衛生福利部
Ministry of Health and Welfare

臺灣 AI 臨床研究倫理審查常見問題說明

衛生福利部資訊處 李建璋 處長

Q1、目前在臺灣已有一個機制，稱為 CIRB，主要用於加速藥物臨床試驗的倫理審查流程， 這樣的審查機制，是否也適合延伸應用到 AI 相關研究？特別是像 AI 模型驗證（validation）、臨床效益評估（clinical impact study），還是模型開發（model development）的研究設計。 AI 研究在本質上與傳統藥物試驗不同，例如具有動態更新、資料依賴性高、以及跨機構資料整合等特性。在這樣的前提下，是否適合沿用 CIRB 這種以「一次性審查、固定 protocol」為核心的機制？

現行的 CIRB（聯合人體試驗委員會）機制是為「標準化、不輕易變動」的流程設計的。根據這個特性，這三種 AI 研究的適合程度如下：

一、 外部驗證研究（Validation Study）— 最適合

- 原因：這類研究通常是「收案、去識別化、跑模型、比對結果」，流程極度標準化。
- CIRB 優勢：AI 模型需要跨院（Multicenter）數據來證明通用性，CIRB 能一次解決多間醫院的行政審查，大幅加速數據收集與驗證的效率。
- 關鍵：只要模型版本在驗證期間保持固定，它是目前最能直接套用 CIRB 的類別。

二、 實務叢集隨機對照試驗（Pragmatic Cluster RCT）— 部分適合

- 原因：這類試驗是在真實臨床流程中比較「有 AI」與「無 AI」的差異，通常以「病房」或「醫院」為隨機分配單位。
- CIRB 挑戰：雖然多中心審查有助於擴大規模，但每間醫院的臨床流程（Workflow）不同，AI 嵌入的方式可能需微調。
- 結論：適合使用 CIRB 建立核心倫理架構，但各醫院仍需保留對其「流程變更」的最終核准權。

三、 AI 模型開發研究（Development Study）— 最不適合

- 原因：開發階段具有高度迭代性（Iterative）。工程師可能每天都在調整參數、更換特徵（Features）或清洗數據。
- CIRB 衝突：傳統 CIRB 要求「一次性核准固定 Protocol」。如果每次微調模型都要跑一次中央變更申請（Amendment），研發進度會被行政流程拖垮。
- 建議：這類研究較適合由單一機構的 IRB 進行彈性滾動式審查，而非透過集中的 CIRB 處理。

研究類型	適合程度	核心理由
外部驗證	★★★★★	流程固定，最需跨院數據整合。
實務 RCT	★★★	倫理框架一致，但需兼顧各院流程差異。
模型開發	★	變動太快，行政審查速度趕不上開發速度。

Q2、在傳統推論型 AI 中，外部驗證相當重要。因此我們於衛生福利部設置了 AI 外部驗證中心，用於收集各類疾病的病患資料，以建立兼具族群平衡與區域平衡的全國性代表性資料集，用來支持 AI 的驗證。請問在這種多中心研究情境下，倫理審查的重點為何？又應如何加速審查流程？

當驗證中心具備「資料合成」、「族群平衡」與「區域平衡」等功能時，IRB 委員會最關注以下面向：

一、 資料代表性與公平性 (Algorithmic Fairness)

- 審查重點：驗證中心如何定義「平衡」？是否涵蓋了罕見疾病、偏鄉族群或不同醫療體系（醫學中心 vs. 基層診所）的資料，以避免 AI 在落地後產生偏見。

二、 去識別化與重識別風險 (Re-identification Risk)

- 審查重點：即使是合成資料 (Synthetic Data)，仍需審查其原始特徵是否可能被「反向推論」出真實病患。必須有第三方技術審計確保去識別化程度。

Q3、AI 可分為傳統推論型（例如 XGBoost、CNN）、生成式 AI（Generative AI），以及代理人 AI（Agentic AI）。請問這三種類型的 AI，在倫理審查上的重點分別為何？

研究階段	推論型 AI（XGBoost, CNN）	生成式 AI（GenAI, LLM）	代理人 AI（Agentic AI）
1. 模型開發 (Development)	重點：**資料正義性** 1. 訓練資料的代表性（有無偏見）。 2. 標註者的資歷與一致性。 3. 原始資料的去識別化與隱私授權。	重點：**隱私與真實性** 1. 避免訓練資料包含敏感個資（防禦逆向推論）。 2. 基礎模型（Base model）的授權與合規性。 3. 幻覺風險的初步測試。	重點：**任務範圍與安全邊界** 1. 定義 Agent 可自主執行的「任務清單」。 2. 推理過程的記錄機制（Traceability）。 3. 模擬環境（Sandboxing）中的安全測試。
2. 模型驗證 (Validation)	重點：外部效能與可解釋性 1. 跨院驗證的穩定度。 2. 醫師是否能理解決策邏輯（SHAP/LIME 等工具）。 3. 效能基準（Baseline）的定義。	重點：產出安全性與幻覺監控 1. 用「挑戰者」的角度去測試：生成錯誤資訊（hallucination） 2. 輸出有害內容（例如仇恨言論） 3. 被誘導違反規範（例如教人做違法事情） 4. 洩漏敏感資訊。	重點：決策鏈與干預機制 1. 測試 Agent 在複雜場景下的決策路徑。 2. 人工接管（Override）的觸發速度測試。 3. 長時間運作下的邏輯穩定性。

		5. 醫療正確性與專家共識對比。 6. 輸出過濾機制（Filter）的有效性。	
3. 模型落地 (Implementation)	重點：臨床效益與偏移監測 1. 對醫師決策行為的影響。 2. 效能漂移（Drift）的定期通報流程。 3. 輔助診斷與最終醫療責任的歸屬。	重點：責任歸屬與醫病溝通 1. 病患是否有權知悉正在與 AI 互動。 2. 醫師對產出內容的簽章核對（Final Sign-off）機制。 3. 誤導性醫學建議的即時下架機制。	重點：自主權限與問責體系 1. 建立權限分級（例如：輔助建議 vs. 自主執行）。 2. 事件發生時的「決策黑盒子」還原機制。 3. 系統自主權對醫療常規造成的倫理衝擊評估。

Q4、關於資料去辨識以及資料不出院的原則，請問在美國，電子病歷相關研究的倫理審查是如何進行的？其具體要求為何？對這些標準是否有嚴格的規範？

一、美國電子病歷相關研究的倫理審查作法

在美國，電子病歷（EHR）相關研究的倫理審查主要由 HIPAA（醫療保險流通與責任法案） 與 Common Rule（聯邦人體試驗保護法規） 兩大架構共同規範。「資料去辨識」與「資料不出院」原則，美國的執行邏輯非常嚴謹且高度標準化，其核心在於「風險分級」與「法律責任」。

i. 安全港法（Safe Harbor Method）— 最嚴格的清單法

- 要求：必須移除 18 項特定的個人辨識碼（包括姓名、詳細地址、電話、傳真、所有早於「年」的日期、MRN、身分證號等）。
- 特性：這是最常見的路徑。其標準極度明確，IRB 審查時只要確認清單中的 18 項確實移除，程序非常快。
- 缺點：會大幅降低資料的臨床研究價值（例如無法分析 30 天內的再入院率，因為日期被刪除）。

ii. 專家判定法（Expert Determination Method）— 基於風險的統計法

- 要求：由具備統計學背景的專家證明「剩餘資料被重新辨認的風險極小（Very Small Risk）」。
- 特性：允許保留部分研究必要的細節（如具體入院日期）。這在 AI 模型開發中非常常見，但需要專家的統計證明報告供 IRB 備查。

二、 資料不出院原則：有限資料集 (Limited Data Set, LDS)

當研究需要保留較多資訊（如 5 碼郵遞區號或完整日期），導致無法完全達到上述去辨識標準時，美國通常採取 「資料不出院 / 受控輸出」 的邏輯：

i. 簽署資料使用協議 (Data Use Agreement, DUA)：

○ 這是法律文件。研究者必須承諾資料僅用於特定目的、不嘗試重新辨認身分、且具備完善的安全保護措施。

ii. 電子安全邊界 (Digital Firewall)：

○ 許多醫學中心（如 Stanford 或 Mayo Clinic）會建置 「安全研究環境」(Secure Research Environment)。研究者透過遠端桌面登入醫院伺服器運算，「原始資料不出院，僅產出結果可下載」。這與您在衛福部推動的外部驗證中心邏輯高度一致

三、 倫理審查（IRB）的具體流程與要求

美國 IRB 在審查 EHR 研究時，重點不在於「是否利用資料」，而在於「知情同意的豁免」。

- 審查重點：
  - ○ 隱私風險極小：是否有完善的去辨識化或 DUA 協議？
  - ○ 研究不具可行性：若不豁免知情同意（例如要找回十年前的萬名病患簽字），研究是否就無法進行？
  - ○ 對受試者權益無不良影響。
- 嚴格程度：非常嚴格。如果研究者未經許可將 PHI 帶離醫院，醫院將面臨巨額民事罰款甚至刑事責任，因此醫院 IRB 通常會要求使用醫院內部的安全運算平台。

項目	**美國標準（HIPAA/Common Rule）**	**臺灣現狀/趨勢**
去辨識標準	Safe Harbor（18項）或 專家判定	依個資法及衛福部去識別化指引
日期處理	嚴格限制（Safe Harbor 僅能留年份）	臨床研究常需具體日期，審查較彈性

資料外傳	必須簽署 DUA，否則極難跨機構	去辨識並詳細說明資料保管，使用權限，使用目的。
技術配套	強調 Secure Enclave（研究沙盒）	規劃推動衛福部資料研究用多中心電子病歷暫存伺服器

Q5、如果是務實性隨機臨床試驗（pragmatic randomized trial），且採用群組隨機設計（cluster randomized trial），在許多情況下逐一取得個別受試者的知情同意有困難。請問目前國際上通行的倫理原則為何？特別是當研究以醫師為單位進行隨機分配（例如分為使用 AI 輔助與未使用 AI 輔助兩組）時，在這種情境下，個別病人是否仍需要簽署知情同意書？此外，國際上（如倫理指引或專業學會）對於此類 cluster randomized trial 的倫理審查作法為何？

針對「群組隨機對照試驗（Cluster Randomized Trial, CRT）」且以「醫師/單位」為隨機分配對象的情境，國際通行的倫理指引主要參考 《渥太華聲明》（The Ottawa Statement）。

一、 核心倫理原則：誰才是「受試者」？

在 CRT 中，倫理審查的第一步是界定「受試者（Human Subject）」的身分。這直接決定了是否需要取得知情同意（Informed Consent）：

- 醫師作為受試者：如果研究僅是觀察 AI 是否改變醫師的開藥行為、診斷準確度，且不涉及收集病人的私密個資，醫師才是主要受試者。此時通常僅需取得醫師同意。

- 病人作為受試者：如果研究會收集病人的預後（如住院天數、死亡率）或介入措施直接影響病人的健康風險，病人即被視為受試者。

二、 何時可以「豁免個別知情同意（Waiver of Consent）」？

根據 CIOMS 指引 與美國 Common Rule，若 CRT 符合以下條件，IRB 通常允許豁免個別病人的書面同意：

i. 風險極小（Minimal Risk）：AI 僅作為「輔助決策」，最終決策權仍在醫師手中，且該 AI 建議符合現行臨床指引（Standard of Care）。

ii. 研究不具可行性（Impracticability）：在務實性試驗中，若要逐一對數萬名門診病人進行招募與簽署，會產生嚴重的「選擇性偏差（Selection Bias）」，導致研究結果失去科學價值。

iii. 不影響受試者權益：豁免同意不會對病人的治療權利或隱私造成實質損害。

iv. 專業實務： 在以醫師為單位的 AI 研究中，常見的做法是「告知而非取得簽署」。例如在診間張貼告示或在掛號系統顯示：「本院正進行 AI 輔助診斷研究，您的醫師可能會參考 AI 建議，如您有疑慮可告知醫師」。

三、 國際專業學會與指引的審查作法

針對 CRT 的特殊性，國際上有幾項關鍵的審查標準：

i. 《渥太華聲明》（The Ottawa Statement）

這是全球首部專門規範 CRT 倫理的指引，強調：

- 區分介入對象：區分「群組成員（病患）」與「群組守門人（醫護/院長）」。
- 群組守門人（Gatekeepers）的角色：當個別病患難以簽署時，必須由醫院主管或科主任作為守門人，審查該研究是否符合該群體（如全體病患）的利益。

ii. 學習型醫療系統（Learning Healthcare System）框架

美國醫學研究院（NAM）提倡，當 AI 驗證已成為提升醫療品質的「常規活動」時，倫理審查應趨向簡化。只要資料受到嚴格保護，且 AI 介入不超過臨床標準，應視為「品質改善（Quality Improvement, QI）」而非傳統研究，從而簡化同意程序。

比較項目	傳統藥物 RCT	AI 務實性群組試驗（pCRT）
隨機分配單位	個人（Individual）	群組（Cluster, 如醫師、診間、病房）
研究環境	高度受控的實驗環境（Explanatory）	真實臨床工作流（Pragmatic / Real-world）
介入性質	實驗性藥物或侵入性治療	資訊介入（AI 建議）
知情同意	必須逐一簽署	常採「豁免簽署」或「告知選擇制」
風險程度	通常較高（藥物副作用未知）	通常為極小風險（Minimal Risk）
受試者定義	僅為病患	包含醫師（介入對象）與 病患（預後對象）
資料來源	專案收集的 CRF 數據	自動化提取 電子病歷（EHR）數據
主要倫理挑戰	受試者安全性與臨床盲性	自主權豁免的理據與群體公平性