

111年度衛生福利部衛生福利資料科學中心

統計軟體與研究實務課程

橫斷研究及病例對照研究

使用統計軟體R及SAS為例

講師：劉品崧 統計分析師

花蓮慈濟醫院

課程大綱

- 課程前導
- 範例主題說明
 - 缺血性中風與出血性中風的住院特性比較(cross-sectional)
 - 缺血性中風與出血性中風與住院天數相關性探討(case-control)
- 軟體實際演練
- 討論與結論

過去課程精華收錄



• 2020年：基礎課程

- R軟體基本應用，使用data.table套件進行資料處理



109年度R基礎課程-劉品崧老師



<https://github.com/PSLiu/NHIRD-R-2020>



• 2021年：進階課程

- 統計繪圖、存活分析、CCI指標、群組資料、預測模型



<https://github.com/PSLiu/NHIRD-R-2021>

2022課程推廣重點：研究實務

- 讓大家瞭解研究設計到統計分析的實務流程
 - 定義問題、研究設計、理解資料、流程管理、統計分析
 - 同時開授SAS軟體以及R軟體的場次
 - 你可以拿老師的code去改，但你一定先理解我原本在做什麼
- 研究設計類型
 1. 橫斷研究與病例對照研究：4/8(五)慈大、4/11(一)成大、4/23(六)北醫
 2. 世代追蹤與重複測量研究 (籌備中)
 3. 混淆因子與傾向分數應用 (籌備中)




課前學員自我準備

- 下載課前寄發材料([Google Drive連結](#))
 - 解壓縮所有材料，**不建議**放在**One Drive** 或 路徑當中含有中文之地方
- 路徑請修改為自己放置資料的地方
- **R場次**的學員
 - 請將RStudio的指令稿設定為**UTF-8**儲存
 - Tools → Global options → Code → Saving → Default text encoding → UTF-8
 - 自行下載課程當中所需的套件，瞭解將會使用的函數
 - 本課程將不會花太多時間講解細部指令語法

課程參與注意事項

- 有問題隨時可以舉手/開麥克風發問，問問題很重要！
- 課程所使用模擬檔數據不得作為研究發表用途
 - 因為模擬數據本來就是假的 XD
- 本課程的操作是基於講師的個人邏輯，不代表所有研究皆可適用
 - 研究方法、生物統計、統合分析有學理基礎架構
 - 依據資料、產業、醫療科別不同，需要對應的domain knowledge
 - 本次課程無法容納之處請大家一定要先看書+看paper瞭解之後再用

課程材料版權聲明

- 本次課程的所有內容依據台灣創用CC條款進行授權
 - 講義
 - 程式碼
 - 但不包含衛生福利部統計處提供之模擬資料檔，僅提供報名學員使用
- 授權類別 
 - 姓名標示—非商業性—禁止改作
 - 本授權條款允許使用者重製、散布、傳輸著作，但不得為商業目的之使用，亦不得修改該著作。使用時必須按照著作人指定的方式表彰其姓名。

範例主題說明

- 缺血性中風與出血性中風的
住院特性比較(cross-sectional)
- 缺血性中風與出血性中風與
住院天數相關性探討(case-control)

研究時間軸圖示

相關因子探討

住院特性比較

Case-control study

Cross-sectional study

1. 分組：住院>7天(case) or ≤7天
2. 暴露：缺血中風 vs 出血中風
3. 配對：年齡、性別
4. 校正：病史、投保金額級距

缺血性中風
Ischemic stroke



出血性中風
Hemorrhagic stroke

1. 住院天數(length of stay)
2. 併發肺炎(pneumonia)
3. 鼻胃管(nasogastric intubation)
4. 尿管(urinary catheterization)

Logistic regression

描述性統計：N (%) or mean ± SD

過去

住院期間當中

衛生福利資料

- 常用資料庫名稱 (檔案代號)

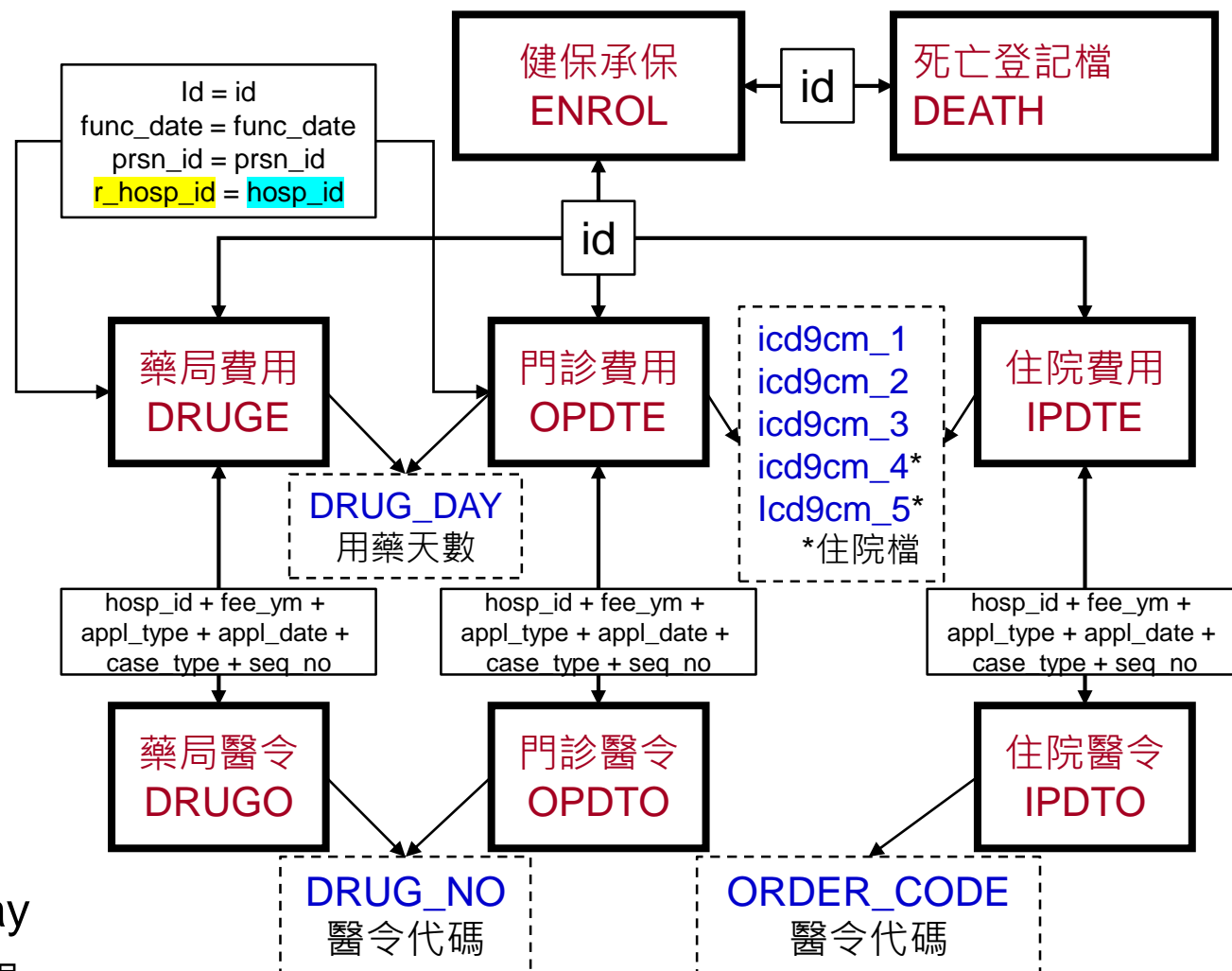
- 門診費用檔 (OPDTE)
- 門診醫令檔 (OPDTO)
- 住院費用檔 (IPDTE)
- 住院醫令檔 (IPDTO)
- 健保承保檔 (ENROL)
- 死因統計檔 (DEATH)

- 資料庫手冊 (也有給各位喔！)

- 用於定義的欄位：icd、order、date、day
- 用於串聯的欄位：身分證號、時間、空間

- 本次範例資料為2014年的10萬人模擬資料

- 迴圈並非本次重點，故老師已經先將整年12月份資料進行合併



操作定義：缺血中風與出血中風

- 缺血性中風 Ischemic stroke
 - ICD-9-CM：433、434
- 出血性中風 Hemorrhagic stroke
 - ICD-9-CM：430、431、432
- 使用住院檔的第一個診斷 (*icd9cm_1*)為主

操作定義：住院特性比較

- 住院天數
 - 急性病床天數(**e_bed_day**)與慢性病床天數(**s_bed_day**)之總和
- 併發肺炎：主診斷以外的欄位有被下肺炎的診斷
 - ICD-9-CM：480 - 486
 - 以住院檔第2 - 5診斷碼欄位(**icd9cm_2** - **icd9cm_5**)判斷
- 鼻胃管：醫令碼(**order_code**)為47017C、47018C
- 尿管：醫令碼(**order_code**)為47013C、47014C

操作定義：相關因子探討(控制變項)

- 性別
 - 承保檔的性別(**id_s**)欄位，分為男、女
- 年齡
 - 住院檔的住院日期(**in_date**)年份 - 承保檔的出生年(**id_birth_y**)
- 投保金額(**id1_amt**)級距
 - 依附投保 (**投保金額** < NTD **15,840**)
 - NTD **15,840** \leq **投保金額** < NTD **25,000**
 - NTD **25,000** \leq **投保金額**

操作定義：相關因子探討(疾病史 ICD-9-CM)

- 住院日(不含)前90天有出現至少2次門診或1次住院診斷碼
 - Diabetes : 250
 - Hypertension : 401 - 405
 - Disorders of lipoid metabolism : 272
 - Chronic kidney disease : 585
 - Cardiac dysrhythmias : 427
 - Stroke history : 430 - 438
 - Pneumonia history : 480 - 486

工作心流

原始資料

住院醫令

住院費用

門診費用

健保承保

處置欄位

住院資訊

診斷欄位

承保資料欄位

放置鼻管
放置尿管中風診斷
肺炎診斷
住院天數

研究樣本index date前90天資料

研究樣本index date前
最近一次資料

標記相關疾病因子

投保金額級距分類

中間資料

join

費用醫令合併資料

疾病史紀錄

加總病史權重

保留符合條件紀錄

歸戶處理

歸戶處理

中風住院歸戶資料

疾病史歸戶資料

投保歸戶資料

join

分析資料

最終分析資料

Final remark

- 思考核心問題

- 研究目的
- 操作定義

- 建構工作心流

- 可以執行定義的材料
- 可以回答問題的設計
- 可以統計分析的資料

- 應用技術實踐

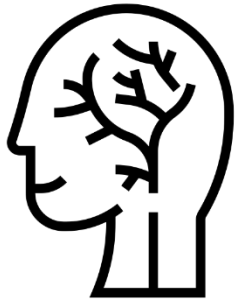
- 劉品崧

- 研究設計
- 統計分析
- 策略諮詢



- **Email** : psliu520@gmail.com
- **PubMed** : Peter Pin-Sung Liu

使用圖片版權來源宣告



Created by Eucalyp



Created by Eucalyp