

真實世界數據：  
評估電子健康紀錄和醫療給付資  
料以支持藥品法規決策指引

中華民國 112 年 10 月

## 目錄 (Table of Contents)

|    |                                                     |    |
|----|-----------------------------------------------------|----|
| 一、 | 前言 .....                                            | 4  |
| 二、 | 目的 .....                                            | 6  |
| 三、 | 概論 .....                                            | 8  |
| 四、 | 資料來源.....                                           | 10 |
|    | (一) 資料來源的適用性.....                                   | 10 |
|    | (二) 資料擷取(Data Capture)：整體討論.....                    | 11 |
|    | (三) 缺失資料(Missing Data)：整體考量 .....                   | 22 |
|    | (四) 驗證(Validation)：整體考量 .....                       | 24 |
| 五、 | 研究設計要素 .....                                        | 30 |
|    | (一) 時間區間(Time Period)的定義 .....                      | 30 |
|    | (二) 研究目標族群的選擇 .....                                 | 31 |
|    | (三) 暴露確認(Exposure Ascertainment)和驗證(validation).... | 32 |
|    | (四) 結果確認(Outcome Ascertainment)和驗證 .....            | 40 |
|    | (五) 共變數確認和驗證.....                                   | 50 |
| 六、 | 數據蒐集、數據度用和轉換為最終特定研究資料集期間的資料品質 .....                 | 53 |
|    | (一) 描繪資料特徵(Characterizing data).....                | 56 |
|    | (二) QA/QC 計畫的紀錄.....                                | 61 |

|                    |    |
|--------------------|----|
| (三) 資料管理過程的紀錄..... | 62 |
| 七、 結語 .....        | 63 |
| 八、 相關文獻.....       | 64 |

## 一、 前言

為加速醫藥產品研發，以更快並更有效地將新藥提供給需要的病人，各國法規單位在近幾年來紛紛建立了針對「評估真實世界證據(real-world evidence, RWE)用來支持已核准藥品取得新適應症核准之潛在用途」的指引。本指引旨在說明法規決策中使用 RWE 的方向。真實世界證據會涵蓋使用真實世界數據(real-world data, RWD)來源之臨床研究，例如來自常規臨床實作的資訊，以得出 RWE。為簡化文字敘述，本指引中所提及之「藥品」是包括人類使用之化學藥品與生物藥品。

以下則為 RWD 和 RWE 之定義：

- RWD 是指常規性收集各種與病人健康狀況或醫療照護的相關資料。
- RWE 是指從 RWD 分析所得出的關於醫療產品之使用、潛在效益或風險的臨床證據。

RWD 的資料包括來自電子健康紀錄(electronic health records [EHR])，例如在電子醫療系統中的病人個人紀錄，包括醫療史、診斷、治療計畫、疫苗接種日期、過敏史、放射學影像、藥品紀錄和實驗室檢查結果等的資料)、醫療給付資料(medical claims data，健康

照護提供者提交給保險公司以收取治療和其他介入費用的醫療給付彙編資訊)、藥品和疾病登錄資料(registry data)、病人產生的資料(包括來自家庭使用及收集自可呈現健康狀況的其他來源之資料,如數位健康技術)等,本指引聚焦在以下兩個來源中提取且由健康照護提供者所記錄的健康相關資料:EHRs 和醫療給付資料。EHRs 和醫療給付資料均為包含病人健康資訊的電子健康照護資料(electronic health care data,一種組織化收集從電腦或其他電子技術平台取得之自動化健康資料的分析性資料),這些資料被廣泛運用於安全性研究中,且越來越常用於有效性(effectiveness)研究中。EHRs 和醫療給付資料可做為各種臨床研究設計的資料來源。

本指引是針對「使用電子健康照護資料執行藥品流行病學安全性研究指引」做補充,前述指引著重在已核准藥品上市後之安全性議題,法規單位可藉由評估來自於「藥品流行病學安全性研究」之證據,考量是否須對已上市藥品之安全性採取緊急的法規行動,此外前述指引中所稱之藥品流行病學安全性研究乃侷限於觀察性研究。

本指引與前述指引最大的區別在於本指引中的 EHRs 和醫療給付資料除了可用於上市後藥品安全性研究外,也可用於評估藥品之有效性(effectiveness),並以此支持藥品之法規決策,例如上市後藥

品之新增適應症，且本指引中的研究除觀察性研究外，尚包含介入性研究。因此本指引會擴展前述指引中與資料來源選擇相關的特定面向，並對在臨床研究中使用 EHRs 和醫療給付資料之適用性和可靠性的評估提供額外說明，此外本指引也對於在臨床研究中使用 EHRs 和醫療給付資料的相關考量提供更廣泛的概述，包括目標是提供法規單位評估藥品有效性(effectiveness)的研究在內。

## 二、 目的

本指引旨在為申請者、研究人員和其他有興趣的利益相關者，提供在臨床研究中使用 EHRs 或醫療給付資料來支持藥品的療效(effectiveness)或安全性之法規決策的考量重點。此處的臨床研究是指所有的研究設計，包含但不限於由計畫書指派治療方式的介入性研究(例如隨機分派或單臂試驗，包括那些使用 RWD 作為外部對照組的試驗)，和由常規臨床照護過程而決定治療方式的非介入性研究(例如案例-對照研究或世代研究)，指引中會以「臨床研究」或「研究」互相代換使用。

本指引會討論有關在臨床研究中潛在可使用 EHRs 和醫療給付資料來支持法規決策的下列主題：

1. 選擇可適當解決研究問題和充分描繪研究族群、暴露、欲研究

的結果(outcomes of interest)和關鍵共變數(covariate，指既不是暴露也不是感興趣結果的變數，但是被測量用來描述族群或因為可能是研究設計或分析中的干擾因子或效應修正因子而被測量的變數)的資料來源。

2. 研擬和驗證研究設計要素(如暴露、結果、共變數)的定義。
3. 數據蒐集(data accrual)、數據度用(data curation)和數據轉換( data transformation)為最終的研究特定資料集(study-specific dataset)整個過程的數據溯源(data provenance)和品質。
  - 數據蒐集(data accrual)：收集資料的過程。
  - 數據度用(data curation)：將標準化編碼應用於資料來源，如以 ICD-10 編碼作為診斷，或以各種編碼代表不良事件、疾病分級、疾病惡化和其他醫療或臨床觀念記錄在 EHR 中。
  - 數據轉換(data transformation)：包括資料提取、清理和整合。
  - 數據溯源(data provenance)：說明一段在資料庫或文件檔案中的資料起源及這段資料如何和為何呈現在目前資料庫中。

對所有研究設計而言，最重要的是確保用來幫助支持法規決策之資料的可靠性(reliability)和關聯性(relevance)。本指引中的可信度

包括資料的準確度(accuracy)、完整性(completeness)、溯源性(provenance)和可追蹤性(traceability，允許了解分析結果、分析資料集、表格資料集和來源資料間的關聯性)。適用性則包括關鍵資料要素(data elements，包括暴露，結果，共變數)的可得性(availability)，和研究中有足夠數量的代表性病人(例如，研究目標為評估後線抗生素 X 的療效，但資料來源為一家沒有設置 ICU 的地區醫院 A 的电子病歷，可推測 A 醫院使用 X 的病人數會比有設置 ICU 的醫學中心少，如此則此研究的適用性不佳)。本指引不會針對研究設計選擇或統計分析類型提供建議，也不會鼓勵使用任何特定種類的資料來源或研究方法。

### 三、 概論

所有將檢送給法規單位用來支持法規決策且使用 EHRs 或醫療給付資料的研究，均應在研究開始執行前檢送計畫書和統計分析計畫。若申請者擬尋求法規單位對該項研究的意見，則應提出諮詢以進行討論。所有研究設計、分析方法、研究執行和最終報告的基本要素均須事前訂定，且計畫書和最終報告中應描述如何從選擇的 RWD 來源中確認這些要素，包括驗證研究(validation studies)。

本指引針對資料來源的選擇提供建議，以期可使來自 EHRs 和醫療給付的資料用於臨床研究之完整性和準確度達到最大化。本指

引針對確定資料可信度和適用性的關鍵議題做說明，而這些內容亦應描述在計畫書中，包括：

1. 用於研究問題和支持關鍵研究要素的資料來源之適當性和潛在局限性。
2. 確認研究設計要素的時間段(time periods)。
3. 研究設計要素(如研究目標族群的納入/排除條件、暴露、結果、共變數等)的概念性定義(conceptual definitions)和操作型定義(operational definitions)，及驗證研究的結果。請參考第五節、研究設計要素段落中關於研究設計要素的概念性和操作型定義之例子。
  - 概念性定義(conceptual definitions)：以一般或定性術語解釋研究結構(例如暴露、結果、共變數等)或特徵。
  - 操作型定義(operational definitions)：指資料的特定操作或程序可讓研究者依循以衡量特定研究中的結構。
4. 在數據蒐集、數據度用和數據轉換為最終研究特定資料集的品質保證和品質管控(quality assurance and quality control，QA/QC)程序。

#### 四、 資料來源

檢送給法規單位的計畫書應確認研究中所有擬使用的資料來源和其他描述性資訊。應評估每個資料來源以確定其可得資訊是否適合用來解決特定的研究假說。由於已存在的電子健康照護資料並非設計用來支持法規決策，因此了解資料來源的潛在局限性非常重要。應在計畫書中描述所選資料來源用於研究目的之歷史經驗，這些描述應包括所選資料來源在擷取研究要素(包括納入/排除條件、暴露、結果、關鍵共變數)方面的表現，以及資料在特定研究活動中可如何被驗證。資料庫建置單位對於資料庫品質及驗證最為了解，若能對該資料庫綜觀說明資料庫內容及資料品質之確保，對於廠商在針對欲研究主題的資料庫選擇與應用上將有很大幫助。

##### (一)資料來源的適用性

世界各地醫療的實際操作和健康照護系統間的不同可能會影響資料來源回答研究問題的適用性，在不同種類的商業或政府健康照護給付方案中的病人特徵(例如年紀、社經地位、健康狀態和風險因子)亦不同。醫療照護系統和醫療給付方案中的各種因子，例如藥品使用排序(例如一線、二線用藥)、處方的決定、病人涵蓋性等，均可能影響在一個醫療照護系統中接受特定治療的病人與在另一個醫療照護系統中接受相同治療的病

人，在疾病嚴重度或其他疾病特徵上的差異程度。此外若欲使用這些資料來源以檢驗研究假說，則確認資料來源中是否涵蓋所有研究相關病人族群亦相當重要，若選擇的資料來源未涵蓋所有相關病人族群，應考慮進行資料連結與合成，或選擇其他資料來源。建議應提供以下資料：

1. 選擇特定資料來源來解決特定假說的原因。
2. 有關健康照護系統的背景資訊(若可取得)，包括任何用於診斷的特定方法、對欲研究疾病的優先治療方式、擬使用資料來源中收集和驗證此類資訊的程度。
3. 對醫療照護系統中處方和醫療實作的描述(若可取得)，包括核准適應症、劑型和用法用量。

若使用國外資料來源，建議提供這些因子可能如何影響研究結果外推至我國族群之合理論述。

## **(二)資料擷取(Data Capture)：整體討論**

EHR 系統的紀錄或醫療給付資料庫只會在病人和健康照護系統有交流(interaction)時才會產生，由於 EHR 和醫療給付資料的收集是在常規醫療照護期間且非依據事前訂定的研究計畫書進行，因此用來解決擬執行研究中特定問題所需的資訊，可能

不會存在 EHR 和醫療給付資料來源中。申請者應證明每個資料來源均包含所需的細節和完整性，以用來擷取研究目標族群、暴露、主要共變數、感興趣的結果和其他適用於研究問題和設計的重要因子(例如暴露的時間點、發生結果的時間點)。

### **1. 納入(Enrollment)和照護的完整擷取 (Comprehensive Capture of Care)**

涵蓋連續性(Continuity of coverage，病人被納入健康照護系統的時間段，此段期間的任何醫療照護或藥品處方均會被擷取進入健康照護系統的電子紀錄系統中)(納入和退出)是使用 EHRs 和醫療給付資料來源時應解決的問題。使用這類資料所得到的研究結果的有效性，有一部分是取決於病人遷進和遷出健康計畫和健康照護組織的紀錄。這些紀錄可用來定義納入期間(enrollment periods，可取得病人資料的時間段)和退出期間(disenrollment periods，無法取得病人資料的時間段)。應在計畫書中探討和記載納入或涵蓋連續性的定義。

應說明在擷取有關研究問題的照護和結果時資料來源的完整性，這個資訊可幫助評估所有暴露和欲研究的結果都會被擷取以用於法規決策的可能性。例如門診資料來源不適合用於可能導致住院結果的研究(舉例說明，若擬評估使用 X 藥發生過敏

的機會，僅使用門診資料來源，則無法擷取到 Steven-Johnson Syndrome 這類會造成住院的結果資料)，或欲研究的結果是取決於特定實驗室檢測頻率的研究，但臨床醫師通常不會依照這個頻率開立檢驗醫囑，因此在電子健康照護資料中擷取到的結果資料可能會有偏差(舉例說明，若擬評估 Y 藥有無造成血糖下降的結果，而能否擷取到血糖下降的結果取決於臨床實作有無開立血糖檢驗，若 A 醫師在用藥後四週檢驗，但 B 醫師在用藥後 8 週才檢驗，而研究時間區間定義為給藥後六週，則將無法取得 B 醫師病人的血糖結果，如此擷取到的結果資料會有所偏差)。

明確說明如何在研究期間擷取相關病人族群、暴露、結果和共變數的資料，特別是資料可得性隨著時間有很大的變動時。在預期與暴露有關的結果可能會發生時，資料來源應包含足夠數量的病人，且有足夠追蹤時間(基於生物學上合理的時間架構)以確認欲研究的結果確實有發生(例如，擬研究 X 藥用於早期乳癌術前新輔助治療的療效，並以無疾病存活[DFS]為評估結果，則依早期乳癌生理學考量，需追蹤到 10 年以上，才足以擷取 DFS 的結果。另一方面若是擬研究 Y 藥用於 R/R ALL 病人的療效，以整體存活為評估結果，則追蹤時間可能約一兩

年。追蹤時間足夠與否應視擬研究問題而定)。由於追蹤長度可透露所選的資料來源是否足夠，或是否需要額外支持性資料以確認可長期潛伏的結果，因此應提供資料來源中病人追蹤長度分布的資訊。

一般而言 EHRs 和醫療給付資料不會系統性的擷取非處方藥品的使用、健康照護計畫中未給付的藥品或工作場所提供的疫苗注射等(例如，病人從一般社區藥局購買的非處方藥、某些由製藥公司經醫師提供樣品給病人的未給付藥品、醫療人員公費流感疫苗注射資料等，均不會紀錄在 EHR 和醫療給付資料中)，因此若研究問題特別與這些暴露相關，則上述的資料來源可能不適合，或計畫書中應描述如何解決這些資訊缺口(gap)。

取得有特定隱私考量之病人(例如性傳播感染、物質濫用、精神健康狀況)的完整藥品涵蓋和醫療照護資料，可能具有挑戰性，若無法完整取得會導致資訊不完整或錯誤，患有這些疾病的病人可能在公立醫學中心或私人醫療診所接受自費治療，如此則不會產生保險給付資料，因此若只使用保險給付資料作為資料來源，將使資訊不完整。此外特定族群病人例如某些腫瘤病人，更常參與試驗性臨床研究，或在製藥公司的協助計畫下取得藥品，因此無法取得完整健康資料。若欲研究的問題與這

些議題相關，應在計畫書中描述如何解決這些議題。

## **2. 資料連結和合成(Data Linkage and Synthesis)**

資料連結可用來增加病人個體隨時間推移的資料廣度與深度，且為驗證目的提供額外資料。若研究涉及內部資料來源(例如母親與嬰兒再同一醫療院所中的電子病歷連結)或外部資料來源(例如生命徵象紀錄、疾病和藥品登錄、人體生物資料庫的資料等，不同機構資料來源的連結)間建立新的資料連結，計畫書應描述每個資料來源、可取得的資訊、連結方法、連結資料隨時間推移的準確性和完整性。若研究涉及產生額外資料(例如訪談、郵件調查、電腦或行動應用程式問卷、數位化健康技術取得的量測結果等)，計畫書應描述上述資料的收集方法和即將收集的資料與電子健康照護資料整合的方法。儘管根據情境的不同採用概率法(probabilistic)和確定法(deterministic)進行資料連結均有其價值，但會產生不同的連結品質，確定法的資料連結是使用完全符合(match)單一或一組常見識別碼(identifier)的紀錄來完成，且可使用單一或多步驟過程來決定資料符合狀態。概率法的資料連結是使用限制較少的步驟，其中識別碼相較下由較少的變數或部分變數組成。若使用概率法，分析計畫中應包含影響符合程度的檢定和結果的穩健性(請參考第六節、數據蒐

集、數據度用和轉換為最終特定研究資料集期間的資料品質段落)。

對於需要結合多個資料來源或研究機構的研究，鑒於不同資料來源中的族群特徵、臨床實作和編碼具有潛在的異質性，建議應說明取得不同來源的資料之可行性及如何取得，並如何在可接受的品質下做整合。

由於人們通常會在多個健康照護機構就診，特別是地理相鄰的區域，納入許多機構的去識別化(de-identification，將個人健康資訊的私人識別碼刪除的過程)資料可能造成單一個人有許多來自不同醫療照護機構紀錄的可能性，同一個人在不同機構存在多個紀錄可導致特定資料量測的過度計算，或從另一角度看，若部分機構的紀錄無法取得，則可導致收集到的個人病史只反映出病人整體健康照護歷史的一部分。對資料度用的特別關注，包括個人層級和族群層級的連結，及對於多對一(many-to-one)和一對一(one-to-one)連結的了解，是評估新資料連結適當性的基礎。而對於橫跨所有機構分享單一病人識別碼的資料來源(例如多分院醫院網絡)，上述情境則不會是一項議題，除非病人在這個網絡之外尋求醫療照護。建議考量並記錄用來解決重複或片段資料議題所進行的度用方法，並記錄無法完全透

過度用而修正的議題，需採取甚麼方法解決(請參考第六節、數據蒐集、數據度用和轉換為最終特定研究資料集期間的資料品質段落)。

### **3. 分散式資料網絡(Distributed Data Networks)**

EHRs 和醫療給付資料系統的分散式資料網絡(由多個分散的健康照護資料機構形成的網絡，有查詢或分析來自任何或所有機構資料的能力)經常透過通用數據模型(Common Data Model [CDM])，將各種電子健康照護資料來源標準化成為通用格式，以確保所有提供資料的機構間具有互相操作性)結合使用，且已越來越多被使用於醫療產品安全性監測和研究目的。使用從多個機構資料轉換為單一 CDM 之分散式網絡的主要優點是能夠對多個資料集執行相同查詢而無需做任何或重大的修改。在某些分散式資料網絡中，查詢可在所有網絡內的機構中同步進行，或在網絡中每個機構間非同步進行，而最終結果可在協調中心合併且回傳給終端用戶。分散式資料網絡使用了多種常用操作模型，有些網絡是由使用相同 EHR 系統或醫療給付資料庫結構的單一企業單位管理，雖然資料在許多地方進行維護，但均以相同的方法建構和處理(例如同一醫療體系中不同分院採用相同的電子病歷)。另一種方法是混合分散式模型(hybrid

distributed model)，來自許多遠端機構的次資料集(subset)被送至中央化儲存庫，以允許直接對整合資料集做研究(例如加入健保的醫療機構，核銷資料均送至健保署資料庫)。第三種常用方法出現在有許多擁有者和許多資料庫結構的資料系統網絡中，此資料的建構和處理方法會因位置不同而不同(如美國 FDA 的 Sentinel 系統，是結合了許多臨床醫療、藥廠、流行病學、統計學、健康資訊、資料科學等領域的美國機構之資料庫，用來監測已核准藥物安全性的國家電子系統，是專用於藥物安全性的最大型多機構分散式資料庫)，在這種模型中，研究問題被分送到各個網絡成員機構，且答案會返回到中央位置以進行收集和報告。

第三一種類型的網絡是由截然不同的資料系統組成並藉由使用 CDMs 而促進其便利性。使用 CDMs 的網絡通常也會提供分析的工具和方法、程度一致的數據度用和資料模型的週期性修訂，以依據研究者的需求納入新的資料概念。此外可將一種 CDM 的資料轉譯至另一種的方法已被研發出來，但這些因涉及額外的資料轉換，因此會存在有更多的品質考量。關於數據度用、資料轉換為一種 CDM、和整體 QA/QC 過程將在第六節中討論。

分散式資料網絡通常由 EHR、醫療給付或登錄資料所組成，然而結合多個資料來源，特別是有額外資料轉換為一種 CDM 時，增加了一層應考慮的複雜性，由於分散式健康資料網絡有許多不同配置，本指引中討論的配置不應被認為是全面完整的。

使用 CDM 將不同的資料庫架構轉換為一種通用的健康網絡，將可允許進行橫跨醫療照護機構的研究，否則這些研究將更為複雜與昂貴。然而 CDM 會對研究者帶來需考量的額外挑戰，許多 CDM 是為了滿足一項特定研究目的而創建的，CDM 中資料擷取的選擇是針對預期用途所需的資料量測類型和細節進行最佳化(例如美國 FDA 的 Sentinel 系統是用於上市後安全性監測為法規決策提供資訊)，因此 CDM 驅動網絡的資料很少包含個別健康照護機構存在的所有來源資訊，且針對特定 CDM 網絡所選擇的資料要素可能無法滿足所有研究目的或問題。此外，CDMs 通常有許多在模型內的資料要素是可選擇的，亦即雖然模型中有這些資料要素可取得來填入資料中，但各機構可選擇是否把它們的原始資料輸入這些可選擇的範圍中。

在使用 CDM 驅動網絡前，應考量 CDM 收集的資料要素—包括模型中是否有所需資料要素的存在、若有則這些要素是必

須的或可選擇的—以決定對研究的合適性，以及辨識出的缺失是否可藉由補充客制化研究的特定資料要素、收集額外資料或使用其他資料集中存在且可合理代替缺失資訊的資料要素來解決。應特別注意的是，此類變通方法會涉及申請者的額外考量，例如涉及驗證代理指標(proxy endpoints)的工作或涉及額外資料之任何人類受試者研究的考量。經常為新用途擴展而更具有彈性的 CDMs 亦可改善其合適性，關於代理變數的資訊請參考第四節(三)缺失資料：整體考量段落。

#### **4. 可運算表型(Computable Phenotypes)**

可運算表型是一種臨床狀況或特徵，可使用一組已定義的資料要素和邏輯運算式(logical expressions)，藉由對 EHR 系統或臨床資料儲存庫的電腦化查詢進行確認，其定義提供了用於辨別感興趣狀況病人族群的規範(例如，擬於多個醫療機構電子病歷中找出 chronic kidney disease (CKD)病人的資料，則 CKD 的可運算表型參數可有多種定義，如 eGFR、24 小時尿量、24 小時尿蛋白量等，至於要採用哪個參數、哪個閾值，則需依擬研究的問題而訂，如想評估某藥品對洗腎病人增加血紅素的療效，則在此研究中 CKD 的可運算表型可定義為：接受透析的病人、eGFR <15、24 小時尿量<100ml 等)。因此標準化的可運算

表型)有助於辨識類似病人族群，且可有效的橫跨多個健康照護系統替大規模臨床研究選擇族群。可運算表型的定義應包括詮釋資料(metadata，一種用來描述其他資料相關資訊的資料)和其定義、預期用途、定義之臨床合理性或研究理由，以及在各種健康照護設定下驗證資料評估的支持性資訊。由資料要素和表型運算法組成的可運算表型定義應在計畫書和最終研究報告中描述，亦應能以電腦可處理的格式取得。計畫書和最終研究報告中應描述可運算表型定義的臨床驗證。更多關於驗證的資訊請參考第四節(四)驗證：整體考量段落。

## **5. 非結構化資料(Unstructured Data)**

EHRs 內的大量主要臨床資料是非結構化資料，不管是自由文字資料段(例如醫師記載的病歷)或其他在電腦檔案中的其他非標準化資訊(例如以 PDF 為基礎的放射學報告)。為了增加資料摘錄的效率，包括現有和新興技術在內的各種方法越來越常被用來將非結構化資料轉換為電腦格式。更近期的創新法包括科技輔助摘要，當中軟體為人類資料摘要者(data abstractors，例如我國的癌症登記技術人員)提供了一種機制，使其能以一致且可擴展方式完成工作。

人工智能(artificial intelligence [AI]，製造智能機器特別是智

能電腦程式的科學和工程學)領域的科技進展可能允許更快速的處理非結構化電子健康照護資料，這些進展包括自然語言的處理、機器學習和特別是對以下領域的深度學習：(1)除了 EHRs 中的結構段落外，額外從非結構化文本中提取資料要素；(2)研發可辨識結果的電腦演算法；(3)評估影像或實驗室結果。本指引不會鼓勵使用任何特定的 AI 技術。

這些方法都是以電腦協助到各種程度，但目前仍需要顯著數量的人工輔助度用和決策，這為最終的特定研究資料集注入了額外程度的資料變異性和品質考量。若計畫書擬使用 AI 或其他衍生法，計畫書應具體說明所使用的電腦演算法之假設和參數、用於建構演算法相關資訊的資料來源、演算法是監督式(亦即使用輸入並由專家審查)或非監督式及與方法驗證相關的度量(metrics)。對資料品質的相關衝擊應記錄在計畫書和分析計畫中。

### **(三)缺失資料(Missing Data)：整體考量**

缺失資料是指要在研究分析中使用，但卻沒有被觀察、收集或評估到的資料，這包括資料來源中有想要收集但缺乏的資訊(例如，研究想分析血糖資料，EHR 中雖有血糖資料欄位，但某些病人可能因為醫師未開立、病人未執行或檢驗結果未輸入

等原因，造成其血糖資料缺乏)，和資料來源設計中沒有想要收集並因此而缺乏的資訊(例如，EHR 中不會常規紀錄病人存活狀態的資料，因此在 EHR 中對於院外死亡的病人，其存活狀態資料是缺失資料)。

資料來源可能在兩種廣泛的情況下缺少資訊，第一種情況是當資訊是預計要收集的(例如 EHR 中存在的結構化段落)但在資料來源中不存在，這是一個傳統缺失資料的例子。第二種情況是當資訊並不是預計要在 EHR 和醫療給付資料中收集的，因此資料不存在。區分這兩種情況並了解資訊在 EHRs 和醫療給付資料中存在或不存在的原因是很重要的，舉例說明，缺乏實驗室檢測結果的資訊可能由以下不同狀況造成：(1)檢測可能不是由健康照護提供者所開立的醫囑；(2)可能有開立檢測的醫囑但未被執行；(3)檢測可能有被執行，但結果未在資料來源中被儲存或擷取；(4)檢測可能有被執行且結果有儲存在資料來源中，但資料並非可取得的格式或在產生最終特定研究資料集的轉換和度用過程中遺失。由於健康照護提供者可能會基於病人特徵而開立實驗室檢測醫囑，因此開立檢測醫囑與否的決定或病人放棄檢測的決定都可能會影響資料用於擬執行研究的合適性。

資料連結是一種解決缺失資料的方法，亦可能辨識出可代理缺失資料的變數，潛在代理變數(proxy variable)的例子為：健康保險中的福保身分可作為低收入資格病人的代理變數。

計畫書和統計分析計畫應基於對資訊存在或不存在的原因有所瞭解而制定，應包括描述性分析以描繪出缺失資料的特徵，應提供關於研究指標和重要共變數在統計分析中缺失資料之假設(例如隨機缺失或非隨機缺失)的支持，並將缺失資料的影響納入考量。

#### **(四)驗證(Validation)：整體考量**

驗證是用來確立方法是可靠的，或資料是正確量測的一種過程，通常是根據參考標準。使用 EHR 和醫療給付資料來源的研究應包含重要研究變數的概念性定義，包括研究族群的納入排除條件、暴露、結果和共變數。概念性定義應反映出對感興趣變數當前的醫療和科學考量，例如：(1)定義族群選擇條件、欲研究結果或共變數的臨床準則；(2)量測藥品攝入量作為欲研究暴露的定義。

操作型定義應根據概念性定義制定，以從資料來源中提取最完整且準確的資料。在許多使用 EHR 或醫療給付資料的研究中，操作型定義會是使用結構化資料要素並以編碼為基礎的電

子演算法。在其他研究中，操作型定義可能源自於從非結構化資料中提取的相關資訊，或建構一項結合結構化和非結構化資料要素的演算法。操作型定義也可在適當的狀況下指定額外的資料收集，例如病人調查。

由於操作型定義通常無法完善地針對每個受試者依感興趣的變數準確地分類，因此產生的錯誤分類(misclassification，將個體或數值錯誤分類或分類至原本應該被分配類別以外的類別)會導致偽陽性和偽陰性(表一)，且可能造成暴露和結果之間產生特定方向和程度的偏差。雖然對感興趣變數的完整驗證(complete verification，指根據所選擇的參考標準，為每個受試者的感興趣變數分配一個準確的數值，例如審視醫療紀錄可與概念性定義結合，以決定受試者是否符合關鍵的納入條件或曾經歷過結果事件，且在不同程度上此過程可能涉及裁決)可將錯誤分類最小化，並使研究的內部效度最大化，但了解潛在錯誤分類對研究內部效度和研究推論的衝擊是決定哪些感興趣變數需要驗證及驗證程度的關鍵步驟，例如在量化藥品療效的研究中，應對內部效度最優化且應最小化關鍵變數的錯誤分類，以準確測量其關聯性。當錯誤分類的存在不預期會改變結果的解讀時，有些研究中的某些錯誤分類是可容忍的，例如用於訊號

偵測，或當假設的效應值(effect size)很大，且錯誤分類對量測關聯性的影響被認為很小的時候。

為瞭解欲研究的變數(例如暴露、結果、共變數)的潛在錯誤分類如何對關聯性的量測和結果的解讀造成影響，申請者應考量：(1)錯誤分類的程度；(2)非均衡性錯誤分類(differential misclassification，指不同組別間的資訊錯誤是不同的，亦即有暴露與沒暴露組別間或有疾病與沒有疾病組別間的偏差是不同的，例如抽菸者比非抽菸者有更高的肺氣腫診斷率，但由於抽菸者更常因支氣管炎而就診，因此較高肺氣腫診斷率可能是因為此族群更常尋求醫療照護，而不一定是這群人真的有較高的得病機會，這是因為「診斷此疾病」與「抽菸者相較於非抽菸者有多常尋求醫療照護」這個變數有關的一種非均衡性錯誤分類)，相對於均衡性錯誤分類(non-differential misclassification，指不同組別間的資訊錯誤是相同的，亦即發生在當暴露或疾病與其他變數無關時，這種均衡性錯誤分類造成的偏差通常是可預測的，例如在研究中詢問病人過去數十年間是否曾經使用特定藥物，由於時間太長可能有錯誤記憶，導致某些疾病或狀況可能與特定藥物產生錯誤關聯，但因每個病人都被詢問了相同可能引發錯誤的問題，因此在研究中每個病人都有均衡性錯誤

分類)；(3)依賴性錯誤分類相對於獨立性錯誤分類(dependent vs. independent misclassification，當兩個變數發生錯誤分類的可能性高於或低於僅考慮一個變數發生錯誤分類的可能性時為依賴性錯誤分類，反之則為非依賴性錯誤分類。例如若暴露錯誤分類的靈敏度為 90%且結果錯誤分類的靈敏度也為 90%，則根據機率計算 $(1-0.9)*(1-0.9)=1\%$ ，亦即當考慮此兩變數時，真正有暴露且有發生結果但被錯誤分類成沒有暴露且沒有發生結果的機率為 1%，若實際數值高或低於 1%則代表暴露與結果間為依賴性錯誤分類)(例如當暴露與結果都在相同的調查中自我通報時，會有暴露和結果的相關錯誤分類)；(4)暴露和結果間關聯性的可能偏差方向。

儘管完整驗證研究變數被認為是最嚴格的方式，但在某些情境中要驗證每個受試者的變數可能是不可行的，且僅評估變數之操作型定義的性能(performance)可能已足夠，例如有一個非常大的研究目標族群或對所有研究受試者缺乏參考標準(reference standard，本指引中指的是可取得的最佳基準，又稱為 gold standard)。根據表一所描述的性能量測，申請者應考量在更大範圍中驗證變數(例如根據操作型定義找出的所有陽性分類受試者)是否是必須的，且應考慮與相關審查單位討論。

由於一項操作型定義的性能取決於各種因子，例如資料來源、研究目標族群、研究時間架構和參考標準的選擇，建議用合理的取樣方法(例如隨機取樣或分層取樣)在擬執行的研究中選取一部分足夠大的研究目標族群樣本，以評估操作型定義的性能。若申請者擬使用的操作型定義已在先前的研究中評估過，則理想上應選擇在相同資料來源和類似研究目標族群中評估過的操作型定義，此外基於疾病、診斷和編碼的長期趨勢，可能必須使用更近期的資料對操作型定義作評估。用來建立靈敏度(sensitivity)、特異性(specificity)和預測值的先前研究之品質始終應進行評估。

表一、計算二元變數的靈敏度、特異性、陽性預測值(positive predictive value [PPV])和陰性預測值(negative predictive value [NPV])示意圖

| 依據擬使用操作                                                                                                                                |   | 依據參考標準的狀況 |         |     |
|----------------------------------------------------------------------------------------------------------------------------------------|---|-----------|---------|-----|
| 型定義的狀況                                                                                                                                 |   | 有         | 無       | 總和  |
| 有                                                                                                                                      | 無 | a (真陽性)   | b (偽陽性) | a+b |
|                                                                                                                                        | 有 | c (偽陰性)   | d (真陰性) | c+d |
| 總和                                                                                                                                     |   | a+c       | b+d     | N   |
| 靈敏度= $a/(a+c)$ 特異性= $d/(b+d)$                                                                                                          |   |           |         |     |
| 靈敏度(sensitivity)：受試者患有疾病時，其分類結果為陽性的概率。<br>特異性(specificity)：受試者未患有疾病時，其分類結果為陰性的概率。<br>PPV：分類結果為陽性時，受試者患病的概率。<br>NPV：分類結果為陰性時，受試者未患病的概率。 |   |           |         |     |

計畫書應包含擬計畫驗證的詳細描述，包括選擇參考標準的合理性、驗證方法(approach)、方式(method)、處理和取樣策略，若擬採用先前評估過的操作型定義應提供額外的資訊，包括使用甚麼資料來源和研究目標族群、在哪個時間架構範圍內執行評估、評估所得的性能量測數值和這個性能量測是否適用於擬執行研究的相關討論。另外也建議在計畫書中涵括預先訂

定的靈敏度分析，以顯示偏差(若存在)是否及如何影響基於驗證資料的研究結果。

關於驗證研究設計要素的討論，請參考第五節(三)5.暴露的驗證；第五節(四)3.結果的驗證；和第五節(五)3.干擾因子和效應修正因子的驗證各段落。

## **五、 研究設計要素**

以下將詳細討論關鍵研究設計要素的確認與驗證。首先應確立欲研究的問題，接著決定最適合解決這些問題的資料來源和研究設計。研究不應設計用於符合特定資料來源，因為特定資料來源的局限性可能會限制研究設計的選擇並限制可得到的推論。

### **(一)時間區間(Time Period)的定義**

建議在計畫書中明確定義與研究設計相關的各種時間區間(例如：辨識研究目標族群、定義納入排除條件、評估暴露、評估結果、評估共變數和追蹤病人的時間區間)，應採用可回答研究問題的時間單位(例如：年、月、日、小時、分鐘)在資料中可得性的充分細節，來明確描述時間標度(time scale)(例如：日曆時間、年齡、自暴露以來的時間)的重點。

計畫書應說明擬使用的時間區間和其對研究效度的潛在影響是適當的，例如應說明暴露前的時間區間是否足夠辨識研究

目標族群和重要基礎共變數、追蹤時間是否足夠觀察到研究結果的發生、受時間影響共變數更新資訊的時間區間是否適合用來獲取這些變數的變化等各項的適當性。此外在考量結果的定義時，可能需要根據研究問題將疾病發生時間(例如早期症狀)和確認診斷做區分，在定義評估結果追蹤時間的起始和結束點時，若結果預期可能會發生且與暴露相關，則要考量生物學上合理的時間架構(time frame)。

計畫書應說明標準照護隨著時間的潛在變化、其他治療的可取得性、診斷標準和其他切合研究問題和設計的各项相關因子，其他相關因子可能包括保險處方的變化、階段治療和實驗室檢查變化等。在發展研究方法前，申請者應與相關法規單位討論資料可獲取這些潛在時間變化的能力及潛在時間變化對內部效度的影響。

## **(二)研究目標族群的選擇**

計畫書應詳細描述決定如何使用納入排除條件(例如人口統計學因子、醫療狀況、疾病狀態、嚴重度、生物指標等)的方法用以從資料來源中辨識出符合這些條件的正確病人。計畫書亦應說明在預計使用的資料來源中，可收集符合納入排除條件資訊的完整性和準確性。

用來選擇研究目標族群的關鍵變數應經過驗證，例如為了評估免疫性血小板低下紫斑症病人的藥品療效，應根據疾病的概念性定義，包括症狀與表徵、血小板數量且排除其他可能造成血小板低下的原因，來驗證經由操作型定義國際疾病分類臨床修訂第九版(International Classification of Disease, nine revision, clinical modification [ICD-9-CM])診斷碼 287.31 所確認出的病例。

某些情況下，需要滿足納入排除條件的關鍵變數(例如懷孕研究中的妊娠年齡)可能是由健康照護提供者在照護時間點可取得的資訊而產生，例如健康照護提供者可在 EHR 中輸入根據孕婦自我通報的最後一次月經週期、超音波資料和其他相關資訊所計算的妊娠週數。若使用此類資料，計畫書應描述資訊來源和健康照護者用來產生資料的方法。

### **(三)暴露確認(Exposure Ascertainment)和驗證(Validation)**

由於介入性研究中需記錄暴露的分配，故本節中所討論的關於醫療給付資料或 EHRs 中的暴露確認主要適用於非介入性研究。

## 1. 暴露的定義

根據本指引的目的，暴露一詞適用於在擬執行研究中所評估的感興趣藥品或處方，感興趣藥品被稱為治療(treatment)，可相較於沒有治療、安慰劑、標準照護、其他治療或上述的任意組合。其他可能影響研究結果的變數被認為是共變數，此部分將在第五節(五)共變數確認和驗證段落中討論。暴露的定義應包含關於預計研究藥品的劑量、配方、強度、給藥途徑、給藥時間、頻率和治療期間的相關資訊，可能還需要描述特定的藥品製造商(例如當不同製造商都使用同一項疫苗的專有名稱時)。

暴露的描述應包含藥品的預期或處方用法(例如數量、頻率和特定劑量)、從暴露的起始點到合理預期可能觀察到療效的最早時間點之間的時間區間和預期的療效期間。這通常需要了解藥品的藥理特性，例如預防骨質疏鬆的單次注射藥品可能有數個月的療效。請參考第五節(三)3.暴露確認：持續期間和第五節(三)4.暴露確認：劑量各段落。

## 2. 暴露確認：資料來源

申請者應能夠呈現在預計使用的資料來源中辨識出特定感興趣藥品的能力，並顯示資料來源包含可用來辨識特定欲研究

藥品的資料文字段和編碼 (例如利用特定健保藥品編碼於健保資料庫中搜尋)，舉例而言僅從計價或診斷編碼不一定能夠推斷出特定的疫苗配方，譬如多項疫苗都使用單一計價編碼。計畫書中應描述使用的編碼系統、呈現的精細度(granularity)程度(例如使用美國國家藥品圖書館發布的 RxNorm 命名系統串聯美國國家藥品編碼[National Drug Code，NDC]系統的辨識碼，由於 RxNorm 命名系統包含藥名、成分、劑量、劑型，因此其精細度會比單純的 NDC 辨識碼更細)，和編碼系統達到的特異性。

倚賴編碼資料時，暴露的操作型定義應根基於所選資料來源的編碼系統，並反映出對資料來源中的藥品處方、調劑和銷毀特徵的了解。例如在台灣，病人使用的藥物可以是由健康照護提供者依健保規定給予，因而使用健保藥品編碼紀錄，但也可能直接從藥局自費購買，而藥局採用藥物治療分類系統 (Anatomical Therapeutic Chemical [ATC])編碼紀錄。

當結合來自結構化和非結構化資料的資訊時，說明操作型定義和結合方法亦非常重要，新興方法可能涉及審視醫療紀錄中的非結構化資訊結合藥局調劑和醫師處方資料與註解，以評估一名病人是否有被處方並接收到感興趣藥品，及病人在持續用藥上是否有問題。舉例說明，曾有一項針對退伍軍人接受常

規大腸鏡檢查的回溯性世代研究，其中使用結構化(電子健康紀錄資料庫)和非結構化(自由文字的病程紀錄)資訊來確認阿斯匹靈的暴露(Bustamente et al. 2019)。

使用醫療給付資料來源時，可能有與保險給付無關的調劑處方無法被擷取，若這些處方與研究暴露相關時則是重要考量。此外非處方藥和營養補充品通常也無法在電子健康照護資料中取得，因此探討這些不完整暴露確認的可能性和對研究效度的影響是很重要的。請參考第五節(三)5.暴露的驗證段落。

### **3. 暴露確認：持續期間(Duration)**

資料來源應擷取相關的暴露持續期間(預計使用藥品的時間)，由於有些藥品是設計為單次暴露(例如疫苗)而其他藥品可能是要用在一段延長的區間，因此資料來源的適用性會隨著所研究的特定藥品而異。建議描述暴露的持續期間和暴露對欲研究結果有影響的區間。持續期間可以指持續暴露或累積暴露，取決於欲研究的問題，有些藥品預計會立即或近乎立即的產生療效，其他藥品則預期會在一段時間後才產生療效(例如促進骨質強度的藥品)。另外建議應考量停止治療後藥品療效的持續期間，以涵蓋整個藥品療效可能發生的區間，例如，疫苗的效果可能在接種後持續數年的時間，且這段區間的人會被認為是有

暴露的，但另一方面一個抗凝血劑的療效則不會持續超過數小時或數天。建議應說明選擇用來估計暴露持續時間單位(例如小時、天)，並確保用這些單位可在資料來源中取得資料。

由於病人通常不會準確地依照時間補充調劑(refill)，或另一方面可能提早補充調劑，因此可能會存在治療間距(gap)或藥品累積(stockpiling)且可能會反映在資料中。建議應在計畫書中描述和說明研究者將如何衡量藥品使用情況、解決資料來源中的潛在治療間距和提早補充調劑時處理藥品累積的合理性。對於間歇性治療(例如用來治療疼痛的藥品是根據病人的需求而使用)或經常提供藥品樣本給病人的治療(例如昂貴藥品、新上市藥品)，在準確評估實際暴露量和暴露持續時間上會存在特殊的挑戰，請參考第五節(三)5.暴露的驗證段落。

#### **4. 暴露確認：劑量**

有關暴露的資料應包含劑量的資訊，根據研究中的暴露和欲研究的問題，描述每次給藥的劑量、每日劑量和估計累積劑量(cumulative dose，在指定時間區間內給予病人欲研究藥品[暴露量]的總量)可能是有用的。

從資料來源中有提供的劑量資訊開始著手是合理的，接著要在計畫書或最終研究報告中討論估計欲研究的暴露劑量所使

用的特定假設，特別是針對兒科病人，請參考第五節(三)6.特殊族群的劑量段落。若有多種劑型可使用，說明不同劑型(例如針劑和口服)如何計入劑量計算也很重要。

## 5. 暴露的驗證(Validation)

除了在醫院環境或輸注設置下給予的藥品之外，電子健康照護資料可擷取藥品處方和交付給病人的藥品，但一般無法擷取病人實際的藥品暴露，因為此部分取決於病人是否取得並且使用處方藥品。

理想情況下，驗證會涉及擬使用資料來源和參考資料來源間的暴露分類比較，並產生可被用於靈敏度分析的錯誤分類估計值(estimates of misclassification)。在某些情況下 RWD 可能是唯一的參考資料來源，例如暴露是以病人有無支付處方來定義，則可使用醫療給付資料且此資訊將作為參考資料來源。

驗證可從定義概念性和操作型定義開始，例如，要定義特定研究中藥品 X 的新使用，概念性定義可為「開始使用藥品 X 且過去 365 天沒有接觸到藥品 X 的暴露」，而操作型定義可為「至少有一次藥品 X 的門診處方保險給付(使用健保編碼 xxx 辨識)且交付處方日期前的 365 天沒有藥品 X 的保險給付」。對於門診設置下使用的處方藥品，藉由記錄已被調劑的處方來反映

藥品暴露的交付或給付資料，會傾向於比大部分 EHRs 更準確，在這種情況下，藉由檢視醫療給付資料來驗證 EHR 處方資料可能是必須的。對於在健康照護設置下才能給予的藥品(例如疫苗、注射藥、血品)，EHR 的給藥紀錄可能提供比醫療給付紀錄更完整的資訊，在這種情況下，藉由檢視 EHR 來驗證醫療給付資料可能是有用的。在某些情況下，當參考資料來源是無法取得的，在相同目標族群中執行的或公開發表在文獻的額外研究，可提供對暴露狀態潛在錯誤分類的估計(例如利用對研究受試者的普查來評估藥品的攝入，經由藥局/工作場所/學校取得疫苗接種人數的公開報告)。

建議應記錄用來計算和驗證暴露持續時間、劑量、藥品轉換和其他暴露特徵的方法，且驗證和錯誤分類議題應在適當的研究文件中說明。

## **6. 特殊族群的劑量**

除了報告關於處方劑量、交付藥品或給藥的驗證資訊外，可能需要額外的資訊以評估劑量是否適合用於特殊族群(例如是否有顯著劑量不足或過量)，舉例說明，要評估給予病人一項主要經腎排除之藥品的使用劑量時，可能需要取得血清肌酸酐、肌酸酐清除率或估計腎絲球過濾率的數值，以評估適當的給予

劑量。另一個例子是在估計兒科族群的暴露時，可能需要取得病人的體重並在其體重類別中描述劑量。給付資料中可能更頻繁的需要額外資料以允許對劑量做正確的評估，但使用 EHRs 時若缺乏必要資料，也可能會發生此狀況。

## 7. 其他考量

選擇適當對照組是臨床研究中很重要的部分，計畫書中應清楚定義提供對照組資料的病人並提供足夠的細節，計畫書應討論選擇以下各項的理由：(1)對照組資料的來源；(2)時間區間(若對照組和治療組非相同時段)。由於對照藥品與欲研究藥品在一項疾病中的特定適應症、禁忌症、安全資訊、使用者疾病嚴重程度或共病症和病人特徵可能不相同，因此確保有足夠資料可讓法規單位評估暴露與對照族群的可比性是很重要的。

若整體評估認為併用藥品存在與否，可能對研究結果有所影響(例如擬研究 A 藥預防心血管疾病的效果，而已知併用 B 藥會加強 A 藥預防心血管疾病的效益)，則應描述相關併用藥品(concomitant medication，與欲研究的藥品或對照藥品同時使用的處方藥、非處方藥或補充品)的使用並從資料來源中做確認。應詳細描述一項研究中對併用藥品的定義，併用藥品的定義可能包括藥品在同一天交付、藥品供應時間有重疊或在研究區間

內有處方調劑兩項以上的藥品。另外也應描述併用藥物確認的局限性(例如非處方藥)。

#### **(四)結果確認(Outcome Ascertainment)和驗證**

選擇資料來源的一項關鍵步驟在於資料來源是否能擷取欲研究的臨床結果，由於健康照護資料通常可擷取會引起健康照護專業人員注意的結果並被記錄在醫療紀錄中，因此症狀輕微的結果或發生在醫療照護機構以外的事件(例如院外死亡)通常無法很好的被擷取到。相反的，已存在問題的惡化(例如憂鬱症、乾癬、關節炎)因為不會導致明顯的事件，因此不連續的結果或急性事件(例如中風、心肌梗塞、新發生的感染)通常比前者更容易被擷取到。不同於傳統臨床試驗，完全使用電子健康照護資料來確認結果的試驗，可能不會有計畫書定義的追蹤訪視，且可能沒有在結果確認所需的時間間隔內監測事件。此外在傳統臨床試驗中對感興趣結果的評估可能更標準化且全面，因此應仔細考量感興趣結果在資料來源中的可得性、準確性和完整性，及是否需要外部資料連結。應在研究開始之前評估資料來源是否及在何種程度上可擷取到感興趣的結果，且須與感興趣的暴露獨立評估。

## 1. 欲研究結果(Outcomes Of Interest)的定義

許多結果涉及醫師作為標準照護一部分所記錄下的診斷，為了使臨床實作在不同醫師間的變化和隨著時間的變化所造成的影響降到最低(例如使用不同的診斷和分類標準、用不同的方法對相同事件作編碼)，建議應適當的依據該狀況的臨床、生物學、心理學、功能學等概念來定義欲研究的結果。欲研究的結果的概念性定義(又稱為案例定義[case definition])應反映出對該狀況在醫療和科學上的了解，且可能因研究不同而異，例如過敏反應(anaphylaxis)的概念性定義可能包括以下臨床標準：突然發生、表徵和症狀迅速惡化、一項以上的重大皮膚學標準和一項以上的重大心血管或呼吸系統標準。計畫書應詳細描述概念性定義，包括可用來確認結果的表徵、症狀及實驗室和影像學檢查結果。

概念性定義應該要能在 RWD 來源中被操作型化，例如腫瘤學的隨機分配對照試驗通常使用以腫瘤為基礎的感興趣結果來設定追蹤評估的特定時間和頻率(例如評估無惡化存活率 [progression free survival, PFS] 通常會訂每三個月評估一次疾病狀態)，且通常會包括一些在臨床實作中不一定會定期監測的分子或生物指標檢測，由於客觀反應(腫瘤縮小)或依標準臨床試

驗準則(例如 RECIST 1.1)定義的腫瘤惡化日期通常無法在 RWD 來源中擷取，因此可能需要使用替代方法(例如以開始腫瘤治療新處方的日期為疾病惡化日期)或多面向定義(例如綜合考量開始腫瘤治療新處方的日期、開始腫瘤治療新處方前一次接受影像檢查日期、病歷明確記載腫瘤惡化日期等多個面向)去探索並說明其合理性。一般而言，若結果是有明確定義的診斷標準(例如中風、心肌梗塞或肺栓塞)，相較於較主觀或本質上是度量化的結果(例如類風溼性關節炎的關節疼痛惡化或憂鬱症狀惡化)，前者通常較容易且可一致性的在 RWD 中被擷取到。申請者應與法規單位討論擬使用的結果定義。

## **2. 結果確認**

為了幫助在選擇的資料來源和研究目標族群中辨識出潛在病例，應基於欲研究結果的概念性定義來制定使用診斷和處置編碼(例如 ICD-9-CM、ICD-10)、實驗室檢查和數值或非結構化資料(例如醫師的會診紀錄、影像學和病理學報告)的操作型定義，若操作型定義包括摘要自 EHR 或其他資料來源的非結構化資料(例如使用在出生證明紀錄中提及的脊柱裂來辨識出嬰兒神經管缺損病例)，計畫書應針對用來處理非結構化資料的方法及工具和這些方法的驗證提供詳細描述並說明理由。更多關於非

結構化資料的資訊請參考第四節(二)5.非結構化資料段落。若擬使用病人或醫師生成的資料(例如主觀指標的資料)以評估欲研究的結果或作為操作型定義的補充，計畫書應敘明如何定義、建構和驗證結果的量測方法(例如表徵評分、嚴重度指數)，並說明如何收集資料。

當結果有錯誤分類時，操作型定義的靈敏度和特異性會是不完美的，鑒於靈敏度和特異性通常不可能完美(亦即 100%)，結果的錯誤分類可能會造成偽陽性和偽陰性，建議在為擬執行研究制訂或選擇操作型定義時，需考量結果的錯誤分類會對研究效度造成的潛在影響，例如當針對在世代研究中不常發生的結果進行研究時，鑒於結果事件的低盛行率，制定操作型定義使其能夠達到高特異性以使偽陽性病例降到最低，並達到高靈敏度以擷取到更多真實案例是很重要的。

由於資料庫有特定的靈敏度與特異性和不同的疾病盛行率，一項資料來源或研究目標族群的操作型定義之制定可能無法用於其他資料來源或族群，陽性預測值(PPV)和陰性預測值(NPV)與靈敏度和特異性相關，而且是一項測量預測值之族群中結果盛行率的直接函數，因此 PPV 和 NPV 會因資料來源和研究目標族群特性(例如人口統計學因子、原有疾病、共病症、

臨床設置情境等)不同而異。

計畫書應詳細描述操作型定義、編碼系統、選擇用來建構操作型定義資訊的合理性與相關限制(例如主要或次要診斷碼的選擇，其順序可能不一定與醫療重要性一致)和對結果的錯誤分類造成的潛在影響。若操作型定義的性能已在過去的研究中評估過，應討論其對於擬進行研究的適用性。此外由於在過去研究中用來建立靈敏度、特異性和預測值的案例定義可能包含與擬進行研究的概念性定義有不同診斷標準，適當使用在過去研究中評估過的性能指標應仔細考量。

### **3. 結果的驗證**

法規單位希望可驗證結果變數以使結果的錯誤分類降到最低。雖然完整驗證結果變數是最嚴格的方法，但在某些情境下驗證每個受試者的結果可能是不可行的，且評估結果的操作型定義之性能可能就足夠。結果的驗證涉及使用臨床上適當的概念性結果定義以決定用操作型定義分類的病人狀態是否真的代表感興趣的結果，通常可藉由審視病人電子或紙本醫療紀錄中的臨床細節來達成。

建議使用標準化的醫療紀錄審查程序，包括使用標準化的工具、記錄程序和訓練人員。一項標準且可複製的程序對於將

評估者個人(intra-)和評估者們之間(inter-)的變化性最小化是很關鍵的，特別是對於醫療紀錄通常無法在不同系統間分享且無法進行中心化醫療紀錄審視的多中心研究，即使有中心化的醫療紀錄審視，標準化程序可幫助確保不同的裁決者或同一裁決者隨時間變化均使用相同標準。報告比較指標(例如 kappa 統計)有助於確保可複製性。計畫書中應合理解釋估計的醫療資料取出率(medical record retrieval rate，從資料來源中估計能成功搜尋出特定醫療資料的概率)，並討論對內部和外部效度的影響。此外對病人暴露狀態的了解可能會影響觀察者並導致非均衡性錯誤分類，因此應考量藉由遮蔽(masking)研究問題或編纂(redacting)暴露資訊，以使摘錄者和裁決者對暴露狀態呈現盲性。計畫書應提供如何處理觀察者偏差的描述。

理想上經由對結果變數的完整驗證，每名受試者可分配到準確的結果變數數值，使得結果的錯誤分類最小化並改善研究的內部效度。但在實作上更常使用的方法是評估驗證性研究中操作型定義的性能，性能指標例如靈敏度、特異性和預測值無法準確的分類案例和非案例，相反的這些指標能告知結果錯誤分類的程度，並能幫助解讀有錯誤分類存在狀況下的結果。

PPV 通常是在驗證性研究中評估，PPV 是操作型定義辨識

出的潛在案例中真正陽性的比例，因此 PPV 可告知在辨識出的案例中包含偽陽性案例的程度。當對偽陰性案例的關注可忽略時(例如當靈敏度被認為足夠高到偽陰性案例數可達最小化)，高 PPV 可能足以對結果變數的效度提供信心，而若是中到低的 PPV 可能必須對所有潛在案例的結果變數進行完整驗證。當對偽陽性案例和偽陰性案例的程度有所疑慮時，申請者應考慮評估定量偏差分析(quantitative bias analysis)所需之所有性能指標，以評估結果錯誤分類對關連性量測的影響，或採取更嚴格的方式藉由驗證所有潛在案例和非案例的結果變數，以準確的分類每個受試者的結果變數。整體而言，所需的驗證程度應取決於確定性的必要程度和潛在錯誤分類的衝擊會對研究造成的影響。

一般而言，在選擇操作型定義時申請者應考量偽陽性和偽陰性案例間的權衡，並訂出適當的結果驗證方法以支持內部效度，例如為了辨識出嬰兒神經管缺損案例，涵括一系列住院和門診診斷代碼的操作型定義可能有高靈敏度、低特異性和低 PPV，限制操作型定義在只納入住院診斷代碼或結合診斷與處置(例如手術修補)代碼可能增加 PPV，但漏失相當比例的真實案例(低靈敏度)，由於漏失真實案例對不常被報告的結果是特

別重要的疑慮，因此可採用一種方法是選擇高敏感度的操作型定義並對所有潛在案例進行完整驗證，以使所有真實案例都被辨識出來的可能性最大化，並藉由驗證使偽陽性案例最小化。與罕病的結果不同，當欲研究的結果涉及一項更常見的事件(例如疾病相關住院)或一個狀況的改善或惡化(例如關節疼痛程度)，由於真實案例和真實非案例都很盛行，因此常見診斷的操作型定義很可能會產生偽陽性和偽陰性案例達到有疑慮的程度，如此可能難以獲得準確和完整的資訊(例如實驗室檢查結果、功能性指標)讓操作型定義可準確分類案例與非案例，對於這樣的結果，單獨測量 PPV 並不足以告知結果的錯誤分類。

在無法對每名受試者進行完整驗證的情境下，應使用合適的取樣策略對擬進行研究的目標族群評估操作型結果定義的性能。如前所述，使用曾在過去研究中評估過的操作型定義，理想上應該用在相同的資料來源和類似的研究目標族群中，因為操作型定義的性能可能隨著資料來源和研究情境不同而有顯著差異，且若疾病、診斷和編碼有長期變化趨勢，則可能需要較近期的資料。應對用來建立靈敏度、特異性和預測值的過去研究之品質作評估，特別是用來建立這些指標的過去試驗中所使用的案例定義，應要能與擬執行研究中所定義的概念性結果定

義一致。應說明將這些指標應用於擬執行試驗中的合理性，且應考量執行靈敏度分析。

若沒有完整的病人資訊和對結果變數的完整驗證，結果錯誤分類對研究的內部效度仍是一項威脅，且對於暴露與結果間關聯性量測的影響取決於暴露組間的錯誤分類程度是否不同。非均衡性錯誤分類涉及暴露組間靈敏度、特異性和疾病盛行率差異的複雜相互作用，且可能使關聯偏差朝向或遠離虛無(null)。由於難以預測偏差的方向，非均衡性錯誤分類對安全性或療效性研究都是重要的疑慮。均衡性錯誤分類則與非均衡性錯誤分類不同，前者傾向於將關聯性的偏差導向虛無，因此在安全性研究中可能會遺漏真正的風險，而在療效性研究中則可能需要更大的試驗目標族群以證明藥品療效。

當結果定義沒有足夠的優化且含有與感興趣暴露間關聯性不均勻的狀況時，可能會發生均衡性結果錯誤分類，例如神經管缺損包括初級神經管形成缺損和神經管形成後缺損，初級神經管形成缺損是直接肇因於初級神經管形成(即神經管關閉)的失敗，大約發生在受精後 18-28 天間，而神經管形成後缺損的病生理則較不清楚，因此若在妊娠期神經管形成的關鍵期有接受到藥品暴露，可能不會以同樣的方式對神經管形成後造成影

響，當結果定義同時包含初級神經管形成和神經管形成後的期間，若存在有初級神經管缺損的風險則可能不會被偵測到。

在暴露狀態是盲性的試驗中均衡性結果錯誤分類可能會最小化，然而即使當資料收集方法似乎已排除均衡性結果錯誤分類時，並無法保證排除在特定研究中真實資料的非均衡性結果錯誤分類，例如，做觀察、診斷和記錄的醫師，與決定哪位病人要接受用來避免特定結果的治療(不論結果是否發生)的醫師可能是同一位，或者醫師可能根據他對病人所接受治療的知識，以不同方法監測疾病進展或治療副作用。有偏差的錯誤分類也可能來自於有安全性疑慮之特定藥品的公開公告，若資料包含發生在公開公告日期後的事件，則使用 RWD 時可能仍無法預測結果錯誤分類偏差的方向。此外當一項研究中存在有超過一種錯誤分類時，申請者應考量他們之間可能有如何的相互關聯，例如儘管均衡性暴露錯誤分類和均衡性結果錯誤分類各自的偏差可能都會使關聯性朝向虛無，但當兩個錯誤分類是相關的，則整體產生的偏差可能會遠離虛無。因此在評估潛在錯誤分類對研究推論所造成的影響時，申請者應避免過度依賴偏差朝向虛無假設的均衡性錯誤分類，在這種情況下，根據擬研究目標族群的暴露狀態來評估操作型結果定義的性能可能是必

須的。

關於結果的驗證，申請者應說明擬採用的驗證方法之合理性，例如驗證所有潛在案例或非案例的結果變數，對比於評估擬使用操作型定義的性能，若採用後者則要說明將評估哪些性能指標及其合理性。計畫書應詳細描述結果驗證的設計、方法、過程和取樣策略。若擬使用過去曾評估過的操作型定義，應提供以下與過去驗證有關之額外資訊：(1)資料來源和研究族群；(2)驗證是在哪個時間區間內執行；(3)性能特徵；(4)評估性能的參考標準；和(5)過去的驗證資料是否適用於擬進行研究的討論。

建議在計畫書中涵括定量偏差分析作為靈敏度分析，以證明結果錯誤分類是否以及如何對研究結果造成影響，計畫書應預先定義將用於定量偏差分析的指標(例如靈敏度、特異性、PPV、NPV)，並描述如何在結果驗證中測量選擇的指標。

#### **(五)共變數確認和驗證**

根據本指引的目的，在特定研究中的共變數可包含兩種類型的要素：干擾因子(confounders 又稱 confounding factor)和效應修正因子(effect modifier)。

## 1. 干擾因子(Confounders)

干擾因子是指在分析中經適當調整後，可用以減少干擾偏差的變數。而干擾則是指由於暴露和其他會影響結果發生的因子具有關聯性，而造成暴露對結果的效應量測失真。

在一項非隨機研究中收集潛在干擾因子的資訊可用來支持在分析中平衡治療組與對照組的成果。

在辨識出研究中的干擾因子後，應對擬使用的資料來源進行評估，以確定其是否足以擷取可能造成干擾因子的重要因子之資訊，這些包括在擬使用資料來源中，可以很好擷取到的干擾因子(可量測的干擾因子)及無法很好擷取到的干擾因子(無法量測或不完全量測的干擾因子)，在電子健康照護資料，特別是保險資料中無法量測或不完全量測的干擾因子的例子包括種族/民族、疾病家族史、生活型態因子(例如抽菸、飲酒、營養攝取狀態、身體活動狀態)、特定身體量測指標(例如身體質量指數 [body mass index, BMI])、沒有保險給付的藥品和使用藥品的適應症。建議考量和其他資料來源或額外資料收集間的潛在關聯，以擴大在原始資料來源中重要的無法測量或不完全測量干擾因子的擷取。

## 2. 效應修正因子(Effect Modifiers)

效應修正因子是一個可在生理上、臨床上、社會上或其他方面改變另一項研究因子效果的因子。藥品有效性或安全性研究通常是報告治療效應的平均值，即使相同治療在不同病人族群間可能會有不同效果，潛在效應修正因子的資訊可用來更好的瞭解治療效應的異質性，亦即一個族群中個體治療效應方向和強度的非隨機、可解釋的變異性。研究中應檢查因人口統計學變異數(例如年齡、性別、種族、民族)或相關共病症產生效應修正的可能性，若辨識出重要且可能對研究結果造成影響的效應修正因子，則在所選擇的資料來源中應可取得相關的效應修正因子，若單一資料來源無法完整取得所需資訊，可考慮使用資料連結等方法克服。

## 3. 干擾因子與效應修正因子的驗證

對於所有的關鍵共變數，建議要在計畫書和最終研究報告中提供操作型定義的效度並說明合理性。若可測量的共變數在病人追蹤期會改變(隨時間變化的共變數)且對分析很重要，則計畫書應描述是否可擷取到隨時間變化共變數的資訊以及多頻繁，特別是因為在 RWD 中擷取隨時間變化的共變數可由疾病嚴重度做區分(例如較嚴重疾病的病人有較多檢測)。

在評估共變數操作型定義的效度時，建議根據共變數的性質找出最佳的參考資料來源，在驗證以醫療事件或採用的處置作為共變數的操作型定義時，適用與「第五節(四)3.結果的驗證」段落中相同的原則。針對驗證有關於使用藥品之共變數的操作型定義(例如併用藥品或過去使用的藥品)的討論，請參考第五節(三)5.暴露的驗證。當評估其他共變數操作型定義(例如疾病家族史、生活型態因子或使用藥品的適應症)的效度時，適當的參考資料可能包括病人或醫療照護提供者的調查或適當的資料連結。

當需要補充資訊以擷取重要共變數或用來進行共變數驗證時，建議應描述獲得整體研究目標族群補充資訊的可能性，若此補充資訊僅可在一部份研究目標族群中取得，建議在相關研究文件中討論對內部效度的潛在影響。

## **六、 數據蒐集、數據度用和轉換為最終特定研究資料集期間的資料品質**

本節討論在資料生命週期(data life cycle)過程中檢查資料品質時需考量的重點。雖然資料生命週期可能隨著資料種類和設置環境(亦即健康照護設置，例如藥局、診所、急診單位和醫院)而異，一般而言生命週期涉及許多階段：從原始來源資料(source data)中進行數據

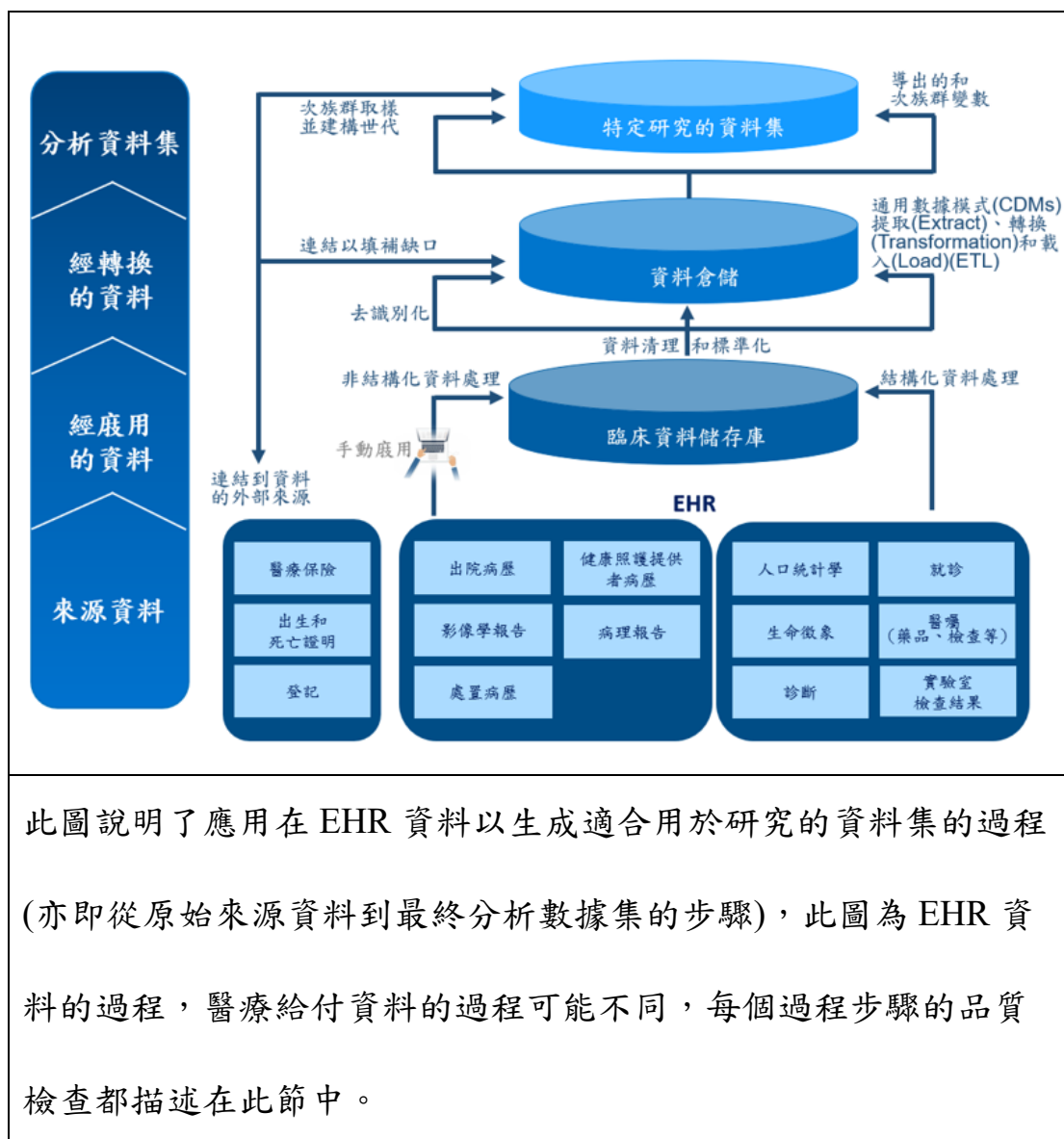
收集；將數據度用進入臨床資料儲存庫(data repository)；將數據轉換並創建資料倉儲(data warehouse)，必要時轉換過程需去識別化(de-identification)；和生成用於分析的特定研究資料集(請見圖一)。

- 來源資料(source data): 在原始紀錄和認證過的原始紀錄副本上的所有資訊，這些資訊與臨床發現、觀察或其他臨床研究活動相關，且是重建和評估研究所需要的。
- 資料儲存庫(data repository): 一個統合來自不同臨床來源資料的資料庫，以提供一名病人所接受的醫療照護之更廣泛面貌。
- 資料倉儲(data warehouse): 由資料儲存庫中經過資料轉換和去識別化的資料所組成。

資料生命週期的概念闡明了檢驗資料品質過程的迭代(iterative，重複反饋的過程)本質，這個過程並非一次性評估，相反的是個持續的過程，會在週期中的每個階段進行資料品質檢查、資料清理(cleansing)和監測，且有些檢查會重複進行(亦即在週期中的多個階段皆發生)。

- 資料清理(data cleansing)：又稱資料洗滌(data scrubbing)，是從資料庫中糾正或刪除不準確的資料、未正確格式化的資料或重複的資料的一個過程。是維持資料品質的基本任務，而需要糾正或移除的資料則稱為錯資料(dirty data)。

圖一、闡明 EHR 資料生命週期的範例



此圖說明了應用在 EHR 資料以生成適合用於研究的資料集的過程 (亦即從原始來源資料到最終分析數據集的步驟)，此圖為 EHR 資料的過程，醫療給付資料的過程可能不同，每個過程步驟的品質檢查都描述在此節中。

評估 EHRs 和醫療給付資料品質的指引主要聚焦在分散式資料網絡，在當中完全不同的資料來源被聚集、連結並處理，以創建一個完整的資料倉儲，雖然法規單位不會為特定的指引或檢查清單之設置背書，研究者應評估資料的完整性、準確性和可信性 (plausibility，資料數值的可信度或真實性)，並根據其原始來源(例如出院病例、病理報告、登錄報告)驗證資料，且使其符合奠基於共識

的資料標準。研究者亦應針對選擇這些標準提供科學性的合理論述，並清楚說明這些標準如何足夠確保相關資料來源的完整性、準確性和可信性。

研究計畫書和分析計畫應明確說明資料溯源(亦即用在整個資料生命週期中的度用和轉換程序)，並描述這些程序會如何影響資料完整性(data integrity，資料的完整性[completeness]、一致性[consistency]和準確性)和研究的整體效度。以下是在資料生命週期中每個步驟檢查資料時應考量的重點，包括：(1)根據資料數值的完整性、符合性(conformance，資料在標準化類型、大小和格式均一致)和可信性描繪資料特徵，(2)記錄包括轉換過程的 QA/QC 計畫，和(3)定義確保資料完整性的一組程序。

### **(一)描繪資料特徵(Characterizing data)**

EHR 和醫療給付資料的格式與溯源可在不同健康照護設置(例如保險公司、醫療實作、照護提供者、資料供應商)間有顯著變化，一般而言研究者應說明用來確保資料完整性和準確性的程序，及在資料生命週期中資料蒐集、度用和轉換的過程。當所選資料來源適用時，建議以標準化方式將包括下列特徵與過程的資料品質報告自動化：

- **數據蒐集(Data accrual)**

1. 資料取出的方法和過程，可使得在照護時間點擷取資料時(例如臨床實作期間手動或自動收集健康照護資料的過程)，讓資料中的缺失資料之提取、不可信數值和資料品質檢查最小化，以確保核心資料要素的準確性和完整性。
2. 核心資料要素的溯源，以允許追蹤回到這些要素各自的起始點，且對可能發生的修改有清楚的紀錄。
3. 資料可得性的即時性、資料橫跨年度和涵蓋連續性(例如納入病人期間中位數)。
4. 資料差異和重複記錄的處理。RWD 可能來自多個資料流或橫跨各種設置和平台，這些可能會造成相同變數卻存在資料差異(例如當相同要素的資訊在不同資料來源以不同方式輸入時)，或甚至在相同資料來源中重複記錄相同病人。
5. 資料持有者在資料收集的相關期間實施資料錯誤更正的原因和時間點。
6. 資料持有者在收集資料的相關期間實施可能影響資料蒐

集和資料品質檢查之程序變更的原因和時間點。

7. 研究期間感興趣變數相關的編碼實作與版本之更新或變化(例如 ICD 診斷碼由 ICD-9-CM 改變為 ICD-10)。
8. 研究期間資料的任何其他變化(例如收集、報告、定義)及其對研究結果的潛在衝擊。

- **數據度用(Data curation)**

1. 從各種來源隨時間定期遷移資料。
2. 各機構間使用的 QA 檢測和品質檢查，及用以決定資料品質技術是否適用於資料預期目標的標準。
3. 根據一致且已知之臨床意義和對資料溯源之了解而作良好定義的核心資料要素，及採用之臨床定義的紀錄。
4. 資料要素完整性和隨時間變化之趨勢的評估。
5. 非結構化和結構化資料處理(例如將非結構化資料摘要和轉換為結構化資料)，包括手動與相對的自動化技術。
6. 跨系統結構化資料的調和(harmonization)。
7. 與開放的、奠基於共識的資料度用標準(若適用)之一致性。

8. 串聯(mapping)的準確性(例如存在有不同編碼系統，將健保藥品編碼與 ATC 編碼串聯)。
9. 其他調和與串聯的考量，若適用(若資料橫跨多個國家，例如同時使用台灣和日本的資料)。

- **數據轉換(Data transformation)**

1. 應用於整個儲存庫族群作為資料倉儲創建一部分之提取、轉換和載入過程的執行。
2. 將病人紀錄去識別化，並保留在原始來源資料中可對特定病人再識別化且不失去可追蹤性之能力。
3. 用於轉換和清洗資料的演算法，及標準操作程序(包括驗證資料的程序)之可用性。
4. 為了達到資料要素和語意學(semantics)內部一致性的資料標準化(例如資料種類、大小、格式)，包括當地編碼的語意學到目標術語(例如實驗室數據)。
5. 將多個資料來源轉變為 CDM 時，資料轉換為 CDM 的處理(例如常用術語和結構)、CDM 的完整性(例如 CDM 是否含有關鍵資料要素)、辨識和處理在資料來源內或之間

重複記錄的方法(例如演算法和方法學)，及限制為 CDM 會對樣本數和病人追蹤時間或藥品暴露時間的潛在影響。請參考第四節(二)3.分散式資料網絡段落。

6. 與資料模型符合性誤差(conformance errors)相關之資料檢查的執行
7. 用以準備資料連結的資料轉換過程。請參考第四節(二)2. 資料連結和合成段落。
8. 紀錄資料連結(亦即連結來自多個資料集的紀錄)和刪除重複紀錄(亦即找出資料集之中的重複紀錄)過程的品質，可能會因用於執行配對的資料準確性和連結演算法準確性而異。
9. 量化可能導致研究結果偏差的誤差(例如錯誤配對、遺漏配對)，這在評估連結品質時很重要。報告連結演算法的細節和適當的指標(例如連結誤差率、配對率、連結和非連結資料特徵的比較)是重要的，其他考量因素尚包括誤差是隨機或非隨機、潛在偏差及對風險估計和研究結果的影響。
10. 裁決連結資料中不一致的程序，和處理連結不一致的計

畫(例如裁決連結誤差的風險估計)。

- **特定研究分析資料集(Study-specific analytic dataset)**

1. 編譯分析資料集時，依照研究計畫書和統計分析計畫所述之資料規範的遵循度。
2. 其他特定研究的資料轉換，例如僅對感興趣病人次族群轉換而不實施於資料倉儲中所有的病人紀錄(例如從非結構化文字病理報告中手動提取資料)。
3. 實施於最終分析資料集的資料檢查，針對資料要素之不可信數值(例如身高、體重、血壓)、如何處理這些數值及關鍵分析變數資料的完整性。
4. 缺失和不可信資料的程度、百分比和模式。根據分析計畫中提出用於處理缺失資料的方法，可能執行插補(imputation)並將其納入最終分析資料集和描述的插補方式中。

## **(二)QA/QC 計畫的紀錄**

應在研究計畫書和分析計畫中描述用來建構分析資料的QA/QC 計畫、預計在分析期間處理品質控制議題的方法及藉由資料要素對區別資料品質程度的仔細考量(及對研究結果的潛在

影響)。一般而言，在研究設計階段研發關於資料行動之前的資料品質確保行為會被視為是 QA，包括如何收集資料的標準化程序。品質控制包括從資料收集到編譯最終分析資料集過程中所做的決定和採取的步驟，用以確保過程符合預先訂定的標準，並確保採用的過程具有再現性。採用多專業方法(包括臨床輸入在內)以確保足夠的資料擷取和處理是必須的，特別是對於本質上結合了醫療照護之細緻度與複雜性的電子健康照護系統。

### **(三)資料管理過程的紀錄**

應從資料收集到撰寫最終研究報告整個期間，完整評估所有手動和自動的資料提取和轉換過程以確保資料的完整性，研究者應確保資料度用和轉換過程不會改變資料的涵義或造成重要的上下文資訊之流失。流程的描述應包括保護措施或檢查，以確保病人資料沒有重複或過度呈現，此外用於挖掘和評估非結構化資料過程的紀錄，應描述用以摘要非結構化資料(例如醫師記載的病歷)和附加結構化資料(例如診斷編碼)的技術(例如自然語言處理)。

應在研究計畫書或分析計畫中，描述用於處理和準備最終特定研究分析資料集的過程。分析師應對資料和編纂分析資料

集使用的軟體有經驗或經過適當訓練。為便於審查，所有提交的程式(由分析師所編寫)都應以註解做完整的註釋，註解應描述程式中所寫的每個資料處理和分析步驟之意圖和目的。

## 七、 結語

本指引的目標是提供在臨床研究中，使用 EHRs 或醫療給付資料來支持藥品的療效或安全性之法規決策的考量重點，此外本指引中所指的臨床研究是指所有的研究設計，包含介入性研究和非介入性研究。由於本指引除安全性外還包含療效之法規決策，因此最重要的是在選擇研究設計、統計分析方法和資料來源時，必須確保用來幫助支持法規決策之資料的可靠性(reliability)和關聯性(relevance)。

所有將檢送給法規單位用來支持法規決策且使用 RWD 的研究，均應在研究開始執行前檢送計畫書和統計分析計畫。所有研究設計、分析方法、研究執行和最終報告的基本要素均須事前訂定，且應描述如何從選擇的 RWD 來源中確認這些要素。由於此類研究使用的是真實世界資料，因此建議於研究規劃階段，盡早與法規單位諮詢討論。

## 八、 相關文獻

1. Bustamante, R, A Earles, JD Murphy, AK Bryant, OV Patterson, AJ Gawron, T Kaltenbach, MA Whooley, DA Fisher, SD Saini, S Gupta, and L Liu, 2019, Ascertainment of Aspirin Exposure Using Structured and Unstructured Large-scale Electronic Health Record Data, *Med Care*, 57:e60–e64.
2. Carreras, G, M Simonetti, C Cricelli, and F Lapi, 2018, Deterministic and Probablistic Record Linkage: an Application to Primary Care Data, *J Med Sys*, 42(5):82.
3. Curtis, M, SD Griffith, M Tucker, MD Taylor, WB Capra, G Carrigan, B Holzman, AZ Torres, P You, B Arnieri, and AP Abernethy, 2018, Development and Validation of a High-Quality Composite Real-World Mortality Endpoint, *Health Services Research*, 53(6)Part I:4460-4476.
4. Daniel, G, C Silcox, J Bryan, M McClellan, M Romine, and K Frank, 2018, Characterizing RWD Quality and Relevancy for Regulatory Purposes, Duke Margolis Center for Health Policy, accessed January 9, 2019, [https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing\\_rwd.pdf](https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf).
5. Girman, CJ, ME Ritchey, W Zhou, and NA Dreyer, 2019, Considerations in Characterizing Real-World Data Relevance and Quality for Regulatory Purposes: A Commentary, *Pharmacoepidemiol Drug Saf*, 28(4):439–442.
6. Harron, KL, JC Doidge, HE Knight, RE Gilbert, H Goldstein, DA

- Cromwell, and JH van der 5 Meulen, 2017, A Guide to Evaluating Linkage Quality for the Analysis of Linked Data, *Int J Epidemiol*, 46(5):1699–1710.
7. Haynes, K, 2019, Mortality: The Final Outcome, *Pharmacoepidemiol Drug Saf*, epub ahead of print Jan 31, 2019, doi: 10.1002/pds.4715.
  8. Kahn, MG, TJ Callahan, J Barnard, AE Bauck, J Brown, BN Davidson, H Estiri, C Goerg, E Holve, SG Johnson, ST Liaw, M Hamilton-Lopez, D Meeker, TC Ong, P Ryan, N Shang, NG Weiskopf, C Weng, MN Zozus, and L Schilling, 2016, A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, *EGEMS*, 4(1):1244.
  9. Lash, TL, MP Fox, and AK Fink, 2009, *Applying Quantitative Bias Analysis to Epidemiologic Data*, New York (NY): Springer.
  10. Mahendraratnam, N, C Silcox, K Mercon, A Kroetsch, M Romine, N Harrison, A Aten, R Sherman, G Daniel and M McClellan, 2019, Determining Real-World Data’s Fitness for Use and the Role of Reliability, Duke Margolis Center for Health Policy, accessed July 24, 2020 [https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd\\_reliability.pdf](https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf).
  11. McCarthy, J, 2007, What Is Artificial Intelligence?, John McCarthy’s Home Page, updated November 12, 2007, <http://www-formal.stanford.edu/jmc/index.html>.
  12. Miksad, RA, and AP Abernethy, 2018, Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-

- Grade Data Quality, *Clin Pharmacol Ther*, 103(2):202–205.
13. National Institutes of Health Collaboratory, 2014, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research, accessed August 27, 2019, [https://dcricollab.dcri.duke.edu/sites/NIHCR/KR/Assessing-data-quality\\_V1%200.pdf#search=Assessing%20data%20quality](https://dcricollab.dcri.duke.edu/sites/NIHCR/KR/Assessing-data-quality_V1%200.pdf#search=Assessing%20data%20quality).
  14. Navar, AM, ED Peterson, DL Steen, DM Wojdyla, RJ Sanchez, I Khan, X Song, ME Gold, and MJ Pencina, 2019, Evaluation of Mortality Data from the Social Security Administration Death Master File for Clinical Research, *JAMA Cardiol*, epub ahead of print Mar 6, 2019, doi: 10.1001/jamacardio.2019.0198.
  15. Porta, M, 2014, *A Dictionary of Epidemiology*, Sixth Edition, New York (NY): Oxford University Press.
  16. Richesson, RL, MM Smerek, and CC Blake, 2016, A Framework to Support the Sharing and Reuse of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications, *EGEMS*, 4(3):1232.
  17. Szklo, M, and FJ Nieto, 2006, *Epidemiology: Beyond the Basics*, 2nd Edition, Burlington (MA): Jones & Bartlett Learning.
  18. Velentgas, P, NA Dreyer, P Nourjah, SR Smith, and MM Torchia, editors, 2013, *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*, AHRQ Publication No. 12(13)–EHC099, accessed January 9, 2019, [https://effectivehealthcare.ahrq.gov/sites/default/files/related\\_files/user-guide-observational-cer-1455\\_130113.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/user-guide-observational-cer-1455_130113.pdf).

19. Wang, SV, S Schneeweiss, ML Berger, J Brown, F de Vries, I Douglas, JJ Gagne, R Gini, O Klungel, CD Mullins, MD Nguyen, JA Rassen, L Smeeth, and M Sturkenboom, 2017, Reporting to Improve Reproducibility and Facilitate Validity Assessment in Healthcare Database Studies V1.0, *Pharmacoepidemiol Drug Saf*, 26(9):1018–1032.
20. Matthew P. Fox, Richard F, MacLehose, Timothy L. Lash, 2022, *Applying Quantitative Bias Analysis to Epidemiologic Data*, Second edition: Springer
21. Stephanie Glen. "Differential & Non-Differential Misclassification" From StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/non-differential-misclassification/>