途、樣本與檢定類型選擇

評估與確定條件

ex: 設定檢定範圍、建立標準品、標準化與優化檢測流程

基本特性演示與常規使用的表現

ex: 分析靈敏度與特異性、診斷靈敏度與特異性、重複性、再現性

確認每批次之效價

ex: 建立檢測平台 (樣本庫)、建立發布規格 (格式與效價)

監控已發布批次之效價

ex: 定期監控計畫

確效

Validation

確效 (validation)

- 目的：
證明檢定 (assay) 之科學合理性、可靠性、再現性，且符合預期用途
- 內容：
完成檢定的開發、優化、標準化及基本特性的演示，並評估常規使用下的檢定表現
完成上述內容可被視為已針對檢定預期用途進行確效
→ 只要按照優化的步驟操作，該檢定結果視為有效

參考資料

800.73_Diagnostic Test Kit Validation (診斷試劑確效)

800.112_Validation of In Vitro Potency Assays (體外試劑確效)

CHAPTER 1.1.6_Principle and Methods of Validation of Diagnostic Assays for Infectious Disease (診斷試劑確效原則與方法)

- 概念化
- 開發與優化
- 驗證
- 實施
- 監控

概念化 (1)

Conceptualization

- 目的：提出檢定與定義其**預期用途**，並預設可能影響檢定表現的各種變量
- 定義預期用途：應定義**目標物種、目標病原 (或狀態) 以及採樣基質**
常見用途包含
 - ✓ 確認一地為非疫區
 - ✓ 確認單一個體無感染或運輸之產品無病原
 - ✓ 協助特定族群清淨疾病或消除感染
 - ✓ 確認疑似或臨床病例之診斷 (包含篩選性檢定結果之確認)
 - ✓ 評估感染或暴露的盛行率或分析風險 (調查、族群健康狀態、疾病控制措施)
 - ✓ 確認免疫狀態 (免疫後)
- 考量檢定能力：
 - ✓ 檢定類型
 - ✓ 檢定系統的物理學、化學、生物學、操作者等因子
 - ✓ **反應與目標物種效力間的關係**
 - ✓ 在**預期濃度範圍內準確測定**分析物或參數的能力

概念化 (2)

Conceptualization

- 樣本：
 - ✓ 詳實記錄樣本資料
 - 病原
株、血清型、基因型、系譜
 - 動物
物種、品系、年齡、性別、繁殖狀態、免疫歷史、獸群病史
 - 疾病階段
臨床症狀、抗體譜 (profile)、病原量 (pathogen load)、排毒
 - 實驗樣本應記錄實驗過程
 - 檢測疾病狀態的試驗結果

概念化 (3)

Conceptualization

- 樣本：
 - ✓ 型態
 - 類型
血液、乳清、鼻腔拭子、尿液、糞便
 - 獨立或合併樣本
獨立樣本來自單一動物且具有代表性 (典型感染進程或階段)
合併樣本來自相似感染階段的 n 個動物，由1個測試為陽性的檢體與 $n-1$ 個檢測為陰性的檢體各取等量混和而成，應於混和前獨立檢測各獨立樣本確認分析物濃度與活性相似，混和後再次確認不會因為混和改變檢測表現
 - 群體樣本：
 - ◆ 從所有群體成員主動收集 (如一槽牛奶樣本)
 - ◆ 從不明確的成員組成被動收集 (例如由繩索獲得口腔粘液)
 - 加樣樣本
加樣已知分析物或強陽性樣本於陰性樣本中，配置涵蓋弱到強陽性的範圍 (典型感染進程)，稀釋用陰性樣本作為背景監控
取得橫跨預期範圍的樣本困難，初期感染階段常見抗體的分型 (isotype) 與親和性 (avidity) 差異較大，一般不適合作為評估分析表現的參考樣本，而中間範圍通常只有短暫出現
 - 樣本組的建立
隨機篩選一組含陽性與陰性的檢體，詳述樣本組的選擇，並提供推定盛行率

概念化 (4)

Conceptualization

- 樣本：
 - ✓ 標準樣本：
 - 必須穩定且無害，並應進行廣泛的特徵分析，其特徵、製備、儲存有刊登在期刊中為佳
 - 包含陰性、弱陽性、強陽性，已知濃度或活性，且在操作範圍內
 - 用於建立檢定表現的上下限、監控隨機與系統性變異 (以不同的控制圖表)
 - ◆ 國際與國家參考標準品
 - 由OIE或其他國際參考實驗室提供「國際參考標準品」含有特定量的分析物，所有試劑都應以此為標準
 - 由國家參考實驗室提供 (盡可能) 根據「國際參考標準品」校正的「國家參考標準品」
 - ◆ 內部標準品 (in-house standard)
 - 通常需根據國際與國家參考標準品校正
 - 若不可行則盡可能依照國際與國家參考標準品的標準進行特徵分析
 - 保存足夠的分裝管用於工作標準品之校正
 - ◆ 工作標準品 (working standard)
 - 又稱分析物或步驟對照
 - 根據國際、國家或內部參考標準品校正
 - 由地方 (目標族群) 選擇或採用地方基質配製
 - 保存大量分裝管用於常規使用的每次操作，盡量模擬田間樣本與常規使用樣本

概念化 (4)

Conceptualization

- 樣本：
 - ✓ 準備
 - 更換樣本可能導致不可控變因並破壞連續性，因此準備大體積的分裝管長期保存做單一使用，數量由參與的實驗室數量與能力測試頻率決定
 - 需要散裝儲存樣本，不時配製工作分裝管，若儲存空間有限建議一次性配製大量分裝管，以減少大體積樣本多次分裝之凍融過程造成降解
 - 各樣本應分裝成多個小管，用於各個重複
 - 較小的動物需要增加動物數量，以確保有足夠體積的參考樣本
 - ✓ 其他考量
 - 干擾物質或交叉反應物質的潛在影響 (內源性或外源性，如：抑制劑、樣本汙染樣本變質)
 - 試劑的可取得性
 - 樣本的選擇、收集、處理、保存、管理、運輸、監管鏈 (chain of custody)、樣品追蹤和實驗室信息管理系統

開發與優化 (1)

Development & Optimization

- 試驗方法設計與概念驗證，由以下三者決定
 - ✓ 含目標分析物的參考樣本 (用於開發至確效的實驗中)
 - ✓ 含目標分析物的基質
 - ✓ 目標族群
- 操作範圍 (operating range)
 - ✓ 可獲得適當的準確性與精確性的分析物濃度或力價的上限與下限
 - ✓ 試驗方法：準備強陽性參考樣本最好與優化用樣本相同
在陰性基質中序列稀釋至反應消失，並繪成反應曲線圖 (濃度-反應關係圖) 建立工作範圍
 - ✓ 若所得操作範圍不適用於預期用途，應再行優化
 - ✓ 典型的校正曲線多為乙狀函數，應轉化為近似線性關係並提供算法
- 優化
 - ✓ 評估並調整重要的物理學、化學與生物學參數，以符合預期用途
 - ✓ 樣本選擇
應選擇至少3個明確定義的參考樣本 (強陽性、弱陽性、陰性)
可代表已知感染與無感染的目標族群個體
參考樣本事先以1種 (建議2種) 替代檢測方法確認
建議準備足夠量的參考樣本於分裝管用於每一次操作，更改參考樣本將對開發與確效步驟的完整性造成影響
 - ✓ 若無法取得則使用加樣樣本，但無法真正代表的基質-病原交互反應
加樣已知量分析物、培養得標的與陽性樣本於陰性樣本中
務必盡量與真實樣本接近
 - ✓ 確認檢定能力：區別分析物含量、區別分析物近似物、優化試劑濃度並完善檢測流程

開發與優化 (2)

Development & Optimization

- 標準化
 - ✓ 材料：需考慮保質期與儲藏條件
 - 化學品
所有化學品、緩衝液配方、試劑皆須描述其純度與等級 (包含水)，並建立可接受的工作範圍與紀錄參數 (如pH、體積莫爾濃度)
 - 生物製品
定義品質、純度、濃度及反應性之標準
確認萃取或抗原洗脫 (elution) 條件
 - ✓ 反應時間與溫度範圍
 - ✓ 必須儀器之操作規範與校正
 - ✓ 資料收集、處理、解讀
 - ✓ 上述所有參數都應於優化後描述於protocol
 - ✓ 紀錄對表現有關鍵影響的步驟
- 基質中的抑制因子
 - ✓ 針對不同基質進行確效
 - ✓ 血清 (尤其是溶血時) 可能含有細胞毒性因子，不利於病毒中和檢定
內源性物質可能影響配體結合與基於酵素的檢定如ELISA
糞便、自溶組織及精液常有更多干擾物質

開發與優化 (3)

Development & Optimization

- 堅固性 (ruggedness) :
 - ✓ 定義：在單一實驗室內的測試情況的微小變化不會影響檢定的能力
 - ✓ 方法：在建立最佳檢測條件後，藉由條件變化對結果的影響來評估堅固性
常見變化：pH、感作時間、感作溫度、材料廠牌、操作員及試劑批號
應考慮檢定的結構和設計的系統可能產生的影響，如：ELISA盤的位置影響
 - ✓ 若微小變化就會造成不可接受的結果，即代表堅固性差
→ 早期得知時，應決定是否繼續確效 (再現性會更差)
- 常態化 (normalization)
 - ✓ 目的
檢定的操作間與實驗室間的 (半) 定量原始數據難以比較
每次操作使用1或多個工作參考品，可增加試驗結果的可比性
 - ✓ 方法
將原始試驗數值轉化成相對於工作標準品的單位
如：陽性對照的百分比與由標準曲線估計濃度或力價
- 初期重複性試驗
 - ✓ 進行一個或多個實驗性盲測，早期得知重複性差時，應決定是否繼續確效
 - ✓ 樣本
至少3個 (5個更好) 參考樣本，代表陰性、弱陽性、強陽性 (自然感染或配製)

驗證

Verification

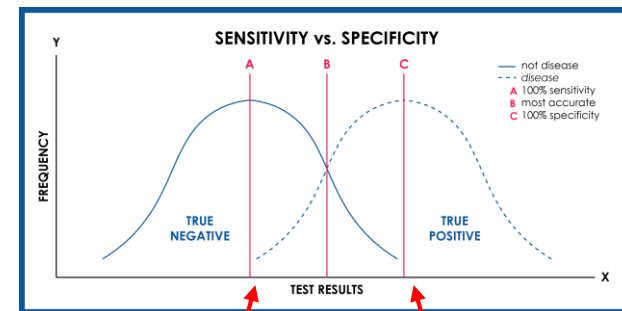
1. 測量尺度與截止值
2. 分析表現
 - a. 重複性
 - b. 分析靈敏度
 - c. 分析特異性
3. 診斷表現
 - a. 樣本選擇
 - b. 診斷靈敏度與特異性
 - ① 有完美參考標準_單一試驗
 - ② 有完美參考標準_比較試驗
 - ③ 無完美參考標準_單一試驗
 - c. 臨時檢定識別
4. 再現性

測量尺度與截止值

Scale of Measure & Cut-off

- 測量尺度 (scale of measure)
 - ✓ 二分法 (binary)：試驗結果只有陽性與陰性
 - ✓ 次序 (ordinal)：有多個等級，等級高低與分析物含量有關 (如：病毒中合力價)
 - ✓ 連續 (continuous)：理論上可能有無數個測量值 (如：ELISA OD值與RT-PCR Ct值)
 - 信賴區間 (confidence interval, CI)：建議DSe與DSp之CI為95%

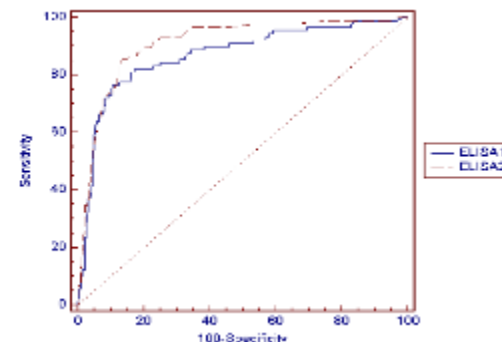
- 截止值 (cut-off, threshold)
 - ✓ 次序與連續必須以截止值區分結果為陽性與陰性才可計算DSe與DSp
 - ✓ 可插入1-2個截止值 (實驗結果分為2-3個分類)
 - 若已知陽性與陰性有重疊，則無法僅插入1個截止值
 - 該情形下2個截止值可以有高DSe與DSp
 - ✓ 截止值可能隨多個原因改變：
 - 預期用途 (篩選性或確認性)、感染盛行率、試驗錯誤的成本、其他試驗的可取得性



篩選性檢定
DSe ↑ DSp ↓

確認性檢定
DSe ↓ DSp ↑

- ROC (receiver-operating characteristic) 分析
 - ✓ 次序與連續DSe與DSp的估計必須有ROC曲線下的面積作為補充
 - 比較多個試驗時，應計算95% CI下ROC分析的差異
 - ✓ 提供截止值之外，可評估整體準確性的方法，呈現為單一數字：0.5 - 1 (無用 - 完美)



分析表現_重複性 (1)

Analytical Performance_Repeatability

- 基本介紹
 - ✓ 定義
在**單一實驗室**以**單一試驗方法**檢測**單一樣本**的結果的在**操作內**與**操作間** (within and between runs) 的一致性 (agreement) , 由重複 (replicates) 間的變異估計
 - ✓ 樣本
需要**至少3個樣本**
 - ✓ 執行
依照檢定的**所有步驟**操作 , 並且**各重複不可混和稀釋**
至少3重複 , 5重複更佳
操作內變異 : 須由**1個操作者** , 以**1個批次**進行
操作間變異 : 須由**2個以上的操作者**於**多天**進行操作

分析表現_重複性 (2)

Analytical Performance_Repeatability

- 連續結果
 - ✓ 用於連續測量尺度時，又稱**精確性** (precision)
 - ✓ 計算一組樣本的重複的**平均值**與**標準差** (standard deviation, SD)，並呈現於散布圖上
 - ✓ 無論SD與平均值是否有正比關係，都應計算計算**變異係數** (coefficient of variation, $CV = SD/\text{平均值}$)：顯示離散程度
應計算CV的不確定度 (如95% CI)
 - 若CV一致，計算所有樣本的95% CI
 - 若CV隨分析物濃度改變，應計算各濃度下的95% CI
(平均值與SD成正比時，通常在分析物濃度低時CV較大)
 - ✓ 若實驗設計包含**多因子評估** (如：不同操作者或操作日期)，應使用其他方法如**隨機效應模型** (variance component model)，以分解變異為多個可以判讀的變異的總和
隨機效應模型也可用於重複性測試

分析表現_重複性 (3)

Analytical Performance_Repeatability

二分法結果

- ✓ 若有定量結果，應以定量結果分析精確性
- ✓ 二分法試驗以**kappa**分析一致性
假設排除偶然的一致性 (agreement beyond chance)
1 (完美排除偶然的一致性)
0 (不排除偶然的一致性)
- ✓ 離截止值越遠的試驗結果，其一致性越佳，故應測試中間/疑似值，以避免高估一致性
- ✓ 加權**kappa**分析可用於次序結果 (如陽性/疑似/陰性)，並發現大的差異 (如2個分類的差異) 較小的差異 (1個分類的差異) 嚴重
- ✓ 應提供kappa分析**95% CI**

Example 1: Kappa calculation based on repeated test results classified as positive or negative

<i>Test result</i>	<i>Positive</i>	<i>Negative</i>
Positive	90	5
Negative	10	95
	100	100

Kappa = 0.85 (95% CI = 0.78 to 0.92)

Example 2: Kappa calculation based on repeated test results classified into three categories (positive, suspicious, or negative)

<i>Test result</i>	<i>Positive</i>	<i>Suspicious</i>	<i>Negative</i>
Positive	80	10	10
Suspicious	15	75	10
Negative	5	15	80
	100	100	100

Kappa = 0.68 (95% CI = 0.61 to 0.75). Weighted kappa = 0.70. (95% CI = 0.61 to 0.79)

分析表現_分析靈敏度 (1)

Analytical Performance_Analytical Sensitivity, ASe

- 偵測極限 (limit of detection, LOD)
 - ✓ 定義：可與背景區別的最少分析物量 (不必定量)
 - ✓ 試驗方法
 - **稀釋至消失 (dilution-to-extinction, DTE) 實驗**
由已知定量的目標分析物在適當的樣本基質 (matrix) 中**序列稀釋後進行檢測**
結果必須分類為陽性與陰性，重複越多越精準
 - 樣本
 - 單位內部或國家/國際參考標準品
 - 分析物已定量之田間樣本
 - 若可能，由**最少5隻動物**收集橫跨整個感染進程的系列樣本 (包含**由感染初期到發展臨床症狀與疾病爆發**)，但不易取得偵測低水平分析物或次臨床感染用的樣本
 - 若採樣是致命的 (採臟器)，減少至最小量，每次採5隻
 - 比較ASe：併行操作候選檢定與其他檢定，檢測自然或實驗感染動物樣本
提供ASe的比較與最早偵測到疾病的時間點比較、自然感染的濃度範圍

分析表現_分析靈敏度 (2)

Analytical Performance_Analytical Sensitivity, ASe

- 偵測極限 (limit of detection, LOD)
 - ✓ 結果判讀
 - 保守估計 (conservative estimate)
選擇一個稀釋倍率序列稀釋 (如10倍) 後進行檢測，結果為100%陽性的最高稀釋倍率
 - 精確估計
再以更低的稀釋倍率 (如2倍) 針對前一實驗0~100%陽性的範圍進行試驗計算
LOD終點一般設為95% (如：19/20)，也可為其他值 (如50%) 但應固定LOD終點，以便比較多個試驗的結果
 - 分析方法
 - Spearman-Kärber 非參數方法
 - 邏輯回歸 (logistic regression)
 - 概率分析 (probit analysis)

分析表現_分析靈敏度 (3)

Analytical Performance_Analytical Sensitivity, ASe

範例1：

10^8 CFU細菌加樣 (spike) 於10 g糞便得 10^7 CFU/g樣本

→ 10倍序列稀釋，進行3重複實驗

→ 結果於 10^3 CFU/g皆可被偵測而 10^2 CFU/g皆未被偵測 → 保守估計LOD為 10^3 CFU/g

→ 更精確的估計再以其他稀釋倍率 (如2倍) 進行，應包含前一實驗0~100%陽性的範圍

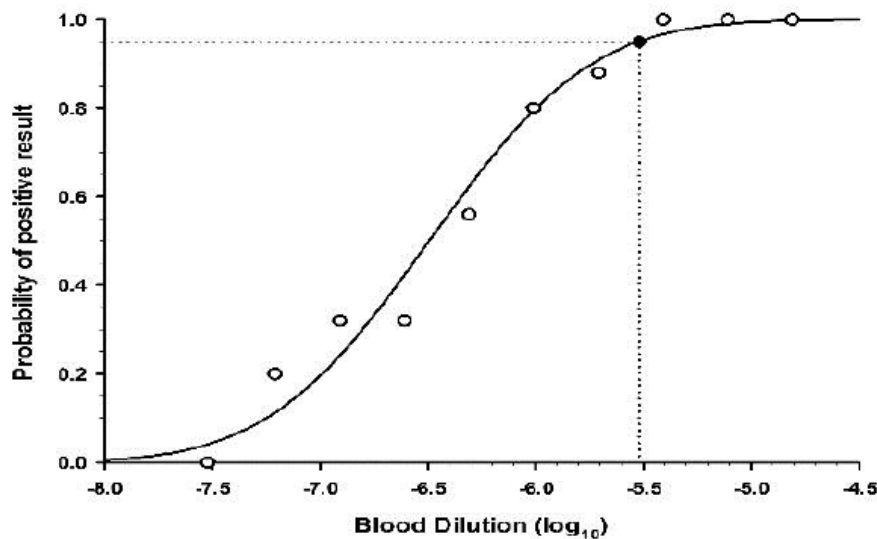
範例2：

非洲馬疫病毒陽性血清 (10^{-3} 稀釋) 進行2倍序列稀釋 (涵蓋檢定的非線性範圍)

重複25次萃取進行qPCR，15個稀釋倍率用於概率分析

95% LOD : 3.02×10^{-6}

對應CT值為35.71



分析表現_分析靈敏度 (4)

Analytical Performance_Analytical Sensitivity, ASe

- 定量極限
在可接受的診斷準確度和精確度之下，可進行定量的分析物的最低和最高濃度
- 信噪比 (signal to background)
評估分析物信號和空白試劑信號的比率

分析表現_分析特異性 (1)

Analytical Performance_Analytical Specificity, ASp

- 定義：從非目標分析物中區別目標分析物的能力
- **選擇性 (selectivity)**
從以下情形中準確定量目標分析物的程度，干擾物可能降低或提高反應
 - ✓ 干擾物 (如基質中的酵素抑制劑)
 - ✓ 降解物 (如毒性因子)
 - ✓ 反應物與固向的非特異性結合 (如ELISA接合物吸附於96孔盤)
 - ✓ **免疫抗體** (可能主動感染之抗體混淆) ex: **DIVA**評估安慰劑與摻有潛在干擾物質的製劑的反應曲線
→理想中，上述製劑**無明顯的劑量反應**，且**背景雜訊細微**
 - ✓ 樣本：
 - 每組至少5個檢體
 - 避免以免疫抗原作為檢定標的，因載體蛋白可能引起非特異性抗體反應，造成偽陽性
 - 未感染/未免疫
 - 未感染/有免疫
 - 有感染/未免疫
 - 有感染/有免疫

分析表現_分析特異性 (2)

Analytical Performance_Analytical Specificity, ASp

- **排他性 (exclusivity)**

檢測目標病原獨特的分析物，並排除其他潛在交叉反應的病原的能力

取得可能有交叉反應的參考樣本，應考慮適用於目標分析物與預期用途的每個譜系、分離株、種或屬

「發生交叉反應的物質」於「所有檢測的潛在交叉反應物質」中的百分比

建立**交叉反應譜**，並隨著潛在交叉反應病原的出現持續擴充

確認性試驗需要更具排他性

- **包容性 (inclusivity)**

檢定檢測到的該物種的一或多個株或血清型、屬內的多個物種、相似且密切關聯的病原或抗體分組的百分比

應考慮適用於目標分析物與預期用途的每個譜系、分離株、種或屬

篩選性試驗需要更具包容性

偵測範圍→適用性

ex: 可用於篩選所有已知基因型或血清型的病毒檢定，其參考樣本須包含所有具代表性的型別，須持續更新新出現的系譜或血清型變異

診斷表現_樣本選擇

Diagnostic Performance_Sample Selection

- 參考動物族群：動物應具有代表性
 - ✓ 目標族群
主要分類：感染標的、暴露於標的、從未感染與暴露於標的
另須注意的條件包含：物種、年齡、性別、品系、感染階段、疫苗歷史、相關的族群疾病歷史
 - ✓ 陰性參考樣本
真陰性樣本 (無可能感染與暴露) 可能難以取得
→ 可從已清淨或未發現該標的的國家獲得樣本，只要樣本來源族群與目標族群相似
 - ✓ 陽性參考樣本
由病原分離確認為陽性樣本較難達到需要的樣本數
以其他準確度高的參考試驗確認為陽性樣本 (如核酸偵測)，建立2×2表格 (黃金標準模型，假設參考試驗為完美)
- 使用實驗感染或免疫動物樣本可了解抗體反應情形，但非典型的自然感染
→ 若必須用實驗動物樣本進行確效，則所得診斷靈敏度與特異性 (DSe & DSp) 應被視為低於真實DSe與DSp

診斷表現_診斷靈敏度與特異性 (1)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)

有完美參考標準_單一試驗

- 樣本組的建立：
隨機篩選一組含陽性與陰性的檢體，詳述樣本組的選擇，並提供推定盛行率
- 靈敏度與特異性的評估：
依照各檢體類型與宿主，確立診斷的靈敏度與特異
建構的樣本組與預期目標群體可能有顯著的差異，包含檢體性質 (如疾病的階段) 與使用目的 (如篩選或確認性診斷)
→產品文獻附上樣本組的描述
- 目測與機器檢測：
由肉眼觀察試驗結果的檢定，使用目測分類方法評估靈敏度與特異性透過儀器 (例如光密度檢測儀) 測量，來協助篩選批次發布的平台樣本，並建立儀器結果與肉眼觀察結果之間的關係
進行目測分類結果時，不能先得知儀器結果

診斷表現_診斷靈敏度與特異性 (2)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)

有完美參考標準_單一試驗

- 陽性與陰性樣本數量取決於檢定的DSe與DSp的可能值 (likely value) 與信心水準 (confidence level) (Jacobson, 1998)

Estimated DSe or DSp	2% error allowed in estimate of DSe and DSp			5% error allowed in estimate of DSe and DSp		
	Confidence			Confidence		
	90%	95%	99%	90%	95%	99%
90%	610	864	1493	98	138	239
92%	466	707	1221	75	113	195
94%	382	542	935	61	87	150
95%	372	456	788	60	73	126
96%	260	369	637	42	59	102
97%	197	279	483	32	45	77
98%	133	188	325	21	30	52
99%	67	95	164	11	15	26

範例：

- 97% DSe, 95%信心水準, 2%誤差 (95-99%)
→ 需要279個陽性
- 99% DSp, 95%信心水準, 2%誤差 (97-99.9%)
→ 需要95個陰性

		Number of reference samples required*	
		Known positive (279)	Known negative (95)
Test results	Positive	270 TP	7 FP
	Negative	9 FN	88 TN
		Diagnostic sensitivity* TP/(TP + FN) 96.8% (94.0 – 98.5%)**	Diagnostic specificity* TN/(TN + FP) 92.6% (85.4 – 97.0%)**

試驗檢定試驗結果

- 診斷靈敏度 (DSe)=真陽性率 96.8 (270/279)
- 診斷特異性 (DSp)=真陰性率 92.6 (88/95)
→ 遠低於99%
→ 複檢樣本數需要707個陰性樣本

診斷表現_診斷靈敏度與特異性 (3)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)
有完美參考標準_比較試驗

- 比較不同族群感染動物：臨床感染與亞臨床感染動物的DSe、不同地區的DSp...
 - ✓ 比較獨立樣本
→ **Pearson's chi-square test**或計算各別95% CI與兩者差異的95% CI
 - ✓ 比較配對 (paired) 樣本
→ **Mcnemar's chi-square test** → 測試Dse與DSp是否同等

範例：

	Infected		
	T ₂ +	T ₂ -	
T ₁ +	124	74	198
T ₁ -	8	243	251
	132	317	449

Sensitivity of T1 = 198/449 = 44.1%

Sensitivity of T2 = 132/449 = 29.4%

Mcnemar's
chi-square test :

DSe有顯著差異 (p<0.0001)

共變異數
(covariance) :

0.147 → 強相關性
(p < 0.0001 by Pearson's chi-square)

計算相關性

	Non-infected		
	T ₂ +	T ₂ -	
T ₁ +	3	27	30
T ₁ -	16	366	382
	19	393	412

Specificity of T1 = 382/412 = 92.7%

Specificity of T2 = 393/412 = 95.4%

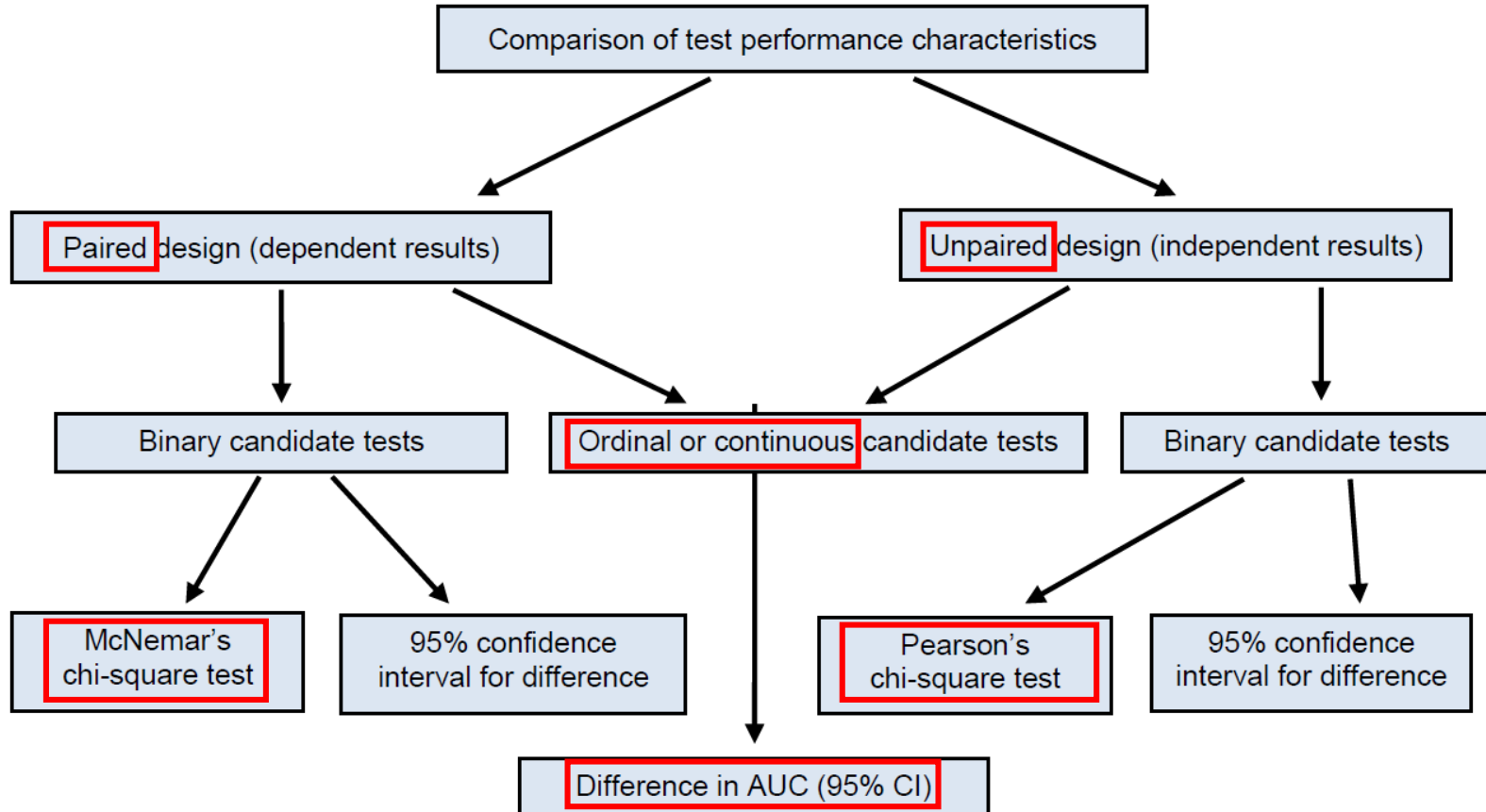
DSp無顯著差異 (p= 0.126)

0.004 → 弱相關性
(p =0.152 by Pearson's chi-square)

診斷表現_診斷靈敏度與特異性 (4)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DS_p)

有完美參考標準_比較試驗



Abbreviation: AUC = area under the receiver operating characteristic curve

Abbreviations: DSe = diagnostic sensitivity; DS_p = diagnostic specificity; ROC = receiver operating characteristic; CI = confidence interval; PI = probability interval.

診斷表現_診斷靈敏度與特異性 (5)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)

無完美參考標準_單一試驗

- 潛在類別 (latent class, 又稱no-gold-standard) 模型
 - ✓ 遭遇問題
若參考試驗不完美，所得DSe與DSp會有瑕疵 → 可用潛在類別模型，假定兩個試驗皆不完美
 - ✓ 計算由多個試驗檢測多個族群的結果 → WinBUGS軟體
最大概似估計 (maximum likelihood estimation)
貝氏分析 (Bayesian methods)
 - ✓ 可識別的條件
在無數次觀察後，理論上可獲得模型參數的真實數值 → 有獨特的一組參數
貝式分析可用於有DSe與DSp的先驗訊息 (prior information) 時
→ 先驗資料與數據的強度會影響最終估計值，可再以無訊息先驗 (non-informative prior) 確認所有參數
- 樣本
 - ✓ 參考族群 (非個體) 必須詳實描述可能在族群內循環的病原
 - ✓ 宿主、病原、疾病進程 (發生率死亡率)

診斷表現_診斷靈敏度與特異性 (6)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)

無完美參考標準_單一試驗

- 單族群潛在類別模型
以3個條件獨立的試驗檢測相同樣本 → 除非有不同的分析物，否則難以施行
- 雙試驗雙族群模型
以2個試驗檢測2個族群的所有樣本 → 成本較少
需要假定：
 - ✓ 在2個族群有一致的DSe與DSp
 - 若一族群為臨床感染，另一族群為亞臨床感染，可能難以驗證與更正
 - 若一族群盛行率為零，另一族群盛行率不為零，前者可用於估計DSp，以利估計DSe
 - ✓ 明確的盛行率
 - ✓ 條件獨立

診斷表現_診斷靈敏度與特異性 (7)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)

無完美參考標準_單一試驗

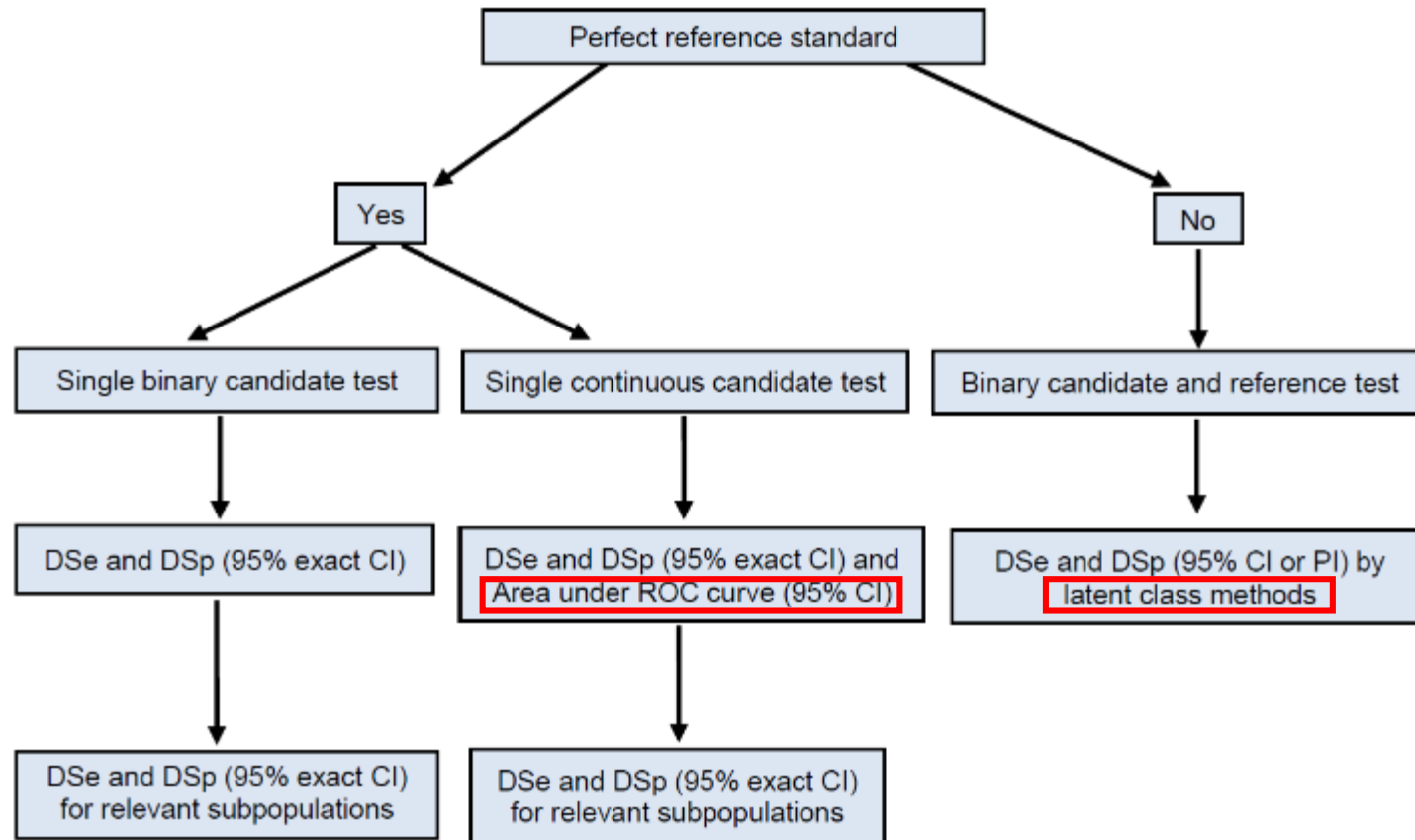
範例：

- 檢測非洲馬疫 (African horse sickness, AHS) RT-PCR與傳統病毒分離的DSe與DSp
→ 雙試驗雙族群貝式模型
 - ✓ 503 個AHS疑似病例
 - ✓ 503 個樣本
- 檢測結果
 - ✓ 疑似病例
 - PCR+VI+ (n=156)
 - PCR+VI- (n=184)
 - PCR-VI+ (n=0)
 - PCR-VI- (n=163)
 - ✓ 健康馬匹
 - 全部陰性
- WinBUGS軟體計算結果
 - ✓ PCR DSe = 0.996 (0.977 – 0.999)
 - ✓ PCR DSp = 0.999 (0.993 – 1.0)
 - ✓ VI DSe = 0.458 (0.404 – 0.51)
 - ✓ VI DSp = 0.999 (0.998 – 1.0)

	PCR+	PCR-	
VI+	156	0	156
VI-	184	163	347
	340	163	

診斷表現_診斷靈敏度與特異性 (8)

Diagnostic Performance_Diagnostic Sensitivity & Specificity (DSe & DSp)
單一試驗



Abbreviations: DSe = diagnostic sensitivity; DSp = diagnostic specificity; ROC = receiver operating characteristic; CI = confidence interval; PI = probability interval.

臨時檢定識別

Provisional Assay Recognition

- 面臨問題
 - ✓ 樣本數可能無法達到需求→常為完成確效的最大障礙
 - ✓ 對於非流行性或不廣泛分布的疾病很難在初期取得足夠的樣本數，但額外的資料可隨時累積，並用於調整截止值或提高信心水準
- 臨時檢定識別
 - ✓ 定義：已完成分析靈敏度、分析特異性及重複性，並初步以較少但資訊完整含有分析物的樣本計算 D_{Sp} 與 D_{Se} 和確認重複性
→即部分完成確效
 - ✓ 所有平台樣本都要來自不同的個體，不可包含單一個體的重複採樣
 - ✓ 理想中需要60個樣本
 - ✓ 確認再現性可提高 D_{Sp} 與 D_{Se} 可接受性
方法：在兩個不同的單位，以同樣的樣本、步驟及試劑，但不同的儀器，進行檢測
→即縮小版本的再現性試驗
 - ✓ 由主管機關認定為臨時檢定識別，代表該檢定未完成診斷表現評估，開發者應繼續增加予評估樣本 (通常此過程應限制大體時間)

再現性 (1)

Reproducibility

- 目的
 - ✓ 評估檢定於多個實驗室的精確性
- 試驗方法
 - ✓ 樣本準備
 - 準備至少20個涵蓋各種預期結果的樣本，約25%為陰性，其餘75%為橫跨檢定操作範圍的陽性 (代表性)
 - 優先取自自然感染的動物的樣本，並包含不超過5個陰性樣本
 - 若為實驗取得材料，每組至少5個檢體，依重複次數分裝樣本多管
 - 若要在多個實驗室常規使用，通常需要至少數百個分裝管 (可測試橫跨多個試驗間隔的相同樣本，以偵測系統性誤差)
 - 由於此類參考材料消耗快速，須持續置換補充
 - 寄送至少3個實驗室
 - ✓ 檢測
 - 每個實驗室需以2個批次的檢定檢測樣本，如批次間結果有差異，不能重新檢測
 - 應進行盲測並註明盲測方法與隨機分配樣本方法
 - 除平台外，參與的實驗室需檢測該實驗室收到的其他檢體，以評估合適性
 - 此評估對於新鮮的樣本，如全血與糞便樣本極具價值，尤其是需要前處理步驟的新鮮樣本
 - 可單獨執行或與重複性研究一起進行

再現性 (2)

Reproducibility

- 試驗方法
 - ✓ 分析方法
 - 與重複性相似，唯須評估分類 (變因) 造成的變異
 - 如：實驗室、操作者、批次、重複次數
 - 例：實驗設計由3個實驗室，各2名操作員，以2個批次的套組，進行2重複檢測
 - 要檢測24次
 - 各分類可為固定或隨機，並應考量其代表性
 - 可評估各分類的方差分量 (variance component) 與類內相關係數 (intraclass correlation coefficient)
- 技術修改 (technical modification)
 - ✓ 完成實驗室內確效後，應考量於不同環境 (如pen side應用) 的使用
 - 可能有極端的改變 (如溫度波動)
 - 可能造成系統性測量誤差 (高估或低估的測量數值)
 - ✓ 評估拆分樣本 (split sample) 於實驗室內與pen side的結果差異的平均值的95% CI
 - 若95% CI排除0，代表確實有系統偏差 (systematic deviation)
 - 兩者無可比性，應進行技術修改或針對pen side應用完全重新確效 (也可用於評估實驗室內的方法改變是否有隨機或系統偏差)

再現性 (3)

Reproducibility

範例：

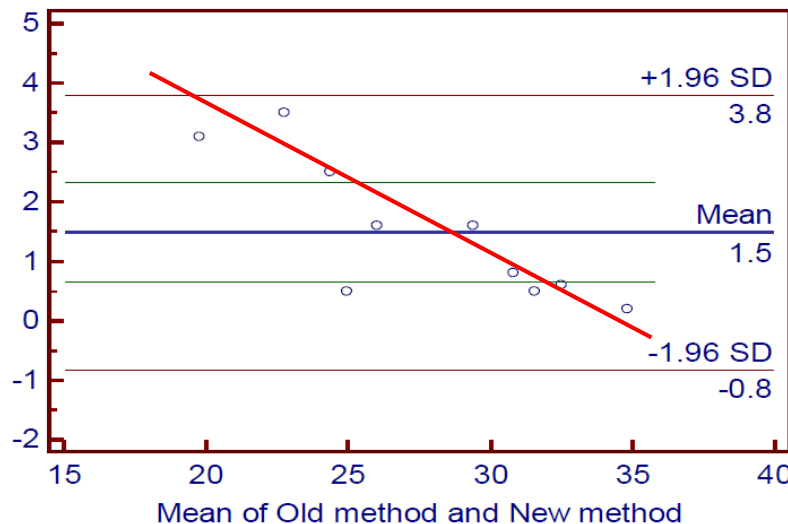
兩種萃取方法進行藍舌病qPCR之CT值的結果平均值 (n=10)

舊方法：25.6, 24.5, 21.3, 26.8, 25.2, 30.2, 31.2, 32.8, 31.8, 34.9

新方法：23.1, 21.0, 18.2, 25.2, 24.7, 28.6, 30.4, 32.2, 31.3, 34.7

做Bland-Altman圖

- 95% CI排除0 → 新方法系統性低於舊方法
- 隨CT值增加，差異變小



兩種方法的差異的平均值 = -1.49
95% CI = -2.33~-0.64 p=0.003

Figure 3. Bland-Altman plot of mean difference (y axis) in CT values as a function of the mean value of the old method and new method (n=10)

實施 (1)

Implementation

- 提供檢定表現的額外有價值的證據
- 適合使用
 - ✓ 應考量研究機構、主管機構及客戶的可接受性，以及實驗室資源的可行性
某些疾病的控制與監控計畫可合併使用多個檢定，因此應評估DSe、DSp、與預估數值的改變
 - ✓ 若無法達到操作需求，同樣會使檢定不適合預期用途
 - 表現成本、儀器可取得性、技術複雜度與解讀技巧、保質期、運輸需求、安全性、生物安全性、吞吐量、周轉時間、品質管理與保證、是否可實際用於其他實驗室
 - 田間使用的檢定強調易於使用，但由於環境不可控，需要更多預防措施以保證適合使用

實施 (2)

Implementation

- 試驗結果解讀

- ✓ 預測值 (predictive value)

- **陽性預測值 (PPV)**：被檢測為陽性的動物確實是陽性的機率
= 真陽性 / (真陽性 + 偽陽性)

- **陰性預測值 (NPV)**：被檢測為陰性的動物確實是陰性的機率
= 真陰性 / (真陰性 + 偽陰性)

- ✓ 預測值受 **DSe**、**DSp** 及 **特定時間點目標族群的真實盛行率** 影響

- ✓ 預測值對疾病控制與清除計畫極為重要 (例如偽陽性高會增加撲殺成本)

- ✓ 疾病控制計畫一般建議 **依據目標族群的感染盛行率與計畫目標改變 DSe 與 DSp**，以符合經濟考量

→ 依 ROC 分析結果設定多個截止值，依各截止值的 DSe 與 DSp 計算撲殺成本

- ✓ 若目的為建立 **疾病清淨證據**，**NPV** 更為重要，且受到 **DSe** 影響

$$PPV = \frac{P \times DSe}{P \times DSe + (1 - P) \times (1 - DSp)}$$

and

$$NPV = \frac{(1 - P) \times DSp}{P \times (1 - DSe) + (1 - P) \times DSp}$$

PPV = Predictive value of a positive test result

NPV = Predictive value of a negative test result

P = Prevalence of infection

DSe = Diagnostic sensitivity

DSp = Diagnostic specificity

P↑ → PPV↑; NPV↓

P↓ → PPV↓; NPV↑

實施 (3)

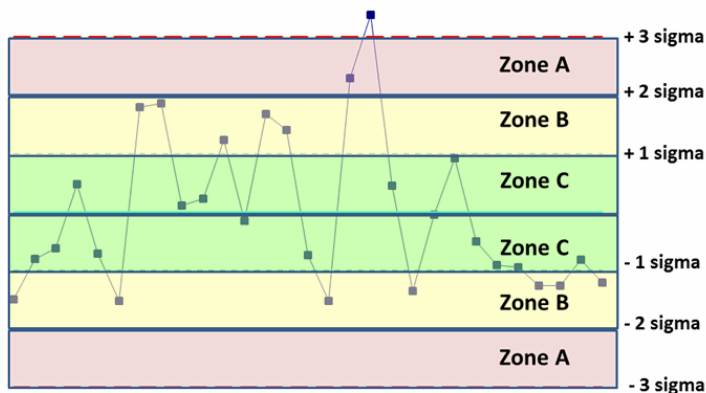
Implementation

- 國際認證
 - ✓ 傳統上，當檢定被指定用於貿易目的的規定或替代試驗時，該檢定即被OIE視為國際認可
 - ✓ 這通常基於它們在國家、區域或國際的用處的證據
 - ✓ 對於已經通過 OIE 確效和認證程序的商品化診斷試劑，最後一步是在 OIE 註冊
 - ✓ 已註冊的檢定如果已完成分析表現、診斷表現及再現性確性，則被證明適合特定目的
 - ✓ 登記冊旨在為潛在的套組用戶提供有關套組及其性能特徵的信息
- 佈署檢定
 - ✓ 檢定有用性的最終證據是其**在其他實驗室中的成功應用**以及被納入國家、區域和/或國際控制或監測計劃，參考實驗室在這一過程中發揮著關鍵作用
 - ✓ 在診斷和/或技術改進的進程中，**新的檢定**將成為與其他檢定進行比較的新標準方法可能會逐步獲得國家、地區和國際認可，並用於開發用於品質控制、能力測試和協調目的的參考試劑這些參考試劑也可能成為國際標準品
 - ✓ 當測試從開發實驗室轉移到現場時，無論是用於當地實驗室還是現場應用，都應重複再現性評估
 - ✓ 應針對可能影響再現性估計的可預測的變化 (例如極端溫度和操作員經驗水平) 進行評估

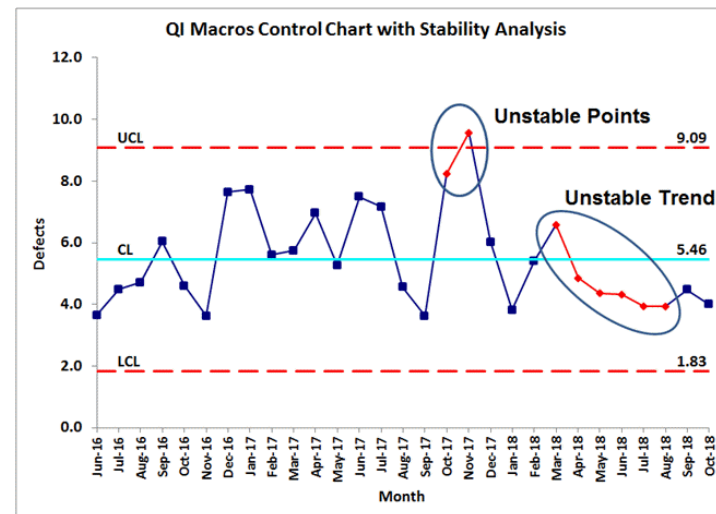
監控 (1)

Monitor

- 監控檢定
 - ✓ 目的
為維持檢定的狀態，必須確定其始終保持與確效試驗時有一致的表現
 - ✓ 品質保證計劃
通過估計內部控制的精確度和準確度以及離群值趨勢，監控檢定的日常表現可繪製控制圖表 (control chart) 來監控表現
控制圖表：呈現控制樣本在不同次操作的重複測量值
應調查與預期表現的偏差以便糾正
這種監控可證明該檢定在實施階段仍保持「確效」的狀態
 - ✓ 外部品質控制計畫 (例如能力驗證)
評估再現性
 - ✓ 若該檢定無法產生與原始確效數據一致的結果，將被視為不適合預期用途
→ 必須持續評估經確效的檢定以確保其保持適用性



<https://www.qimacros.com/control-chart/stability-analysis-control-chart-rules/>



監控 (2)

Monitor

- 修改與增強
 - ✓ 當預期用途、目標分析物或技術改變時，檢定需進行修改，以改善檢定效率與成本效益
 - 若改變預期用途，應從診斷表現開始重新確效
 - 若檢定應用於其他地區或族群，需要針對特定目標族群重新確效
 - 病原的譜系或亞譜系可能有變化，尤其是核酸檢測 (NAD) 系統
 - 點突變常發生在許多病原 (尤其是 RNA 病毒)，引物 (primer) 或探針位點內發生的突變會影響檢測效率，甚至使已建立的檢定無效
 - 建議定期確認所選基因組區域的目標序列，以用於國家或地區的病原分離株，以確保它們保持穩定不會影響檢定之 DSe 和 Dsp
 - 類似的問題可能出現在免疫學檢定的抗原或抗體
 - 若現有病原的新亞型出現，可能需要修改現有的檢定

監控 (3)

Monitor

- ✓ 技術修改與可比性評估
 - 已確效檢定的技術修改，如儀器、萃取流程或改用半自動或全自動系統，通常不需要全面重新確效，但要進行方法比較研究以確定是否影響表現
 - 樣本在試劑耗盡前完成替代試劑的配製與重複測試
與原樣本同時於檢定測試多次，以建立比例關係，一次只替換一種試劑
 - 通過多次並排操作修改後的和原先的流程來建立可比性
檢測相同且可代表兩種流程的完整操作範圍的樣本
 - 根據預先指定的標準，如果確定修改後的和原先的流程的結果具有可比性，則修改後的檢定對其預期用途仍然有效

監控 (3)

Monitor

- ✓ 生物學修改與可比性評估
 - 在某些情況下，可能需要改變某些生物製品
 - ▣ 樣本 (如待測組織或物種)
 - ▣ 試劑 (如在 ELISA 中用重組抗原替代細胞培養衍生的抗原，或替換抗體接合物)
 - 應考量是否需要在實驗室和現場進行完全重新確效
必須先評估確認分析表現達標，再考慮使否重新完整確效診斷表現
確效診斷表現需要至少60個獨立樣本
 - 可比性評估 (診斷表現)
以一組已知陽性和陰性診斷樣本比較原先 (參考) 檢定與修改後的 (候選) 檢定
如果診斷性能未發生變化，則可將修改後的檢定用於常規使用
如果觀察到 DSp 和 DSe 的差異，則修改後的檢定需要額外的診斷表現或
田間確效才能被採用
- ✓ 耗盡試劑替換
 - 當對照樣本或工作標準品等試劑接近耗盡時，必須在耗盡前準備並重複測試替代品
 - 預期對照樣本應與原始對照進行多次平行試驗，以建立它們的比例關係
一次只更換一種試劑，以免評估多個變量的複合問題

監控 (4)

Monitor

- 增強信心水準
 - ✓ 許多宿主變量對檢定的診斷表現有影響
 - 需要增加樣本的數量，用於參考樣本或潛在類別分析
 - 樣本數量的增加提高了 DSe 和 DSp 總體估計值的精確度，並可根據年齡、疾病階段和病原量等因素計算 DSe 估計值
 - 新樣本的採樣設計、採集、運輸和測試環境應與原先的確效研究相同
 - 新數據應每年包含在相關測試檔案中
- 驗證現存檢定
 - ✓ 如果實驗室正在考慮使用已確效的商品化套組或基於已發表的具有驗證數據的文獻的候選檢定，則需要確定該檢定是否符合製造商或作者在預期應用的聲明
 - 需要以可用的參考材料對 ASp 和 ASe 進行有限的驗證，無論是從外部的和/或本地目標族群獲得
 - ✓ 若實驗室確信該檢定的描述符合分析表現後，應在預期應用和目標族群考慮進行有限的診斷表現確效，才能將檢定投入常規診斷使用

監控 (5)

Monitor

- 能力試驗：
 - ✓ 目的：評估實驗室對已建立且確效的檢定的表現
 - ✓ 檢定於多個實驗室的常規使用需要持續監控，外部品質控制計畫，由實驗室間比較得實驗室能力的測量
 - ✓ 結果多為定性
 - ✓ 單一稀釋檢定
 - ✓ 半定量結果提供評估非偶然誤差的額外資料
 - ✓ 至少5個樣本，代表陰性、弱陽性與強陽性

可比性

Comparability

檢定改變

- 小改變需要確認檢定表現的可比性，並評估是否需要重新確效
- 大改變 (生物學基礎被改變)，需要重新確效

改變類型	檢定改變	目標族群或檢體改變
小	<ul style="list-style-type: none"> • 耗盡試劑替換 (PC、新批次的抗原、盤子與接合物) • 儀器與平台 (讀盤機、感作箱/震盪機、PCR機) • 從個別包覆到事先包覆的ELISA 盤 • 從手動操作到機器操作 • 核酸萃取步驟 • 修改引子 (primer) 或探針 (probe) • 增加額外的探針 • PCR作用條件改變 (相同引子或探針) • 改變探針化學性質 	
大	<ul style="list-style-type: none"> • 以細胞培養抗原替換重組抗原 • 從非直接型ELISA到競爭型ELISA • 改用不同標的的引子或探針 	<ul style="list-style-type: none"> • 不同物種 (牛與水牛、家禽與野鳥) • 不同檢體類型 (氣管與泄殖腔採樣、血液與精液、不同組織或臟器)

可比性 Comparability

設置對比實驗

1. 根據預期用途決定實驗
2. 決定判斷指標與允收標準
3. 進行盲測
4. 圖像評估與統計學分析
5. 評估可比性 (95% CI、精確性、準確性、ROC)
6. 最終決定與其他考量

可比性 Comparability

可比性試驗範例

Figure 5. Layout of 96-well plate to assess analytical Se, diagnostic Se and Sp and repeatability.

	1	2	3	4	5	6	7	8	9	10	11	12
A	10^{-1}	10^{-1}	10^{-1}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
B	10^{-2}	10^{-2}	10^{-2}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
C	10^{-3}	10^{-3}	10^{-3}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
D	10^{-4}	10^{-4}	10^{-4}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
E	10^{-5}	10^{-5}	10^{-5}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
F	10^{-6}	10^{-6}	10^{-6}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
G	10^{-7}	10^{-7}	10^{-7}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C
H	10^{-8}	10^{-8}	10^{-8}	Neg	Neg	C++	C++	C+	C+	C+++	C+++	C

計算重複間的結果
以評估重複性

10倍序列稀釋
陽性樣本3重複
(分析靈敏度)

陰性或感染其他
病原的樣本2重複
(診斷特異性)

不同強度的田間
陽性樣本2重複
(診斷靈敏度)

評估交叉汙染

可比性 Comparability

圖像檢查

散佈圖

- 相關性：線性或對數關係
- 有無離群值、缺失值及人為干擾
- 無法提供一致性評估 (除非都在對角線上，即100%一致)

Figure 2. Scatter diagram of a weak positive control sample after being tested 28 times in two different Hendra TaqMan assays, M1 and N1 (results expressed as cycle threshold [Ct] values).

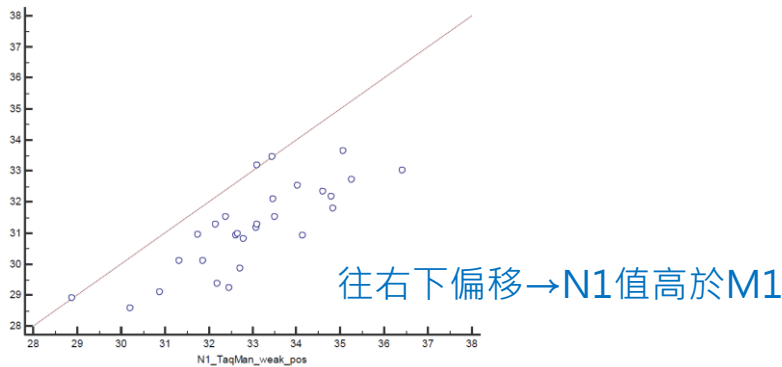


Table 2. Statistical analysis of a weak positive control sample after being tested 28 times in two Hendra TaqMan assays, M1 and N1

Variable Y	M1 TaqMan weak positive
Variable X	N1 TaqMan weak positive
Sample size	28
Correlation coefficient r	0.8015
Significance level	$p < 0.0001$
95% confidence interval for r	0.6112 to 0.9042

強陽性相關性
結果為隨機的機率低

直方圖

- 可觀察左傾、右傾或雙峰

Figure 3. Histogram of a weak positive control sample after being tested 28 times in two Hendra TaqMan assays, M1 and N1 (results expressed as cycle threshold [Ct] values).

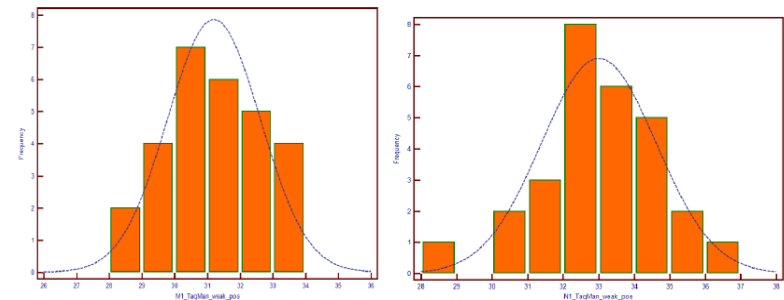


Table 3. Statistical analysis of a weak positive control sample after being tested 28 times in two Hendra TaqMan assays, M1 and N1 (results expressed as cycle threshold [Ct] values)

	M1 Taqman weak positive	N1 Taqman weak positive
Sample size	28	28
Lowest value	28.58	28.88
Highest value	33.63	36.42
Mean	31.19	33.00
Median	31.22	32.94
Standard deviation (SD)	1.42	1.62

可比性 Comparability

重複性

並行操作多次並計算平均值與SD，觀察兩者結果是否符合允收標準 (如： $\pm 2 \sim 3SD$ 、 $\pm 2 \sim 3Ct$)

Bland-Altman plot (Tukey 均值差圖)

- 同時顯示和分析比較研究的結果 (配對設計)
- 差異與平均的關係→評估系統性偏差與可能的離群值

Figure 4. Bland-Altman plot showing differences of cycle threshold (Ct) values in two Hendra TaqMan assays, M1 and N1 for a weak-positive control sample after being tested 28 times.

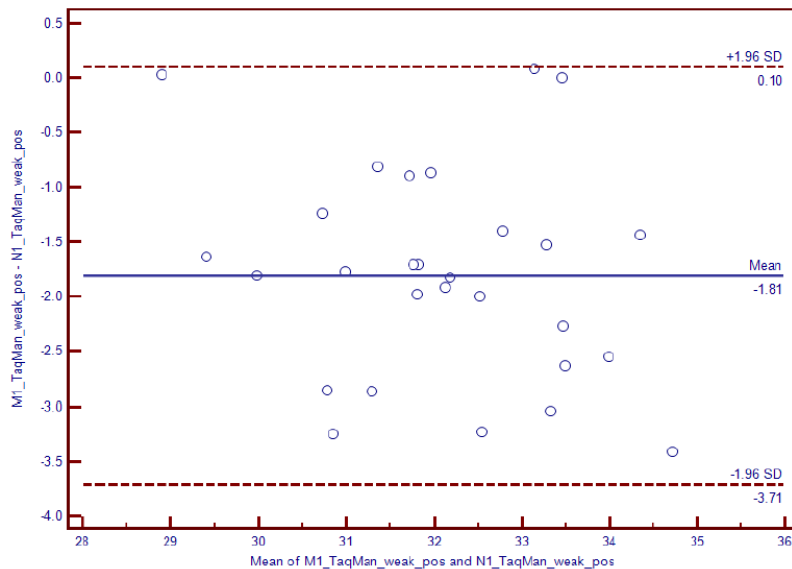


Table 4. Statistical analysis for Bland-Altman plot

Method A	M1 TaqMan weak positive
Method B	N1 TaqMan weak positive
Differences	
Sample size	28
Arithmetic mean	-1.8054
95% CI	-2.1831 to -1.4276 排除0→無可比性
Standard deviation	0.9741

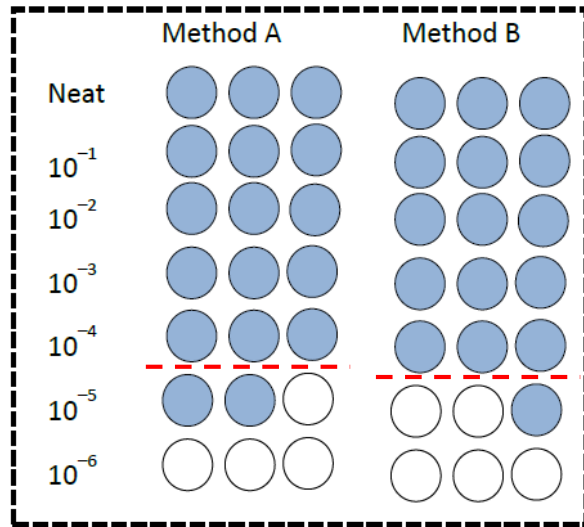
可藉由調整截止值平衡系統性偏差

可比性 Comparability

偵測極限 (LOD) 實驗

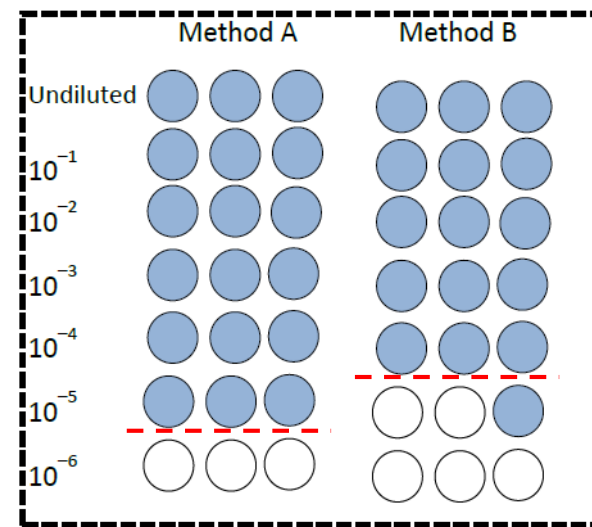
- 保守估計LOD：100%陽性的最低稀釋倍率

Figure 6. Example of a limit of detection (LOD) experiment with an acceptable outcome.



10^{-5} 以下的稀釋倍率被認為無可比性
兩個方法有可比性

Figure 7. Example of a limit of detection (LOD) experiment with a non-acceptable outcome.



結果顯示兩個方法無可比性
應重複實驗多次以確認可比性
或使用其他稀釋倍率

可比性 Comparability

ROC曲線比較

- 評估與比較整體診斷準確性
比較多個檢定於不同截止值的結果

0.5：無用

0.5-0.7：低準確性

0.7-0.9：中等準確性

0.9-1：高準確性

Figure 8. Comparison of receiver operating characteristic (ROC) area under the curve (AUC) for four different ELISAs to detect antibodies against influenza virus in pigs.

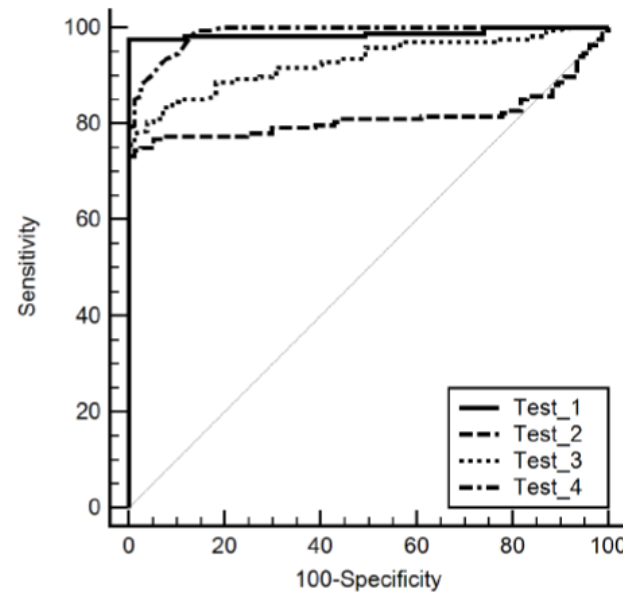


Table 5. Receiver operating characteristic (ROC) comparison of area under the curve (AUC) and p-values for four different ELISAs to detect antibodies against influenza in pigs

Parameter	Test 1	Test 2	Test 3	Test 4
Area under ROC curve	0.988	0.814	0.929	0.988

TEST 1準確性最高

Table 6. Pairwise comparison of ROC curves

	Test 1 vs 4	Test 1 vs 3	Test 3 vs 4	Test 2 vs 3	Test 1 vs 2	Test 2 vs 4
Difference between AUC	0.0002	0.055	0.0589	0.115	0.173	0.174
95% CI	-0.016 to 0.016	0.025 to 0.092	0.027 to 0.090	0.069 to 0.161	0.117 to 0.230	0.118 to 0.229
Significance level	$p=0.9808$	$p=0.0007$	$p=0.0003$	$p<0.0001$	$p<0.0001$	$p<0.0001$

TEST 1與TEST 4差異最小

可比性 Comparability

討論與結論

- 可比性：統計分析與客觀評估
- 其他考量：成本/儀器、吞吐量、周轉時間、品質保證能力、技術成熟度、監管或科學界的接受度和結果判讀能力
- 範例：
 - ✓ 篩選性試驗需要高DSe
 - ✓ 有複數針對不同標的的檢定可以減少錯過新變異株的機會 (於致命人畜共通疾病為重要考量)
 - ✓ 手動操作改以機器操作時要考量汙染問題

允收標準可有彈性 (如：低於1~3 Ct、10% 差異、1~3 SD)，但其表現不可明顯低於已確效檢定

最終由有權者 (技術經理、實驗室指導者、品質經理) 根據預期用途決定不同方法是否有可比性

最終結果與決定過程要完整記錄以利審計追蹤