

科目：生物醫學研究導論

系所組：生物醫學海量資料分析
碩學位學程

台灣地區全民健保自 1995 年開辦以來，全民健保資料庫累積了台灣兩千三百萬人口，自健保開辦以後的所有健保就醫資料。中央健保署自 2000 年起，委託國家衛生研究院發行全民健保資料庫，提供學術界使用全民健保資料庫來進行臨床醫學與公共衛生的研究，成為台灣地區學術界研究的寶藏。請針對以下使用台灣地區全民健保資料庫所進行之醫學研究問題作答。

一、因應社會之高齡化，帕金森氏症人口逐年增加，造成醫療負擔的增加。再者，帕金森氏症是慢性、退化性且不可預測過程的疾病，對於病人、醫生與照顧者都是一大挑戰，而憂鬱症是帕金森氏症常見的合併症，約 40% 至 50% 的患者合併有憂鬱症，若患者合併憂鬱症，將會大幅降低其生活品質，其重要程度可是不容小覷。若以台灣地區全民健保資料庫來探討服用抗帕金森藥物之患者，其藥物服用種類與其發展成憂鬱症之間的關係：

1. 你會採用何種研究設計來探討此一研究主題？(10 分)
2. 此一研究所面臨之 confounding by indication 的偏差為何？你要如何處理此一偏差？(40 分)

二、自從荷爾蒙補充療法被發現會提高乳癌發生風險後，學界對於中藥中所含的植物性荷爾蒙同樣會有提高乳癌風險的顧慮，尤其是最常使用在改善婦科更年期症狀的當歸。在當歸相關研究中，早期有些學者發現當歸萃取液會刺激乳癌細胞的增生。但後續試驗中所觀察到的結果卻非呈現一致性，甚至有觀察到當歸萃取物有抑制雌激素活性或促進乳癌細胞凋亡等抑癌作用。也因為缺乏關鍵性的人體對照實驗佐證，因此在當歸與乳癌風險的議題上，目前學界還是無共識。若要利用台灣地區全民健保資料庫來探討當歸等含植物性荷爾蒙藥材對於女性乳癌發生之關聯性，你會怎麼做？(50 分)

- ※ 注意：1. 考生須在「彌封答案卷」上作答。
2. 本試題紙空白部份可當稿紙使用。
3. 考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。

科目：海量資料分析導論

系所組：生物醫學
海量資料分析
碩士學位課程

1. If...else 是常見的程式語言，請解釋其涵義並舉例說明。(20分)
2. 請解釋統計假設檢定中的顯著水準、type 1 error 和 type 2 error 涵義?(20分)
3. 2018年 John P.A. Ioannidis 博士在 JAMA 發表了一篇「The Proposal to Lower P Value Thresholds to .005」的文章，請解釋何謂 p 值? 您認為降低 p 值的閾值對於海量資料分析研究有何影響?(20分)
4. 目前有許多商業標榜藉由 Artificial Intelligence (AI) 所發展的疾病預測模型 (prediction model) 可準確預測個人長期罹病風險，您認為這樣的產品若能實際應用於醫學，從統計學上的觀點需要經過哪些評估過程或驗證?(20分)
5. 資料分析軟體眾多，請分享您使用過的軟體，並擇一說明其優缺點。(20分)

※ 注意：1. 考生須在「彌封答案卷」上作答。

2. 本試題紙空白部份可當稿紙使用。

3. 考生於作答時可否使用計算機、法典、字典或其他資料或工具，以簡章之規定為準。