



臺北榮民總醫院

大數據中心

雙週報



期別：0001期
出刊日：2022.8.3

<https://wd.vghtpe.gov.tw/bdc>

臺北榮總COVID-19大數據研究資料集公告

各位同仁大家好：

分享好消息!「臺北榮總COVID-19大數據研究資料集」已經準備完成，所有有想法有興趣卻苦無資料同仁可以著手撰寫IRB提出申請案，COVID-19研究領域一展身手!!

過去兩年防疫期間臺北榮總不只辛苦守護台灣，院長指示我們應該更進一步，分析北榮累積寶貴研究資料(real world data)，讓榮總在世界上扮演重要領先地位。在陳院長、候副院長與高副院長指導下，大數據中心主動與資訊室、家醫部、兒童醫學部、醫企部、醫研部一起努力將本院2021/1-2022/6疫情期間資料整理並初步分成三個資料集:

1. COVID-19疫苗資料集
2. 新冠篩檢資料集
3. COVID-19確診資料集

這些資料都是珍貴的真實世界證據(real world evidence)，在疫情期間全世界很少有單位有如此完整資料，可以針對疫苗與篩檢的可靠性完整性、確診後症狀相關性等許多議題進行創新研究，歡迎同仁多加利用。

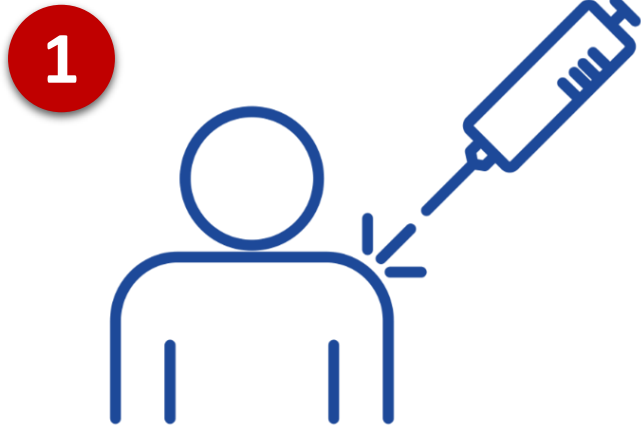
這個資料集誕生並不容易，尤其是疫情期間政策滾動式修正，每個時間點資料不一致需要統整。在院長、副院長指導下，許多同仁攜手接力串聯才有辦法完成。

首先必須瞭解資料產生過程：感謝家庭醫學部彙整並提供兒童醫學部如何在門診系統中輸入疫苗、產生疫苗紀錄、上傳系統等細部過程；感謝醫企部提供產生新冠篩檢紀錄、COVID-19確診病患如何編碼..等細節，感謝資訊室工程師們從主責單位開始到資訊室追蹤資料流(data flow)流向將病歷紀錄轉換成院內資料。接著進行資料驗證確保資料正確性，工程師必須抽樣反向從院內資料逆向回推到資訊室大型主機資料庫，以確保資料完整性，接著由大數據中心進行匿名化編碼(pseudo-anonymity)保護病患隱私，資料分割(data partition)，將資料由大型資料庫導出成研究資料表。最後大數據中心會依照各主持人IRB申請，資料欄位再次處理加密後，切分成一般研究者可以分析之文字檔案。

大數據結果往往影響深遠，提醒各位朋友，在發揮創意使用資料前請向人體試驗委員提出申請，與專家討論最適當的資料處理策略，保護主持人義務與病患隱私唷！

臺北榮總 COVID-19 大數據研究資料集

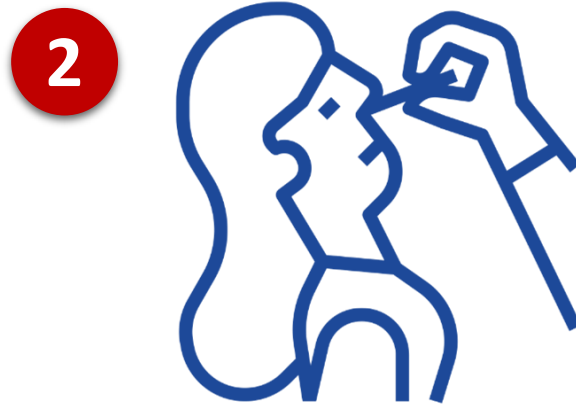
2022.8.3



疫苗資料集

141,813 人次

2021/1-2022/6 曾於本院接種 COVID-19 疫苗者，包含四種廠牌 (AZ、莫德納、BNT、高端)，五種劑型(基礎劑1 2、追加劑1 2、加強劑)



PCR篩檢資料集

89,319 人次

2021/1-2022/6 曾於本院接受過 PCR 篩檢者



COVID-19 確診資料集

3,717 人次

2021/1-2022/6 於本院門診住院 診斷碼為 COVID-19 U07.1 者

**注意: 1. 研究者須先申請 IRB 同意後再至大數據中心取得資料
2. 去個資匿名化資料，請勿嘗試破解以維護病患隱私**

1 臺北榮總 COVID-19 疫苗資料集

2022.8.3



疫苗資料集

141,813 人次

研究族群：

2021/1-2022/6 曾於本院接種 COVID-19 疫苗者(包含成人與兒童)

篩選條件：

- 成人疫苗：曾經於 COVID 成人疫苗門診(科別 0CR, 1CR)就診者

涵蓋疫苗：

- 四種廠牌(AZ、莫德納、BNT、高端)
- 五種劑型(基礎劑1、基礎劑2、追加劑1、追加劑2、基礎加強劑)

可能包含檔案內容：

- 實際取得資料與欄位以 IRB 核准為準
- 病人基本資料檔(ID)、
- 門診處方與治療明細檔(CD)、門診處方醫令明細檔(OO)
- 住院醫療費用清單明細檔(DD)、住院醫療費用醫令清單明細檔 (DO)
- 檢驗結果歷史檔(LAB)

注意: 1. 因各研究主題影響個資與隱私程度不一，研究者須先申請 IRB 同意後再至大數據中心取得資料，因此各研究主題可取得資料可能不一樣 2. 去個資匿名化資料，請勿嘗試破解以維護病患隱私

2 臺北榮總 COVID-19 PCR 篩檢資料集

2022.8.3



PCR 篩檢資料集

89,319 人次

研究族群：

2021/1-2022/6 曾於本院接受 COVID-19 篩檢者(包含自費與公費)

篩選條件：

- 自費篩檢：曾經於新冠自費篩檢門診(科別0CV, 1CV)就診者
- 公費篩檢：曾經於急診、綠色通道門診(科別2HN)就診者

可能包含檔案內容：

- 實際取得資料與欄位以 IRB 核准為準
- 病人基本資料檔(ID)、
- 門診處方與治療明細檔(CD)、門診處方醫令明細檔(OO)
- 住院醫療費用清單明細檔(DD)、住院醫療費用醫令清單明細檔 (DO)
- 檢驗結果歷史檔(LAB)

注意: 1. 因各研究主題影響個資與隱私程度不一，研究者須先申請 IRB 同意後再至大數據中心取得資料，因此各研究主題可取得資料可能不一樣 2. 去個資匿名化資料，請勿嘗試破解以維護病患隱私

臺北榮總COVID-19確診資料集

2022.8.3



COVID-19確診
3,717人次

研究族群：

2021/1-2022/6 曾於本院因COVID-19就醫者

篩選條件：

- 2021/1-2022/6 於本院門診住院診斷碼為COVID-19 U07.1 者

可能包含檔案內容：

- 實際取得資料與欄位以IRB核准為準
- 病人基本資料檔(ID)、
- 門診處方與治療明細檔(CD)、門診處方醫令明細檔(OO)
- 住院醫療費用清單明細檔(DD)、住院醫療費用醫令清單明細檔 (DO)
- 檢驗結果歷史檔(LAB)

注意: 1. 診斷碼篩選之個案可能有遺漏 2. 因各研究主題影響個資與隱私程度不一，研究者須先申請 IRB 同意後再至大數據中心取得資料，因此各研究主題可取得資料可能不一樣 3. 去個資匿名化資料，請勿嘗試破解以維護病患隱私

常見問題 (FAQ)

Q1. 『資料庫』有什麼？

A：大數據中心已串聯藥物、申請健保品項和檢驗值（格式請參考本中心網頁中『文件下載』處 新版譯碼簿excel檔），依照本院申請流程，基於保護病患隱私和PI的研究合法性，有需求者請申請IRB，由各研究團隊設計研究及經IRB委員審查，針對研究主題決定可取得欄位後，大數據中心資訊組（資訊室）將據此提供資料，為了掌握時效，建議主題可以鎖定在這個資料集準備好的項目：所有的藥物、申請健保品項和檢驗值。

Q2. 可以幫我串連xxx數值嗎？

A：大數據中心資訊組很努力替大家進行資料串聯，但是串聯資料依照複雜度不等需要時間處理。為了掌握時效，建議主題可以鎖定在這個資料集有的項目：所有的藥物、申請健保品項和檢驗值，更多需求或有資料問題者請洽：大數據中心-資訊組 高子凱工程師（分機7264轉370） | 朱原嘉 博士/高級工程師（分機7264轉501） | 陳育群 中心主任（公務機：8 # 1149） | 申請行政流程請洽 大數據中心 楊豐源 醫檢師（分機4065） | 何達凱 行政助理（分機1244） | 陳育群 中心主任（公務機：8 # 1149）

Q3. 回溯性研究有規定，限制申請範圍為申請日之前6個月的資料，請問此資料集是否也有此限制？

A：因為各個研究依照主題風險不一，資料釋出的時間 和 欄位也可能不一樣。由IRB委員審查時給予建議後執行。

感謝



指導

- 陳院長、侯副院長、高副院長指導

資料篩選條件

- 資訊室郭振宗主任、薛筱萱工程師、廖彥琪工程師、劉智方工程師、蔡采珈工程師、施榮達工程師、鄭竣元工程師
- 家庭醫學部林明慧主任、李易錚總醫師、林怡萱總醫師、李秉豪總醫師、黃哲葦專員、楊川慧研究助理
- 兒童醫學部牛道明主任、洪妙秋主任
- 醫務企管部彭家勛主任、莊媿婷專員、張祐瑄組員、彭玉娥組員

資料集製作

- 大數據中心資訊組朱原嘉高級工程師、高子凱工程師、黃欣儀工程師

文件/行政製作

- 大數據中心楊豐源醫檢師、何達凱行政助理

行政支援

- 大數據中心、醫學研究部

大數據中心 Big Data Center



大數據中心 Big Data Center

聯絡資訊

- 大數據中心-資訊組 高子凱 工程師 (分機7264轉370)
朱原嘉 博士/高級工程師 (分機7264轉501)
陳育群 中心主任 (公務機：8# 1149)
- 申請行政流程請洽 大數據中心 楊豐源 醫檢師 (分機4065)
何達凱 行政助理 (分機1244)
陳育群 中心主任 (公務機：8# 1149)

