

113年度衛生福利資料科學中心統計軟體推廣課程

2024.07.01 (一) 13:00-16:00 @ 高雄醫學大學 IR503電腦教室

健康資料研究與R軟體

《衛生福利資料進階班：世代追蹤研究》

劉品崧 統計諮詢暨分析師

花蓮慈濟醫院

課程大綱

- 課程說明
- 研究議題
- 實作練習

根據研究議題完成資料管理及統計分析

- 學員背景

- 已經稍微接觸過健保資料庫、R軟體、生物統計，**但不完全熟悉觀察性研究設計**

- 課程設計

- 理解計畫：閱讀執行計畫書內容
 - 研究設計：時間軸、時間軸、時間軸
 - 資料來源：門診、住院、藥局、承保、死亡
 - 關鍵變數：日期、診斷、醫令、藥品、ATC code、正規表達式
 - 連結變數：身分證號、日期、費用 / 醫令檔串聯6變項

現今AI工具
尚未克服的障礙

- 未來應用

- 發想計畫、給予定義、親手分析

世代追蹤研究 Cohort study

- 目標族群的介入組 (intervention)

Population + exposure

- 目標族群的對照組 (comparator)

Population + non-exposure / other expose / active comparator

- 追蹤兩組未來的事件發生狀況 (outcome)

Interested event during follow-up

- 探討介入分組與未來事件發生的相關性 (association)

Statistical analysis

探討NOAC與Warfarin對病人存活率之相關性

- 目標族群的介入組 (intervention)

NOAC使用者 with 過去中風/心律不整病史

- 目標族群的對照組 (comparator)

切記不可以用追蹤過程的事件進行分組
否則會導致immortal-time bias

Warfarin使用者 with 過去中風/心律不整病史

- 追蹤兩組未來的事件發生狀況 (outcome)

死亡事件

- 探討介入分組與未來事件發生的相關性 (association)

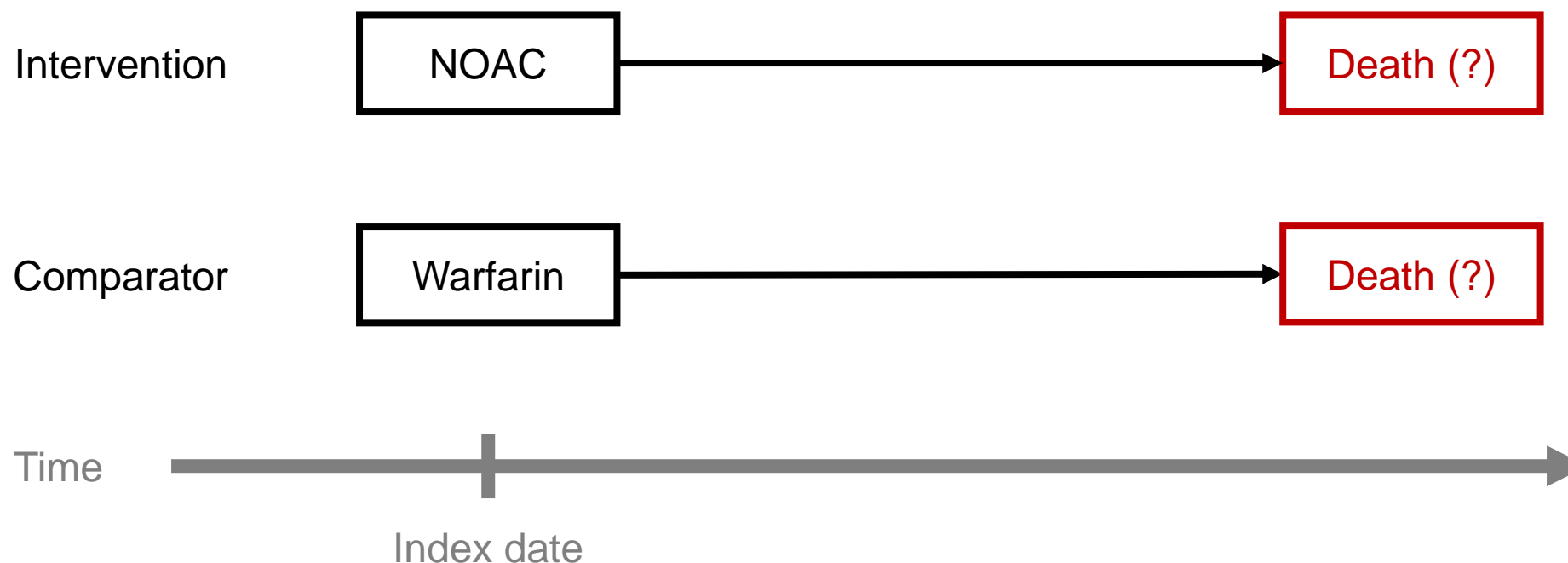
發生率、KM plot、Cox regression model

探討NOAC與Warfarin對病人存活率之相關性

- 課程資料夾：`study-design-cohort`
- 研究執行計畫書
 - `study-design-cohort-研究執行計畫書.pdf`
- 檔案資料夾
 - `codebook`：衛生福利模擬資料檔案之說明文件
 - `base`：衛生福利模擬資料檔案、藥品清單檔案 (`nhi_drug_list_20240229.csv`)
 - `temp`：資料整理過後從R軟體輸出到硬碟的存放位置
- 課程練習程式碼
 - `study-design-cohort-ST.R` (Encoding = UTF-8)
 - 要將挖空部分 (#####) 補齊







將核心理念進行圖像化







- 探討NOAC與Warfarin對病人存活率之相關性




衛生福利模擬資料檔 (虛擬10萬人2014年資料)

- 原始型態
 - 不同月份資料分開儲存
 - 實際模樣
 - 需要迴圈處理
- 堆疊型態 (本日使用)
 - 整個年度一起儲存
 - 教學使用
 - 不要執著技術層面

 h_nhi_enrol201401.csv
 h_nhi_enrol201402.csv
 h_nhi_enrol201403.csv
 h_nhi_enrol201404.csv
 h_nhi_enrol201405.csv
 h_nhi_enrol201406.csv


 h_nhi_enrol201407.csv
 h_nhi_enrol201408.csv
 h_nhi_enrol201409.csv
 h_nhi_enrol201410.csv
 h_nhi_enrol201411.csv
 h_nhi_enrol201412.csv

 h_nhi_enrol2014.csv


Warning !!!!!!!
教學使用虛擬資料
不得作為研究素材


衛生福利模擬資料檔及譯碼簿

- 模擬資料檔案
 - base資料夾
- 使用手冊（譯碼簿）
 - codebook資料夾


 h_nhi_opdte2014.csv


 h_nhi_ipdte2014.csv


 h_nhi_drugte2014.csv


 h_nhi_opdto2014.csv


 h_nhi_ipdto2014.csv


 h_nhi_drugo2014.csv


 h_nhi_enrol2014.csv


 h_ost_death2014.csv


 模擬資料庫使用手冊Health01_全民健保處方及治療明細檔_門急診.pdf


 模擬資料庫使用手冊Health02_全民健保處方及治療明細檔_西醫住院.pdf


 模擬資料庫使用手冊Health03_全民健保處方及治療明細檔_藥局.pdf

 模擬資料庫使用手冊Health04_全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診.pdf

 模擬資料庫使用手冊Health05_全民健保處方及治療醫令明細檔—住院.pdf

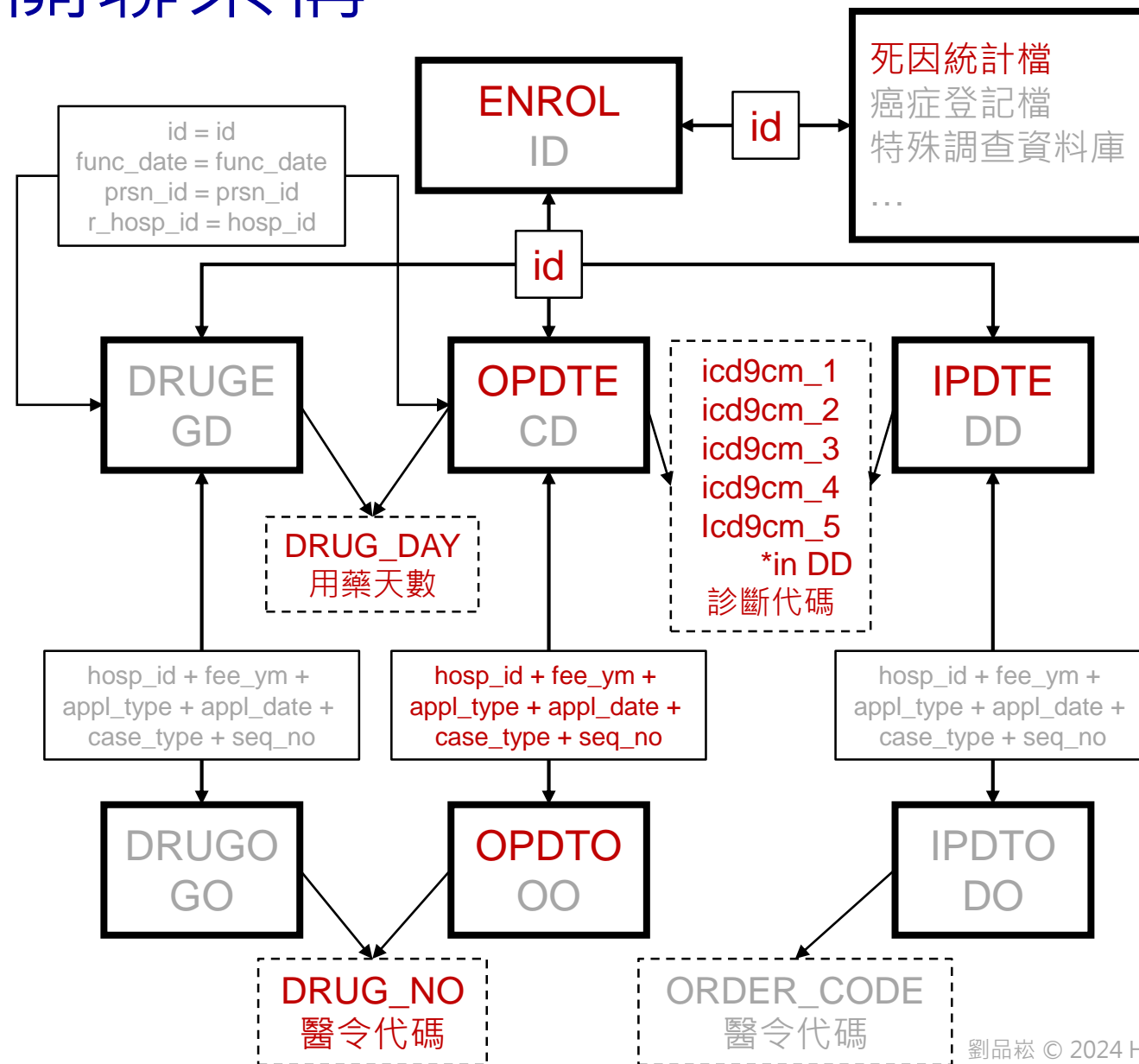
 模擬資料庫使用手冊Health06_全民健保處方及治療醫令明細檔—藥局.pdf

 模擬資料庫使用手冊Health07_全民健保承保檔.pdf

 模擬資料庫使用手冊Health10_死因統計檔.pdf

衛生福利模擬資料檔關聯架構

- 預設各位已經認識檔名與架構
- 本次課程實際使用範圍
 - 門診費用檔 (OPDTE)
 - 門診醫令檔 (OPDTO)
 - 住院費用檔 (IPDTE)
 - 住院醫令檔 (IPDTO)
 - 藥局費用檔 (DRUGE)
 - 藥局醫令檔 (DRUGO)
 - 健保承保檔 (ENROL)
 - 死因統計檔 (DEATH)



正規表達式 Regular expression

找出icd_code字串裡面為433開頭或是434開頭的觀察值

```
dt[grepl("^433|^434", icd_code)]
```

找出icd_code字串裡面為43開頭，後面接著3或4的觀察值

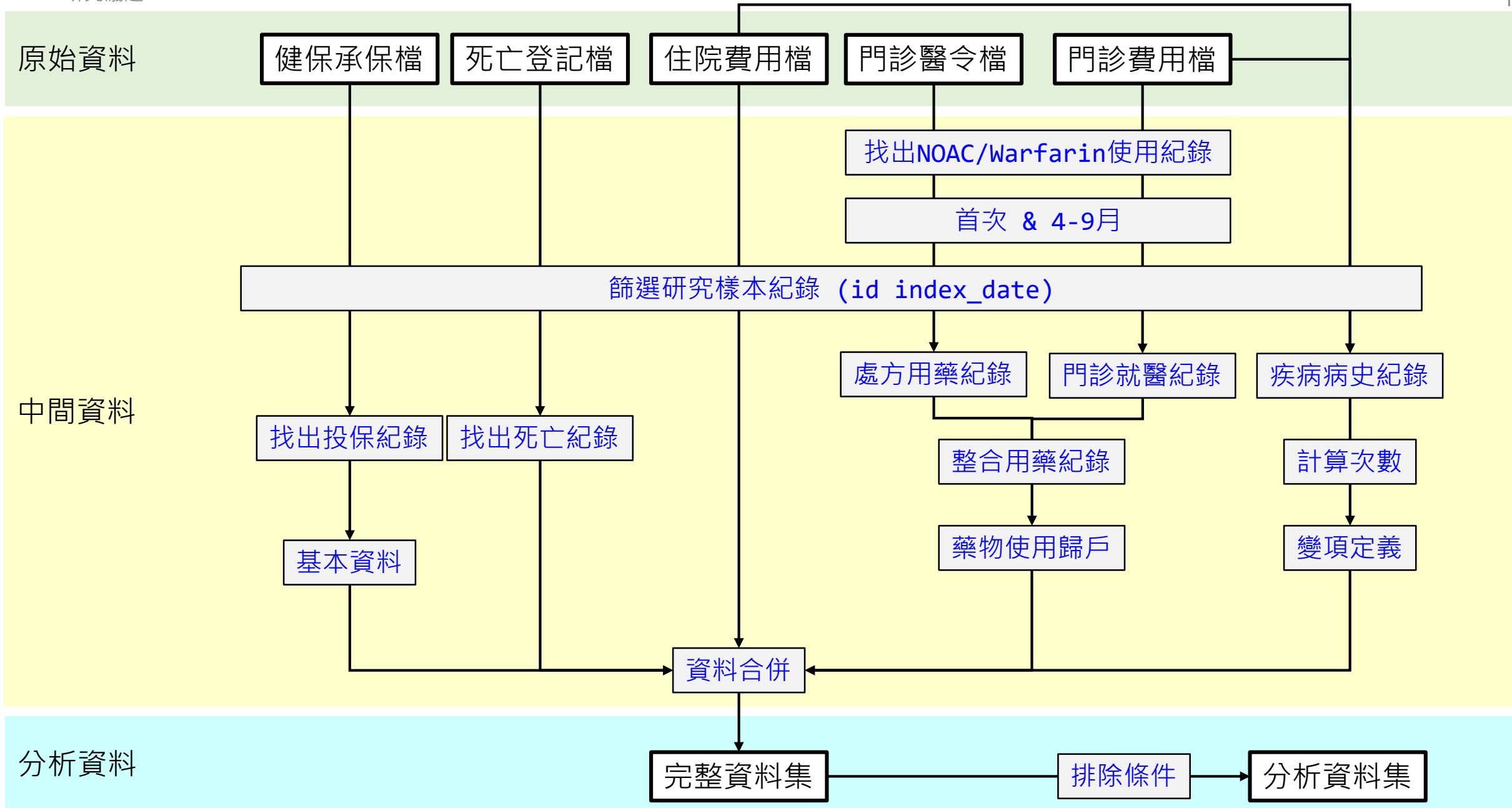
```
dt[grepl("^43[34]", icd_code)]
```

找出icd_code字串裡面為430-438開頭

```
dt[grepl("^430|^431|^432|^433|^434|^435|^436|^437|^438", icd_code)]
```

找出icd_code字串裡面為43開頭，後面接著0 ~ 8的觀察值

```
dt[grepl("^43[0-8]", icd_code)]
```



Summary

- 研究議題
- 實作練習
 - 訂定主題
 - 研究設計
 - 給予定義
 - 思考流程
 - 撰寫計畫
 - 資料管理
 - 統計分析
- 開放提問時間
- Facebook : 劉品崧
- PubMed : Peter Pin-Sung Liu
- Mail : psliu520@gmail.com
- GitHub : <https://github.com/PSLiu/>