



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

## 美國 FDA 針對電腦輔助偵測軟體(CADe)的獨立效能 與臨床效能之評估要求

林事曄<sup>1</sup>

### 前言

在工業革命之後，電子計算機的發明曾是改變科技發展的大突破，而在這個群眾已經習慣智慧型手機、電腦、軟體等電子化產品的時代，下一個備受矚目的發展當屬人工智慧/機器學習 (artificial intelligence/machine learning, AI/ML) 技術。在各個層面的應用中，隨著每天無數的醫療數據產生，AI 技術帶來的可能性促成了醫療照護新穎進步的重要轉變，例如廣泛應用於放射影像疾病偵測、診斷及病情分類/通知的醫療器材。美國 FDA 為提供病患高品質且安全有效的 AI/ML 醫療器材，以透明且具實證的證據支持器材效能的評估，乃針對較為使用的 AI/ML 醫療器材建立法規指引，例如電腦輔助偵測軟體醫療器材(Computer-assisted detection device, CADe)，在在展現其在醫療器材法規的領導地位。自 1998 年美國 FDA 通過第一個 CADe 器材的上市前核准(premarket approval, PMA) (P970058, M1000 Imagechecker, Hologic, Inc.)之後，已陸續核准牙齒、胸腔、乳房攝影及超音波等相關 CADe 器材，並於 2012 年發布 CADe 器材申請上市<sup>[1]</sup>與臨床效能評估<sup>[2]</sup>指引。當時分類為 21 CFR 892.2050 的 CADe 器材屬於 Class II，以 510(k)途徑(上市前通知, premarket notification)申請；而分類為 21 CFR 892.2070 的 CADe 器材則歸屬於 Class III，需要申請 PMA 核准上市。而隨著法規經驗與科學知識的進步，美國 FDA 於 2020 年 1 月將原屬於 Class III 的 CADe 器材 (21 CFR 892.2070)降低其風險等級為 Class II，將其上市申請途徑從 PMA 改為 510(k)，並改版臨床效能評估指引<sup>[3]</sup>，作為現今美國 FDA 對 CADe 器材安全性及有效性評估之準則。改版後的 CADe 臨床效能評估指引<sup>[3]</sup>刪除了申請 PMA 的相關說明，並載明指引內容適用於分類屬 21 CFR 892.2050 及 21 CFR 892.2070 之 CADe 器材，除了這些行政相關內容因上市途徑調整而改變之外，在進行臨床效能評估時所需符合之實質要求仍維持與舊版<sup>[2]</sup>要求一致。於本篇文章中，將綜合整理及說明現今美國 FDA 對 CADe 器材申請上市之臨床效能要求，包含器材的描述、獨立效能

<sup>1</sup> 財團法人醫藥品查驗中心 醫療器材組

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

評估、臨床效能評估、以及兩種效能評估間之差異比較。

## 定義

醫療使用的放射影像包含由超音波、X 光、核磁共振造影(MRI)、電腦斷層掃描(CT)及正子斷層造影(PET)等設備所取得之影像及數據，而將 AI/ML 技術使用在這些放射影像及數據所產生的電腦輔助醫療器材軟體，依效能主要可分為電腦輔助偵測(CADe)、診斷(CADx)、偵測及診斷(CADe/x)、病情分流及通知(CADt)四種，其預期用途及現今美國醫療器材分類分級整理於表一。其中電腦輔助偵測軟體(CADe)是具備圖形辨識與數據分析功能的電腦計算系統，可從病患的放射診察數據資料中辨別、註記、強調或其他任意方式，使臨床醫師注意特定圖片區域或特定內容，用以呈現病患數據資料中的異常狀況，在美國屬於第二等級醫療器材。而電腦輔助診斷軟體(CADx)則可補充 CADe 器材的功能，提供診斷輔助訊息，例如良惡性病變、病況嚴重度、期別等。電腦輔助病情分流及通知軟體(CADt)則是在發現疑似特定症狀(例如腦出血)時，通知醫師並建議優先針對含此疑似症狀之影像進行評估，以協助醫事人員減輕或免除相同的臨床程序。這些不同的電腦輔助醫療器材軟體，皆以偵測影像中特定內容為基礎，再延伸其他效能，故本篇文章雖主要說明 CADe 器材在效能評估之要求，仍可供作其他電腦輔助醫療器材軟體開發的參考。

表一、放射影像電腦輔助醫療器材軟體之分類

軟體分類	預期用途	美國醫材分類	美國醫材分級
電腦輔助偵測 Computer Aided Detection, CADe	同時或次序使用以幫助 辨識可能的疾病	510(k) under 21 CFR 892.2050、892.2070	II
電腦輔助診斷 Computer Aided Diagnosis, CADx	同時或次序使用以幫助 分類癌症可疑病變	510(k) under 21 CFR 892.2060	II
電腦輔助偵測及診斷 Computer Aided Detection & Diagnosis, CADe/x	結合偵測及分類可能的 疾病	510(k) under 21 CFR 892.2090	II
電腦輔助病情分流及通知 Computer Aided Triage & Notification, CADt	發現時間敏感的放射影 像時通知臨床人員	510(k) under 21 CFR 892.2080	II



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

## 電腦輔助偵測軟體申請美國 510(k)之要求

美國的 CADe 器材分類於 21 CFR 892.2050 「Picture archiving and communications systems」(product code : NWE、OEB、OMJ) 與 21 CFR 892.2070 「Medical image analyzer」(product code : MYN)，屬於第二等級醫療器材，依循 510(k)途徑申請上市。申請時除了需符合 510(k)的一般控制外，亦需符合各分類的特殊控制(Special control)要求。美國 FDA 公告之 CADe 器材申請 510(k)上市指引<sup>[1]</sup>與臨床效能評估指引<sup>[3]</sup>，便提供這些 CADe 器材於申請 510(k)時對器材要求的相關建議，以利製造商參考及依循。應注意的是，CADe 申請 510(k)上市指引<sup>[1]</sup>，並不適用於預期用途為輔助侵入性手術過程使用的 CADe 器材以及屬於 CADx、CADt 之器材。以下就 CADe 器材申請 510(k)上市指引<sup>[1]</sup>針對 CADe 產品之器材描述與獨立效能評估之相關建議與要求作說明。

### 1 器材描述(Describing the Device)

CADe 器材是將輸入的影像或數據資料經由演算法計算後，轉換為在影像上的標記等資訊，以提醒臨床人員注意影像上的特徵，故演算法為 CADe 器材的重要核心，申請時應針對 CADe 演算法提供總覽及描述如下。

#### (一) 一般資訊(General Information)

包含 CADe 器材預期用途的目標族群(如病症、患者、使用者等)；符合使用的放射影像類型；目標病症的臨床實務；CADe 器材的臨床工作流程(包含臨床使用之標示及使用時機、特殊宣稱對判讀時間的影響)，並與類似品比較；使用 CADe 器材協助影像判讀所帶來的影響(考量 CADe 器材判讀結果為真陽性、真陰性、偽陽性及偽陰性之情況下)，並與類似品比較；與類似品相比 CADe 器材的限制等。

#### (二) 演算法的設計與功能(Algorithm Design and Functions)

包含偵測標記的所有規格(例如形狀、大小、與目標病變之相對位置、邊緣及顏色等)；用以辨識演算程序、特徵、模組與分類器的流程圖；演算法/軟體的名稱、版本及軟體平台的相關特徵；簡短描述每個演算階段的設計與功能；製表比較與類似品技術特徵之異同等。



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

### (三) 處理程序(Processing)

應描述所有演算程序、相關演算法的流程圖與參考標準、以及用以建立適當校正與標準化之技術。

### (四) 特徵(Features)

特徵乃指電腦或人員在放射影像資料上用來計算，以擷取整體影像、特定區域或其像素上的特點，申請時應提供演算法擷取的特徵種類(例如族群、生物學特徵、形狀及幾何特徵等)以及特徵的描述；並簡述特徵選擇(用以減少候選特徵)以及幾何縮減(用以減少特徵空間面向)之技術。

### (五) 模組及分類器(Models and Classifiers)

模組定義為用以綜合資訊的任意方法或規則，而分類器則是人員定義或以數學定義的模組，用以在影像上擷取或分類出疾病/狀況/異常的區域。此定義的模組是依據一組特定的參數，手動或自動地對影像特徵與疾病/狀況/異常區域之間的關聯性提出假設。模組及分類器通常代表某種病徵辨識的程序，包含使用特殊特徵的單一閾值到使用多重繁複的分類器。在演算法的每個程序中，應描述所使用的模組與分類器、每個模組與分類器輸入特徵的數目以及使用順序。

### (六) 演算法訓練(Algorithm Training)

演算法訓練是用來建立參數與閾值的程序，包含參數的過濾(Filter)、特徵的選擇、模組及分類器的調整等，這些程序可以透過手動(例如：由程式開發者或醫學專家執行)、自動或是兩者兼具的方式完成。申請者應提供演算法訓練的描述與摘要(包含是由前述提及的哪一種方式完成)，以及用於決定參數設定的標準(criteria)及效能要求(performance metrics)，並摘要說明在合適的演算法開發中間階段(intermediate stages)，隨著不同階段遞增的演算法效能。

### (七) 資料庫(Databases)

資料庫指的是用來訓練及測試演算法的放射影像或數據組，可為電腦模擬數據、模擬假體數據(phantom data)或真實病患數據。若使用模擬數據或模擬假體數據，應描述所用之方法及其數據與真實病患數據間的關係。若使用真實病患數據時，則應提供以下資訊：



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. 收集病患數據的納入及排除條件；
2. 病患的族群資料(例如年紀、性別、種族)；
3. 病患與使用 CADe 器材相關的醫療紀錄；
4. 病患的病程分期及於放射科診察的醫囑；
5. 病患進行放射科診察的情況，例如使用的放射造影技術及觀察區域；
6. 描述如何收集影像數據，例如造影器材的製造商及型號、造影流程；
7. 收集影像的機構；
8. 進行影像處理(例如：影像數位化)的機構(若有)；
9. 病例數，包含疾病與正常之案例數；以及用來決定疾病狀態、位置與範圍的方法(此部分請參考下述「(八) 參考標準」段落)；
10. 依據干擾因素(confounders)或修飾作用(effect modifiers)分組的案例分佈，例如病變型態(例如：大腸瘰肉為增生型或腺瘤型)、病變大小、病變位置、疾病分期、器官特徵、共病(concomitant disease)、影像硬體(製造商/型號)、造影流程、影像收集與處理機構等；
11. 比較病患數據與 CADe 器材目標族群間於臨床、造影及病理之特徵；
12. 訓練與測試資料庫之數據收集與使用的歷程。

同時若 CADe 器材預期與已販售影像器材併用，應提供影像器材之名稱，法規狀態以及該影像器材的規格。

### (八) 參考標準(Reference standard)

參考標準或稱黃金標準(golden standard/ground truth)，是在病患影像數據資料中，用以決定病患資料是否存在疾病/狀況/異常，且可包含疾病/狀況/異常的範圍或位置。在病人數據資料上定義參考標準稱為辨真過程(truthing process)。CADe 器材的開發與評估經常依賴具有參考標準的放射影像或數據資料庫，藉此來指出每個病人的數據資料是否存在疾病/狀況/異常。申請時應提供決定參考標準的程序及其合理性，並說明參考標準的建立是否基於下列項目：

1. 從其他器材輸出的結果；
2. 已建立的臨床診斷結果(例如切片或實驗室測試)；
3. 後續臨床影像診斷，或；
4. 後續除影像外的醫學診斷或臨床人員之判讀。



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

申請者應描述決定參考標準的方法，且應在開始器材效能評估前即完成參考標準的決定。若影像數據的參考標準由臨床人員判讀，另應提供：

1. 參與的臨床判讀人員數；
2. 人員資格；
3. 人員之專業與經驗程度；
4. 參考標準制定前提供給判讀人員的說明；
5. 所有提供給判讀人員的病患臨床資訊；
6. 任何用於制定參考標準的特殊允收條件。

當有多位臨床判讀人員參與制定參考標準時，應描述如何結合多個判讀結果來完成單一參考標準的程序，及在多個判讀結果產生不一致時如何應對。應注意的是，參與制定參考標準之判讀人員不可與參與器材臨床效能評估之判讀人員相同。

#### (九) 計分方法(Scoring)

決定 CAdE 器材輸出結果與參考標準間相應關係(例如病變位置)的過程，稱為計分，申請時應提供計分過程的科學性說明，且應在開始 CAdE 器材效能評估前即決定計分方法。值得注意的是，此處與參考標準比較來進行計分的對象是 CAdE 器材輸出的標記，僅使用於 CAdE 器材的獨立效能評估，與使用於臨床效能評估的對象(即為臨床判讀人員之標記)不同；有關獨立效能評估之詳細內容請參見本文「2 獨立效能評估」段落，而臨床效能評估請參見本文「3 臨床效能評估」段落。申請方應說明效能評估之計分是否基於下列方法：

1. 電子化或非電子化；
2. 標記範圍與參考標準的重疊性，包含邊緣、面積或體積的重疊狀況的評估；
3. 標記中心到參考標準邊緣或空間位置的相關性；
4. 參考標準中心到輸出標記邊緣或空間位置的相關性；
5. 判讀人員之判讀結果，或；
6. 其他方法。

若計分方法採用判讀人員之判讀結果，應參考前「參考標準由臨床人員判讀」段落提供判讀人員相關資訊與描述說明。

#### (十) 其他資訊



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



FDA  
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

一般 CAdE 器材的風險等級為中等(moderate)·FDA 建議 CAdE 器材的軟體應依此風險等級提供軟體確效文件·且若 CAdE 器材包含現成(Off-the-shelf)軟體·亦應提供此現成軟體風險控制的相關文件·詳細內容及要求可參考 FDA 所公告之含軟體醫療器材申請上市指引<sup>[4]</sup>及含現成軟體醫療器材指引<sup>[5]</sup>。另若 CAdE 器材為搭配特定影像器材使用的軟體·應提供影像器材資訊以及 CAdE 器材輸出的檔案格式。

## 2 獨立效能評估(Standalone Performance Assessment)

因為每個新的 CAdE 器材皆代表著新演算法及軟體的實行·美國 FDA 認為新的 CAdE 器材(以及已核准的 CAdE 器材在軟體及其他設計、技術、效能上的改變)皆與已核准類似品具有不同的技術特點·且對 CAdE 器材的安全及有效性具有重要影響·因此判定此 CAdE 器材與類似品的實質等同性·須根據所檢附的臨床或科學性資料佐證此 CAdE 器材與類似品是否具有實質等同的安全及有效性。因此建議製造商即使僅是改變或調整已核准的 CAdE 器材·應諮詢 FDA 是否需要申請 510(k)。而為了輔助判定一個新的 CAdE 器材(或已核准 CAdE 器材的重大安全及有效性變更)之實質等同性·FDA 建議應提供此 CAdE 器材的獨立效能評估。

CAdE 器材的獨立效能評估·即指在無臨床人員判讀的情況下·評估 CAdE 器材本身獨立效能的表現·用以評估 CAdE 器材本身標記影像上已知異常區域及避免標記非異常區域的能力。FDA 建議以相同的數據資料組·比較新的 CAdE 器材與已核准類似品兩者的獨立效能評估結果;或確保新的 CAdE 器材及已核准類似品所用之資料組·兩者的特點具備可比性。

在進行獨立效能評估之前·應事先建立參考標準的定義、計分過程及分析方法·且應包含主要及次要的效能評估指標(endpoint)。所有的效能宣稱皆須經由預先擬定的分析計畫·以及共變異數分析(covariate analysis)結果來佐證。FDA 建議獨立效能評估應依據重要的共變異數(例如:病變型態、大小及形狀)個別進行分組(stratify)評估;這種個別的獨立效能評估·可提供使用者額外資訊·以利使用者更容易解讀 CAdE 器材輸出標記代表之意義。

CAdE 器材申請 510(k)指引<sup>[1]</sup>對獨立效能評估的各項要求·分別說明如下。

### 2.1 研究樣本群(Study Population)

進行獨立效能評估所使用的測試數據資料(testing data)·與 CAdE 器材開發及訓



練數據資料(training data)不可相同，且同時需具有目標族群及目標疾病/狀態/異常的代表性，以符合 CADe 器材的預期用途，因此 FDA 建議申請者應提供資料案例收集的計畫書。

一種較廣泛被接受的收集方式是從參與收集案例的醫療機構中，連續性地收集符合納入及排除條件的案例。測試數據資料應包含疾病案例、異常案例與正常案例。雖然為了有效不繁冗地收集具代表性的案例資料，可允許在測試數據資料中選擇性地擴增(enrich)疾病/異常案例，但這樣的做法可能會影響獨立效能評估結果，譬如造成測試數據資料組無法代表預期用途之族群。此外，FDA 建議進行壓力測試(stress testing)，意即刻意納入屬於預期用途的病例，但具備較罕見之影像特徵，可挑戰臨床人員判讀能力的影像。進行壓力測試時，應基於預期用途來選擇擴增的測試資料類型，例如：對於預期用途為偵測大腸瘰肉大小的 CADe，可選擇性地擴增具備小型瘰肉的影像數據。而不應該因為特定案例可增加 CADe 的效能，選擇性的擴增此類案例之數據資料。

進行獨立效能評估時，應根據臨床相關的干擾因素、修飾作用及共病來定義子群組，使各個重要的子群組皆包含有效的樣本數，並據此獲得這些子群組獨立效能的評估結果與信賴區間(Confidence Intervals, CIs)。然而除非新的 CADe 器材對特定子群組的效能有特殊宣稱，否則並不需要強求這些子群組的效能評估結果皆具有統計意義。綜合上述要求，研究中同時包含「擴增相關干擾因素案例」與「從每個參與的醫療院所收集連續病例」是較好的研究設計。

效能評估的樣本應具有足夠的數量來代表所宣稱效能在統計上具有意義，若是在子群組上有特殊效能宣稱，則應考慮其統計意義並決定樣本數。當有多個效能宣稱時，應預先擬定這些多個子群組的效能評估在統計上的調整。

作為獨立效能評估的一部分，應如前述「器材描述-(七) 資料庫」提供測試數據資料庫的資訊說明，且效能測試應包含偵測準確度(detection accuracy)及定位準確度(localization accuracy)。

## 2.2 測試數據資料的重複使用(Test Data Reuse)

雖然收集 CADe 器材效能評估所使用的數據資料並不容易，且在同一數據資料上進行多次測試有助於優化 CADe 器材的演算法，但重複使用測試數據(即在同一數據上



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

進行多次測試)在進行獨立效能評估時可能會產生問題，因此 FDA 並不建議在評估 CADe 器材效能時重複使用數據資料，而應採用新收集的獨立樣本。若考慮在獨立效能評估時重複使用測試數據，應證明這些數據資料任一部分的重複使用在評估 CADe 器材效能時皆不會導致非預期的偏差，且仍保有測試數據資料的可靠性。當考慮重複使用測試資料時，應控制相關人員取得效能評估結果，例如必須建立防火牆來阻隔法規評估人員以外的其他人員(例如演算法開發人員)，使其無法獲得與放射影像與數據資料相關的知識，包括效能評估結果摘要以外的所有結果。

為了最小化對數據資料進行優化的風險及保持數據資料的可靠性，當考慮重複使用任何測試數據時，應建立數位稽核軌跡(audit trail)並實行下列控制要求：

1. 從一個隨時間而擴增、較大的資料庫中隨機選擇數據資料；
2. 不再使用無法代表現今臨床實務程序或儀器所收集的數據資料；
3. 對每個案例設定一個小而固定的限制使用次數；
4. 保持前述的防火牆設定，嚴格限制法規評估人員以外人員(尤其是演算法開發相關人員)取得資料。法規評估人員之外的他人僅能得知效能評估結果摘要；
5. 應維持及追蹤每次數據資料存取的紀錄，紀錄應包含存取的人員、測試的條件及效能結果的摘要。

建立數位稽核軌跡的目的有：1) 保證適當地定義訓練資料與測試資料，使兩者互不相通；2) 在進行測試前確定已預先擬定新的 CADe 演算法；及 3) 於測試其他 CADe 演算法或設計時，提供關於相同數據資料的使用資訊。這些控制手段是用以實質降低在評估新的 CADe 器材時，使用到先前 CADe 器材已採用的數據資料的風險。

當重複使用測試數據時，除了提供整體效能的評估報告外，應同時提供相關子群組之效能評估報告，這些子群組包含：1) 用於新的 CADe 演算法的測試數據資料與之前使用過資料相重複的部分，以及 2) 測試資料中從未使用過的部分。因為這些子群組的資料數較整體資料少，因此信賴區間也會變大，然而其平均效能趨勢將有助於得知，新的 CADe 演算法是否有針對過去使用過的測試數據資料部分進行調整。

### 2.3 偵測準確性(Detection Accuracy)

FDA 建議依循前述計分方法評估 CADe 器材的獨立效能，定義 CADe 器材輸出標記的真陽性、真陰性、偽陽性、偽陰性應與 CADe 器材的預期用途具有一致性。例如，



若 CAdE 器材的預期用途是偵測所有異常(良性及惡性)，則其真陽性標記應定義為標記所有異常。反之，若 CAdE 器材的預期用途僅用於偵測部分異常(例如僅含有某些影像特徵的部分病變)，則應據此定義陽性及陰性的標記。當參考標準是依據臨床判讀人員之判讀結果來建立時，應一併考量辨真過程中臨床判讀人員的差異，以及在多位專業臨床判讀人員中建立各種規則並取得共識與一致性。

FDA 建議應基於病變、病患或其他相關評量點的獨立效能敏感性(Sensitivity, Se)及偽陽性率(False positive marks/case, FPs/case，即每個案例中平均偽陽性標記數)提供評估報告，且應依據不同的干擾因子/修飾作用(例如病變大小、病變型態、病變位置、疾病期程、造影或掃描程序、影像或資料特徵)進行分組分析。獨立效能的偽陽性率，應分別從正常及異常的病患資料中取得。若 CAdE 器材允許臨床人員選擇或更動操作點(operating point)，FDA 建議提供 CAdE 器材每個可選擇的操作點或可能操作點範圍的效能。此外，偵測準確性是依據計分方法的允收標準及閾值來決定，因此可採用數個不同距離與比率的計分閾值(若有)來評估器材的效能。偵測準確性的評估方法應在評估試驗開始前即擬定，包含選擇主要及次要的效能評估指標。所有效能評估結果都需提供其信賴區間，且應描述信賴區間計算方式以及其臨床重要性。

FDA 建議申請者說明偵測準確性以及針對分析結果進行臨床解讀時，應提供自由反應接收者操作特徵曲線(free-response receiver operating characteristic curve (FROC)，即效能敏感性(Se)對 FPs/case 作圖的曲線)，以及相應的信賴區間。FDA 也建議在效能評估時，當 CAdE 器材輸出標記有包含異常所在的位置及異常的範圍，應考量兩者各自的差異(例如 CAdE 器材標記區域與參考標準區域兩者交集與聯集的比率)。

## 2.4 通用性測試(Generalizability Testing)

FDA 建議應闡明 CAdE 器材標籤上所列的造影技術對於 CAdE 器材效能評估結果的影響，例如研究不同造影技術及造影參數在 CAdE 器材效能上的影響，以了解 CAdE 演算法的穩定性。進行此研究時應採用適當的方法且不可對病患產生額外風險(例如增加輻射或對比劑的暴露量)，且應包含測試不同造影技術的干擾因素(如影像硬體、造影或掃描程序及影像/數據資料特徵等)的影響。

CAdE 器材通用性的測試可包含下列：



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



FDA  
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. 當用於開發及測試的數據資料反應出造影的要求時，對預期會在結果中造成變異的參數進行分組分析；
2. 提供技術性的可靠論證或實驗資料，以支持 CAdE 器材與使用相似技術之不同影像系統間的相容性；
3. 當 CAdE 器材是以數位影像數據輸入時，以同一組影像比較不同掃描器數位化結果的效能；
4. 從同一位病患身上以同一種影像裝置(例如 MRI 或超音波)取得多張影像後進行效能評估，且應注意不可對病患產生額外風險；
5. 當模擬數據資料已經經過驗證，使用模擬數據資料替代真實影像進行效能評估。

而在通用性評估中，FDA 建議提供：

1. 評估預期會造成結果變異的參數(例如掃描的特徵、影像硬體的組成及模型、取得影像的程序及參數等)，以及這些參數對 CAdE 器材效能的影響；
2. 研究執行的描述；
3. 評估的結果與統計分析。

FDA 建議註明在驗證測試中使用的特定器材、技術與取得影像的技術等，並將這些相容的資訊刊載在 CAdE 器材的標籤中。

## 2.5 類似品比較(Comparison with Predicate Device)

當可行時，FDA 建議直接比較 CAdE 器材與類似品的獨立效能評估，亦即使用**同樣的評估程序及資料組**進行效能比較。有時候與其他製造商製造的 CAdE 器材進行獨立效能評估的比較是較為困難的，然而當申請同一 CAdE 器材改版的 510(k)時，其獨立效能評估應與前一版本進行直接比較。進行類似品比較時，應描述此比較的分析、比較的假說、樣本數的估算以及試驗評估指標，並提供比較結果，包含：

1. 獨立效能 FROC 曲線下面積的差異及相關統計分析；
2. 在每個操作點中，獨立效能偵測敏感性及每個案例中偽陽性標記數的差異；
3. 新/改版的 CAdE 器材與類似品比較真陽性及偽陽性標記的位置及範圍。

因具有相似的獨立效能偵測敏感性及偽陽性率的兩種 CAdE 器材，可能會在同一影像中標記不同的影像構造，造成實質上不同的臨床效能，故建議進行此類似品比較



來決定新的 CADe 器材標記與類似品標記的一致性。

## 2.6 電子資料(Electronic Data)

FDA 建議申請時應以電子方式檢附任何研究中統計分析的資料，包含病患資訊、疾病與正常狀態、病變大小、病變型態、造影與掃描設定、以及影像與數據資料的特徵等。

## 3 臨床效能評估(Clinical Performance Assessment)

FDA 建議當新的 CADe 器材與已核准類似品可在獨立效能評估**直接**比較時，意即與類似品**以相同的評估程序及測試資料庫進行獨立效能評估**，則可免除臨床效能評估；此時，獨立效能評估之結果須同時符合下列至少一項條件：

1. 新的 CADe 器材在獨立效能評估中具備更好的敏感性指標，且具有較類似品同等或更好的偽陽性率指標；
2. 新的 CADe 器材在獨立效能評估中具備更好的偽陽性率指標，且具有較類似品同等或更好的敏感性指標；
3. 新的 CADe 器材在獨立效能評估中具備較類似品更好的 FROC 曲線；

並且新的 CADe 器材其真陽性與偽陽性的標記與已核准類似品相比，在輸出標記的位置以及範圍上須具備合理的一致性。

多數可免除臨床效能評估的情形，為改版或修正的 CADe 器材與前一個版本進行直接的比較。除此之外，也許在某些情況下可僅以獨立效能試驗來證明與類似品的實質等同，但當新的 CADe 器材無法與類似品進行直接比較、或其預期用途有所改變、或與類似品相比具有實質上不同的技術特點時，應考量進行臨床效能評估來證明與類似品的實質等同性，因此當製造商認為新的 CADe 器材不需要進行臨床效能評估時，應先尋求 FDA 之建議，且不論是否進行臨床效能評估，FDA 建議皆應進行 CADe 器材的獨立效能評估。

在多數的情況下 FDA 建議提供臨床效能評估來證明與類似品的實質等同性，例如：

1. 資料庫的建立及特徵與類似品無法比較且造成臨床上的疑慮時；
2. 獨立效能評估的結果較類似品差時；
3. CADe 輸出標記的位置及範圍與類似品實質上不同且造成臨床上的疑慮時；



4. 參考標準的定義、計分程序、分析方法或效能評估指標與類似品不同，且對臨床人員或病患的重要性及影響並無清楚的理解或清楚的描述時；
5. 演算法的設計與類似品不同且造成臨床上的疑慮時；
6. 採用新的數據資料擷取技術或程序，改變了 CADe 器材的輸入數據資料型態，且造成臨床上的疑慮時。

因此在臨床效能評估指引<sup>[3]</sup>中，FDA 提供設計與執行臨床效能評估的建議以供製造商參考及依循，並建議製造商在開始執行獨立及臨床效能評估前，先提供 FDA 試驗計畫書進行審查。

### 3.1 臨床試驗設計(Clinical Study Design)

臨床效能評估的目的是呈現 CADe 器材在其預期用途下，其於臨床上的安全性及有效性。因 CADe 在設計上，是將其輸出結果交由其使用者(即臨床判讀人員)進行判讀，故 CADe 器材輸出結果與臨床判讀人員間具有相當複雜的關係；有鑑於此，FDA 建議任何 CADe 器材的臨床試驗設計，皆需考量以下幾項重點。

1. 使用者參與判讀 CADe 器材輸出結果的時間點，例如：使用者為同時進行判讀或進行二讀(second read)；
2. CADe 器材輸出標記的物理特徵，例如大小、形狀、標記邊緣的形式(實線、虛線、圓圈、等高線形式等)、輸出標記與異常區域的接近程度；
3. 使用者對 CADe 器材標記為異常區域的相關知識；
4. CADe 輸出標記的數量。

臨床效能評估應使用控制良好的試驗設計，包含或限制可能影響 CADe 器材安全性及有效性的偏差，尤其是臨床效能評估因不易複製臨床環境，而導致需在實驗室中進行時(即並非在臨床場域執行)。下列試驗設計可用於評估 CADe 器材的臨床效能：

1. 實地試驗(field test)或前瞻性判讀者試驗(pro prospective reader study)：此類試驗可評估 CADe 器材在真實臨床情況下的效能，例如隨機控制試驗(randomized controlled trial)。然而，當目標疾病的盛行率低時，常需要增加病患的收納人數，因此有時在實務上實地試驗並不容易執行。
2. 回溯性判讀者試驗(retrospective reader study)：回溯性地收集多量的疾病/異常案例，作為實地試驗的替代方法。



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

3. 壓力試驗(stress study)：壓力試驗亦為一種回溯性試驗，是在病患案例中納入具挑戰性的影像特徵(或其他影像資料)，而這些高挑戰性的影像在一般臨床實務上並不常見，但仍落在預期目標族群中。應注意人為選擇擴增任何案例，皆可能造成 CADe 器材在真實臨床實務與效能評估試驗中的效能差異。

在臨床效能評估中，案例的判讀建議採用多判讀人員/多案例(multiple reader multiple case, MRMC)的設計，即一組臨床判讀人員在多種判讀情境下(例如：進行不經 CADe 器材協助的判讀，與經 CADe 器材協助的判讀)，進行放射影像的評估。每位判讀人員在所有情境下獨立判讀每件案例，且所有的判讀人員皆需完成判讀所有的案例(fully-crossed)。MRMC 設計可在給定的案例數量下，達到最大的統計檢定力。某些情況下，每位判讀人員要完成所有案例的判讀並不容易實行(non-fully crossed)，但仍可被接受。不論是否完成判讀所有案例，FDA 皆建議採用 MRMC 方法搭配上所述的試驗設計來進行臨床效能評估，並預先提供統計分析方法的描述如下：

1. 試驗設計的描述；
2. 影像數據資料收集方法的描述；
3. 試驗計畫書，包含：
  - 試驗假說與試驗評估指標；
  - 任何用以確認統計假設之驗證計畫；
  - 未能滿足統計假設時的替代方法或程序；
  - 試驗成功的標準，即假說應符合之條件，以此判定臨床試驗成功與否；
  - 敘明選擇案例數目在統計及臨床上的合理性；
  - 敘明選擇判讀人員數在統計及臨床上的合理性；
  - 影像判讀方法與臨床實務的關聯性；
  - 隨機取樣的方法；
  - 判讀人員的執行任務(包含計分標準)。
4. 判讀人員之資格及其經驗；
5. 判讀人員訓練之描述；
6. 統計分析計畫，包含：
  - 辨真過程；
  - 計分技術的細節；



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

- 任何支持該試驗設計的前導試驗結果。

完善的臨床效能評估試驗設計，則應包含：

1. 研究族群(包含正常與異常案例)可適當地代表目標族群；
2. 試驗設計應排除干擾 CADe 器材效能的干擾因素；
3. 收納案例數的規模應足以代表效能宣稱；
4. 辨真過程需適合用於評估效能，且試驗分析已確切地說明參考標準的不確定性；
5. 資料組中具有合適的資料族群；
6. 判讀人員的選擇足以代表預期的使用者；
7. 影像硬體與目前臨床實務具有一致性。

### 3.1.1 評估模型與試驗評估指標(Evaluation Paradigm and Study Endpoints)

為了證明 CADe 器材的有效性，應選擇適當的試驗終點(study endpoints)，即試驗的評估指標。主要及次要評估指標的選擇，須依 CADe 器材的預期用途而有所不同，且試驗評估指標應在評估開始前即確認且不可變動。一般而言，臨床上的試驗評估指標會採用 CADe 器材的敏感性(Se)、特異性(Specificity, Sp)，以及接收者操作特徵曲線(Receiver operating characteristic curve，ROC 曲線)來定義，或是採用 ROC 曲線的變形曲線，如自由反應接收者操作特徵曲線(free-response receiver operating characteristic curve，FROC 曲線)或定位接收者操作特徵曲線(location-specific receiver operating characteristic curve，LROC 曲線)。敏感性(Se)即為真陽性率，特異性(Sp)即為真陰性率，ROC 曲線乃為 Se 對 1-Sp (即偽陽率)作圖之曲線，可呈現 CADe 器材或臨床判讀人員診斷效能的總結摘要。而其變形之 FROC 曲線，則是以 Se 對每組放射影像中偽陽性標記的數目(FPs/case)作圖，適用於分析中需要考慮每位病患的疾病位置及可能有複數疾病區域時總結診斷效能。

評估判讀人員使用 CADe 器材的效能時可以採用不同的總結指標，例如：

1. ROC 曲線的面積、部分面積或該曲線的其他量測指標；
2. FROC 曲線的面積、部分面積或該曲線的其他量測指標；
3. LROC 曲線的面積、部分面積或該曲線的其他量測指標；
4. 判讀人員的 Se 及 Sp (或 Se 及病患召回率，此處病患召回率是指正常及異常



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

的病患後續被要求進行額外臨床處置的比例)；

5. 判讀人員標記影像上異常區域的準確度。

FDA 建議使用 ROC 總結效能指標作為主要試驗評估指標，同時以 Se 及 Sp 作為次要試驗評估指標，提供 Se 及 Sp (或 Se 及病患召回率)可對 CAde 器材在臨床實務上的影響有更多的了解。若使用不同的效能指標，則建議先與 FDA 聯繫確認。

FDA 建議應描述統計評估的方法，並提供下列結果：

1. 整體判讀人員的效能；
2. 干擾因素或修飾作用(例如病變型態、病變大小、病變位置、掃描程序、影像硬體、共病)的分組分析效能；
3. 判讀人員、案例及辨真的變異性或其他可能存在的變異性的信賴區間。

同時 FDA 亦建議需鑑別及驗證在效能評估時所使用的分析軟體，且真陽性、真陰性、偽陽性及偽陰性 CAde 器材輸出標記的定義應與 CAde 器材的預期用途及參考標準一致。

### 3.1.2 控制組(Control Arm)

進行臨床效能評估時應包含控制組，一般以傳統臨床判讀結果(即無 CAde 器材協助下的判讀結果)作為 CAde 器材臨床效能評估的控制組最為常見。若 CAde 器材預期作為第二判讀者(second reader)，則可採用兩位判讀人員分別判讀做為控制組。一般採用這些控制組或將類似品作為控制組進行直接比較皆適用於證明實質等同性，或者使用其他控制組亦可能被接受。FDA 建議在進行臨床評估前，先與 FDA 討論控制組的合適性。

CAde 器材臨床效能評估的控制組，應該與效能評估組使用相同的判讀方法且與臨床實務具有一致性。所有的組別應含有相同的樣本族群，以降低評估上可能的偏差；當各組別中含有不同的案例時，應提供如何隨機指派病患及判讀人員的描述及流程圖。

### 3.1.3 判讀情境與隨機性(Reading Scenarios and Randomization)

臨床效能評估中的判讀情境，應與 CAde 器材的預期用途具有一致性。當進行臨床效能評估時，FDA 建議如下的判讀情境：

1. CAde 器材僅作為第二判讀者(次序性的判讀設計)時：



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

- 無器材協助下進行傳統的臨床判讀(即僅由判讀人員判讀結果)；
  - CADe 器材進行第二判讀，並在人員判讀後立即顯示 CADe 器材的判讀結果(此第二判讀情景可與傳統臨床判讀情境在同一時段進行)。
2. CADe 器材用於同步判讀(交叉性的判讀設計)時：
- 無器材協助下進行傳統的臨床判讀(即僅由判讀人員判讀結果)；
  - CADe 器材與人員進行同步判讀，而 CADe 器材判讀結果可在判讀過程的任一時間點呈現(此同步判讀情景應與傳統臨床判讀情境在不同的時段進行)。

為了降低效能評估時的偏差，應隨機分配判讀人員、案例及判讀情境，FDA 建議提供隨機分配方法的描述及流程圖。如果有多個相同的案例在不同的判讀時段進行判讀時，每個判讀時段應間隔至少四週以避免人員記憶造成的偏差。FDA 建議採用更長的時間間隔，且無論採用更短或更長的時間間隔，皆應提供支持此時間間隔的佐證資料。

### 3.1.4 評分級距(Rating Scale)

進行效能評估時，應採用傳統的醫療判斷並報告病變的位置、範圍及病患的管理。當判讀人員在進行病患病變位置及疾病狀態的評分時，若使用 ROC 曲線的試驗評估指標，則可採用更細緻的評分級距來收集資料(例如 7 分或 100 分級距)。FDA 建議對採用的評分級距提供判讀人員訓練。

### 3.1.5 計分方法(Scoring)

在臨床效能評估中，與參考標準比較所進行的計分，其評估對象是臨床判讀人員所作之標記(即臨床判讀人員的判讀結果)。FDA 建議應清楚描述「臨床判讀人員判讀結果與參考標準之間是否相符合」的計分過程，包含其合理性、定義與允收條件。臨床評估的計分方法，應與 CADe 器材標記的異常及器材的預期用途具有一致性，且在評估試驗開始前即須選定及描述清楚計分過程。FDA 建議依據前述「器材描述-(九) 計分方法」中之說明，清楚描述臨床效能評估之計分方法。

### 3.1.6 臨床判讀人員的訓練(Training of Clinical Readers)

FDA 建議應針對 CADe 器材的使用及如何參與此臨床評估，提供臨床判讀人員適當的說明與訓練，且應包含器材的描述及如何使用的說明。針對特殊的判讀說明或規



則(例如當得知 CADe 標記結果後改變原先判讀結果的規則)，應描述其與判讀任務、臨床流程及醫療實務的相關性。

FDA 也建議應提供判讀人員關於評分級距的訓練，尤其是當使用臨床實務上不常見的評分級距時。這些訓練將可避免不正確或無法判讀的結果產生。判讀人員訓練應包含根據試驗設計方法來評估一組具有代表性的正常及異常案例，且這些案例不應被納入在臨床評估的資料庫中。

### 3.2 研究樣本群(Study Population)

病患的資料(例如影像案例)可採用良好定義的納入/排除條件，以前瞻性或回溯性的方式，從 CADe 器材預期用途的目標族群中收集，使案例組包含適當範圍的疾病/異常及正常案例，並於申請時提供案例收集的計畫書。所有臨床評估所用的案例，皆必須獨立於 CADe 器材開發時所用的案例。從參與收集案例的醫療機構中，連續性地收集符合允收條件及排除條件的案例，是較被接受的一種收集方式。

其他有關研究樣本群的考量請參考獨立效能評估段落之「2.1 研究樣本群」。

#### 3.2.1 數據資料的整合性(Data Poolability)

當使用非美國的數據資料時，應說明非美國資料用於美國人族群的適用性，包含疾病發生率、疾病特徵、醫療實務及臨床人員勝任性。同時為了設計良好的臨床試驗，應就統計及臨床的角度說明從多個醫療機構收集數據資料的整合性。建議當使用非美國臨床資料時，先與 FDA 聯繫。

#### 3.2.2 測試數據資料的重複使用(Test Data Reuse)

與前面描述獨立效能評估重複使用資料之情況相同，當採用重複使用的資料時，判讀人員及 CADe 演算法可能會明顯或隱晦地產生優化，而造成效能評估的偏差。當兩次試驗間沒有演算法訓練時，若欲重複使用數據資料，則可由一組對前次測試結果完全未知的判讀人員來進行判讀。但當試驗之間有演算法訓練，或是為不同的演算法時，FDA 建議採用新收集的獨立資料進行評估。原則上，FDA 並不建議在 CADe 器材效能評估時採用重複的數據資料；若考慮重複使用測試資料，應證明這些資料任一部分的重複使用，在評估 CADe 器材效能時皆不會導致非預期的偏差，且仍保有測試資料的可靠性。



台灣藥物法規  
資訊網法規公告



台灣藥品  
臨床試驗資訊



FDA  
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

其他有關測試數據資料的重複使用的考量，請參考獨立效能評估段落之「2.2 測試數據資料的重複使用」。應注意的是，在建立數位稽核軌跡時，除 2.2 段落所要求實行之 5 項控制外，另應在每個臨床效能評估中使用新的判讀人員。

### 3.3 參考標準(Reference Standard)

請參考前述「器材描述-(八) 參考標準」提出參考標準選擇的合理性及其描述。

### 3.4 結果報告(Reporting)

效能評估結果報告之格式與內容，可依循 FDA 發布的評估結果報告統計指引<sup>[6]</sup>撰寫。FDA 建議以電子方式提交統計分析所使用的任何資料，並應包含：

1. 病患資訊；
2. 疾病或正常狀態；
3. 共病；
4. 病變大小；
5. 病變型態；
6. 病變位置；
7. 疾病病程；
8. 器官特徵；
9. 影像硬體；
10. 造影或掃描的程序；
11. 影像及資料的特徵；
12. 統計分析。

### 3.5 類似品比較(Comparison with Predicate Device)

FDA 建議將 CAdE 器材的臨床效能評估結果，與其宣稱實質等同類似品(例如改版的前一版本)的臨床效能評估結果進行比較，以佐證其實質等同性。但在臨床效能評估中將 CAdE 器材與類似品進行直接比較，有些時候不一定可行；故亦可將 CAdE 器材臨床效能評估的統計結果，與類似品在相似的臨床效能評估中的統計結果進行比較，作為無法直接比較時的替代方案。此兩者在各自的臨床效能評估中，相較於各自的控



致力法規科學  
守護生命健康

Regulatory Science, Service for Life

制組應在統計上具有更好的效能，例如使用這些 CADe 器材協助判讀人員判讀的結果，在統計意義上應優於未使用這些 CADe 器材協助判讀人員判讀的結果。

為了幫助讀者了解上述 CADe 器材臨床效能評估試驗的重點，於表二彙整 FDA 臨床效能評估指引中之各項主要建議。

表二、美國對 CADe 器材臨床效能評估的建議

項目	建議	注意事項
試驗設計	<ul style="list-style-type: none"> <li>● 採用前瞻性、回溯性或壓力試驗設計</li> <li>● 採用 MRMC 方法判讀案例</li> </ul>	試驗應具備合適的影像硬體、判讀人員、案例、參考標準等，足以代表臨床實務及其宣稱的預期用途
評估模型與試驗評估指標	<ul style="list-style-type: none"> <li>● ROC 總結效能指標作為主要試驗評估指標</li> <li>● Se 及 Sp 作為次要試驗評估指標</li> </ul>	應描述統計分析的方法，包含整體效能、分組效能及變異數的信賴區間
控制組	<ul style="list-style-type: none"> <li>● 採用無 CADe 器材協助的判讀結果做為控制組</li> <li>● 或將類似品作為控制組進行直接的比較來證明實質等同</li> </ul>	所有組別應採用相同的統計分析方法及樣本族群，且與臨床實務具有一致性。
判讀情境與隨機性	<ul style="list-style-type: none"> <li>● 依預期用途採用次序性或交叉性判讀情境</li> <li>● 應隨機分配判讀人員、案例及判讀情境，並說明隨機分配方法</li> </ul>	當有多個不同的判讀時段而產生相同的案例在不同時段進行判讀時，每個判讀時段應間隔至少四週
評分級距	<ul style="list-style-type: none"> <li>● 採用傳統醫療評分判斷</li> <li>● 基於 ROC 曲線的試驗評估指標可支持用更細緻的評分級距來收集資料</li> </ul>	應就評分級距提供判讀人員訓練
計分方法	<ul style="list-style-type: none"> <li>● 清楚描述計分過程的合理性、定義及允收條件</li> <li>● 依建議說明計分方法採用的方式</li> </ul>	計分過程應與 CADe 器材標記的異常及器材的預期用途具有一致性



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

臨床判讀人員的訓練	<ul style="list-style-type: none"><li>● 應提供判讀人員此 CDe 器材的描述及使用說明</li><li>● 訓練應包含具代表性的正常及異常案例</li></ul>	訓練判讀人員的案例不可用於臨床效能評估之資料庫
研究樣本群	<ul style="list-style-type: none"><li>● 較常從參與收集案例的醫療機構中連續性地收集符合允收及排除條件的案例</li><li>● 應從預期用途族群中選擇案例，且研究樣本族群(及/或子群組)應有足夠的案例數以證明其效能宣稱</li></ul>	臨床評估之案例不可與開發器材時所用案例重複
數據資料的整合性	<ul style="list-style-type: none"><li>● 應說明非美國人族群資料的合適性</li><li>● 應說明從多個機構收集案例的資料整合性</li></ul>	
測試數據資料的重複使用	<ul style="list-style-type: none"><li>● 不建議重複使用數據資料</li><li>● 若重複使用資料應證明不會產生非預期的偏差且具有資料的可靠性，並建立防火牆及數位稽核軌跡來限制人員得知效能評估結果</li></ul>	數位稽核軌跡應依要求進行控制
參考標準	<ul style="list-style-type: none"><li>● 說明參考標準定義程序的合理性及依要求進行描述</li></ul>	當多位判讀人員參與時，應描述如何總結多個判讀結果成為單一參考標準；以及當多個判讀結果不一致時將如何處理
結果報告	<ul style="list-style-type: none"><li>● 依要求以電子方式提供統計分析所用的資料</li></ul>	
類似品比較	<ul style="list-style-type: none"><li>● 應與其宣稱實質等同的類似品進行臨床效能評估的比較</li></ul>	可在同一臨床效能評估試驗中進行直接的比較；或將 CDe 器材在臨床效能



評估中的統計結果，與類似品在相似的臨床效能評估中的統計結果進行比較

#### 4 獨立效能評估與臨床效能評估之比較

FDA 在醫療器材軟體的臨床評估指引<sup>[7]</sup>中，說明製造廠可依循三個步驟：有效的臨床關聯性、分析/技術確效及臨床確效，來獲得合適的臨床評估資料<sup>[8]</sup>。而 CAdE 器材作為以演算法軟體為核心的醫療器材，藉由 510(k)途徑申請與類似品的實質等同來證明有效的臨床關聯性、經由獨立效能評估及臨床效能評估來取得分析/技術確效及臨床確效，亦符合醫療器材軟體的臨床評估指引<sup>[7]</sup>的要求。

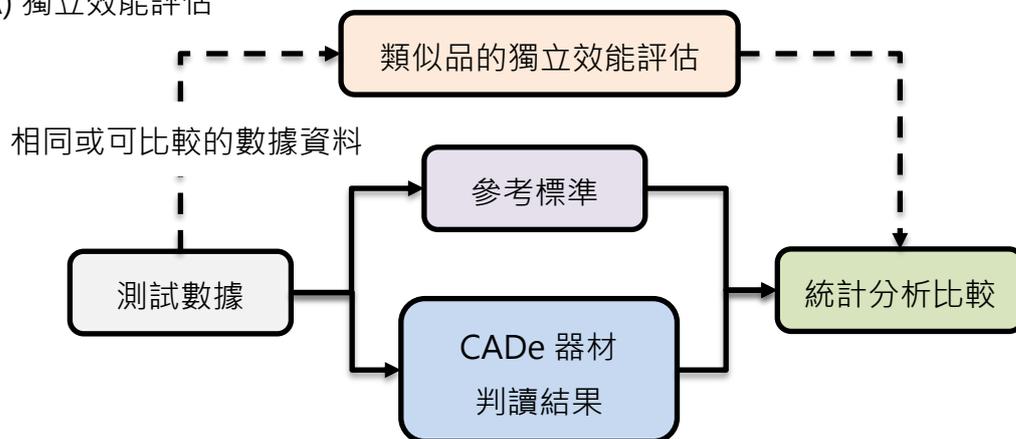
作為 CAdE 器材效能評估的重點，獨立效能評估與臨床效能評估各自具備不同的評估目的。獨立效能評估是將測試資料經 CAdE 器材偵測結果與參考標準進行統計分析，藉以評估其基本效能(包含信賴區間)，並與類似品的獨立效能評估進行比較以協助實質等同性的判定。而臨床效能評估則是將判讀人員經由 CAdE 器材協助判讀的結果(實驗組)與判讀人員未經 CAdE 器材協助的判讀結果(控制組)、以及參考標準三者之間的統計結果進行比較，以得到 CAdE 器材在真正臨床上協助其預期使用者判讀時的臨床效能，並與類似品的臨床效能評估進行比較，以證明其實質等同性(圖一)<sup>[9]</sup>。



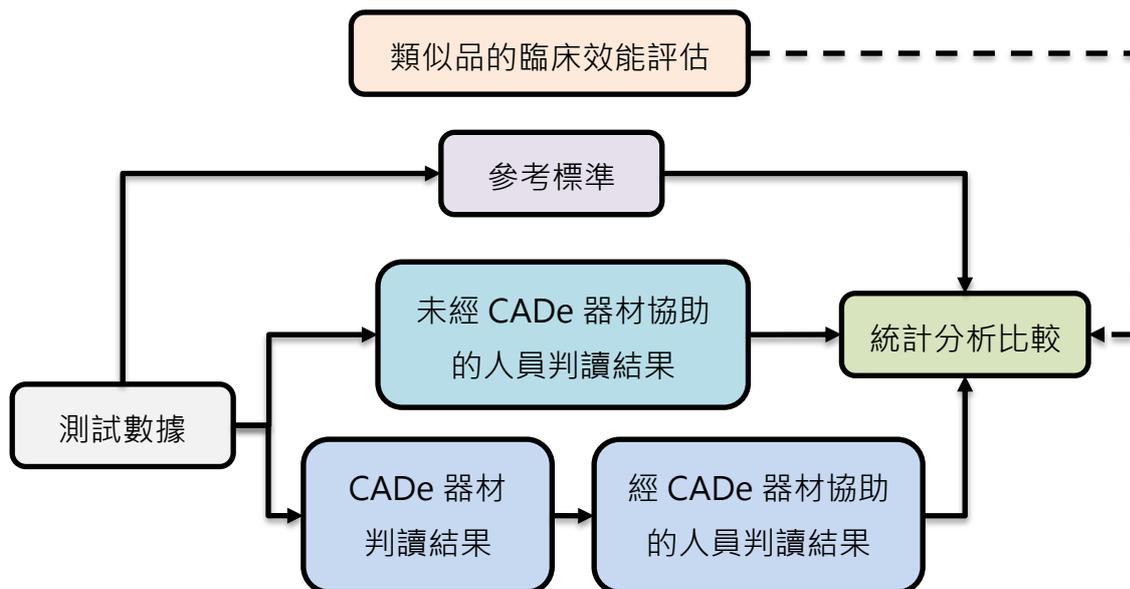
致力法規科學  
守護生命健康

Regulatory Science, Service for Life

(A) 獨立效能評估



(B) 臨床效能評估



圖一、CADe 器材的效能評估流程<sup>[9]</sup>。(A) 獨立效能評估；(B) 臨床效能評估

討論

雖然 CADe 器材的申請 510(k)上市指引<sup>[1]</sup>與臨床效能評估指引<sup>[3]</sup>不適用於 CADx 及 CADt 器材，但這幾類 AI/ML 技術的電腦輔助醫療器材軟體，應用於放射影像上時具有相似的工作流程，即由影像數據輸入、經演算法判讀、再輸出結果標記的特點，因此不難從 FDA 在上述兩份指引的各項建議中，理解 CADx 及 CADt 器材在效能評估上的幾個重點，包含：



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

- 從預期用途的目標族群中收集足夠數量的樣本來建立測試數據資料庫；
- 建立樣本的參考標準；
- 建立計分方法來定義參考標準與輸出結果之間的關係；
- 事先擬定適當的試驗評估指標，用以解讀效能評估的結果；
- 執行獨立效能評估與臨床效能評估，並與類似品比較。

故雖然目前 FDA 尚未擬定 CADx 及 CADt 器材效能評估的相關指引，但上述這些 CADe 器材效能評估要求的重點，仍可作為製造商開發及評估 CADx 及 CADt 器材效能時的參考。

## 結語

由本文可知，美國 FDA 對於電腦輔助偵測軟體效能評估的重點，主要分成獨立效能評估及臨床效能評估兩部份，且針對兩種效能評估在執行之前，應針對評估過程預先擬定合適的計畫，並先定義清楚如何判定評估結果。FDA 建議所有 CADe 器材皆應進行獨立效能評估及臨床效能評估，若製造商認為 CADe 器材不需要進行臨床效能評估時，應先向 FDA 尋求建議。近年我國在資通訊產業上的堅強實力以及在智慧醫療領域的投入，各種 AI/ML 技術相關的醫療器材開發蓬勃發展。本篇文章提供我國製造商在電腦輔助醫療器材軟體的法規參考，期盼得以幫助我國製造商在相關的醫療器材開發上更具備競爭力。

## 參考文獻

1. US FDA. Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions. Guidance for Industry and Food and Drug Administration Staff. Document issued on July 3, 2012.
2. US FDA. Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data – Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions. Guidance for Industry and FDA Staff. Document issued on July 3, 2012.



3. US FDA. Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data in - Premarket Notification (510(k)) Submissions. Guidance for Industry and FDA Staff. Document issued on January 22, 2020.
4. US FDA. Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices. Guidance for Industry and FDA Staff. Document issued on May 11, 2005.
5. US FDA. Off-The-Shelf Software Use in Medical Devices. Guidance for Industry and Food and Drug Administration Staff. Document issued on September 27, 2019.
6. US FDA. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests - Guidance for Industry and FDA Staff. Document issued on March 13, 2007.
7. US FDA. Software as a Medical Device (SAMM): Clinical Evaluation. Guidance for Industry and Food and Drug Administration Staff. Document issued on December 8, 2017.
8. 醫療器材軟體之臨床評估. 當代醫藥法規月刊 97 期. 2018 年 11 月.
9. US FDA. Nicolas Petrick. Pre-and Post-Market Evaluation of Autonomous AI/ML: Lessons Learned From Prior CAD Devices. Public Workshop - Evolving Role of Artificial Intelligence in Radiological Imaging. Workshop on February 25-26, 2020.