

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

世代追蹤及重複測量研究

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：DM病人的胰島素使用與低血糖急症
 - 研究設計1：世代追蹤(cohort study)
 - 研究設計2：重複測量(repeated measurement)
- 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：4/8(五)慈大、4/11(一)成大、4/23(六)北醫
2. 追蹤研究：8/15(一)高醫、8/16(二)成大、8/17(三)國衛院、8/18(四)台科大
3. 混淆因子 (籌備中)

111年度衛生福利部
衛生福利資料科學中心

統計軟體 研究實務課程

課程資訊：

111/8/15 (一) 09:00-12:00、13:00-16:00
高雄醫學大學 國研大樓五樓IR503電腦教室
講師：花蓮慈濟醫院高齡暨社區醫學部 統計分析師-劉品崧

111/8/16 (二) 09:00-12:00、13:00-16:00
國立成功大學 計算機與網路中心75209電腦教室
講師：花蓮慈濟醫院高齡暨社區醫學部 統計分析師-劉品崧

111/8/17 (三) 09:00-12:00、13:00-16:00
國家衛生研究院 國資大樓5樓L5203電腦教室
講師：花蓮慈濟醫院高齡暨社區醫學部 統計分析師-劉品崧

111/8/18 (四) 09:00-12:00、13:00-16:00
國立臺灣科技大學 圖書館LB001教室
講師：花蓮慈濟醫院高齡暨社區醫學部 統計分析師-劉品崧
國立臺灣科技大學資訊工程系 助理教授-陳怡伶

QR code連結
R軟體線上基礎課程QR code連結

報名網址：
<https://forms.gle/BvCA1xjidedUhyCdA>

課程大綱：
存活分析及重複測量研究—
使用統計軟體R及SAS為例
上午09:00-12:00 R軟體
下午13:00-16:00 SAS軟體

指導單位：衛生福利部統計處
主辦單位：衛生福利資料科學中心、慈濟大學研究分中心
協辦單位：衛生福利資料科學中心、國立成功大學藥學系、國立成功大學臨床藥學與藥物科技研究所、
成大醫院健康數據科學中心、國家衛生研究院生醫資源中心、衛生福利資料科學中心、國家衛生研究院研究分中心、
衛生福利資料科學中心、高雄醫學大學研究分中心、高雄醫學大學醫學資訊與統計中心、國立臺灣科技大學資訊工程學系、
教育部智慧創新領域人才培育聯盟計畫


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，**先不要開啟script**，先修改RStudio設定
 - 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

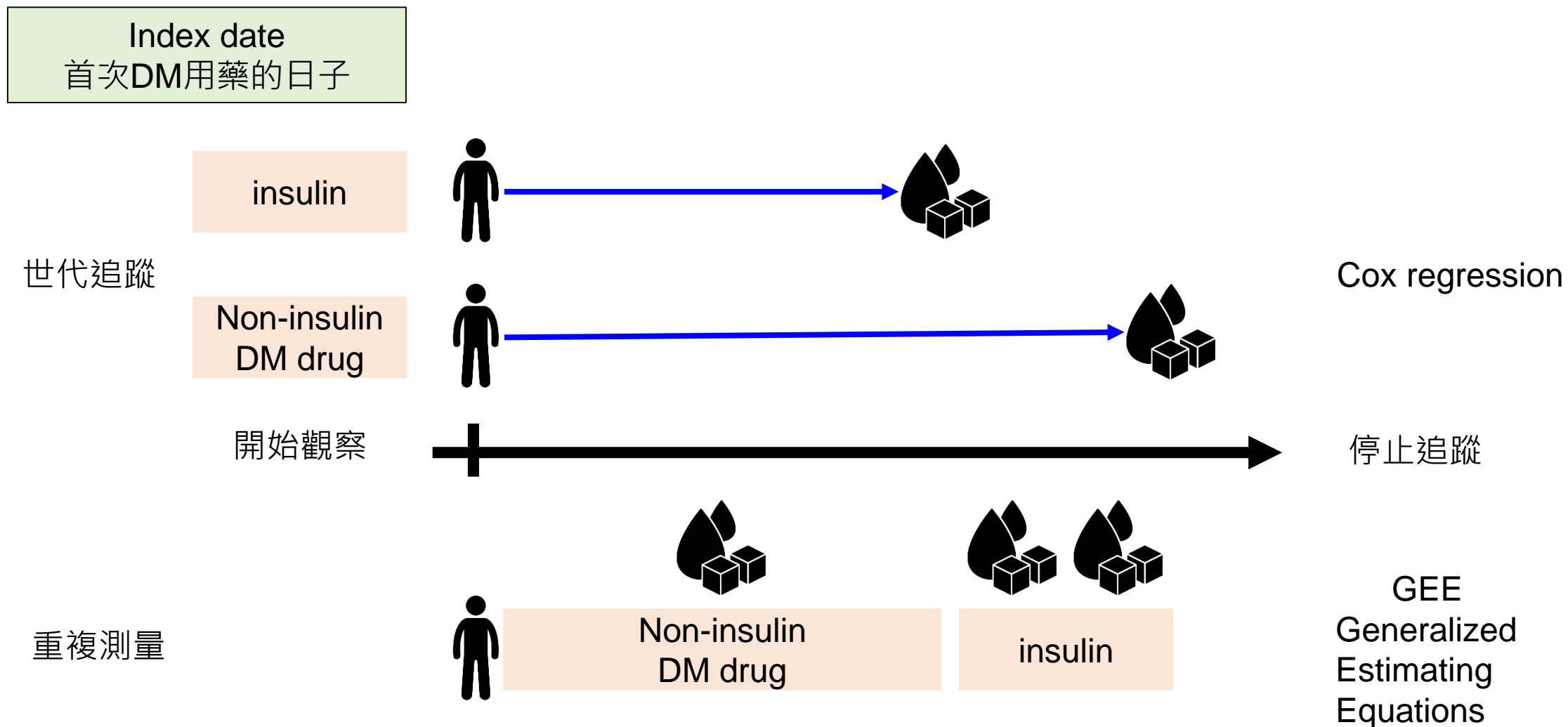
課程材料版權聲明

- 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

- 胰島素使用與低血糖急症的世代追蹤研究(cohort study)
- 胰島素使用與低血糖急症的重複測量設計
(repeated measurement)

研究時間軸圖示



衛生福利資料

- 常用資料庫名稱 (檔案代號)

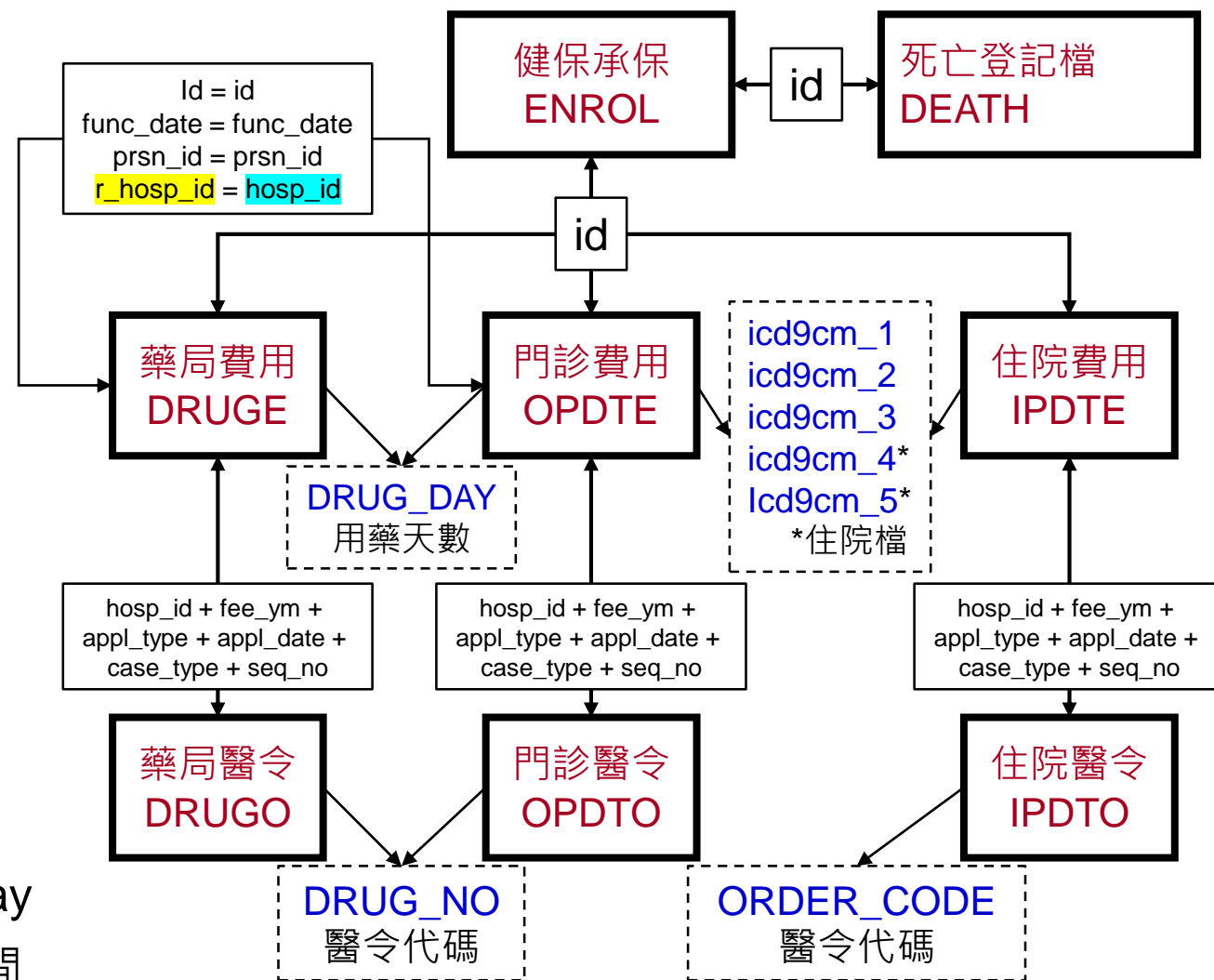
- 門診費用檔 (OPDTE)
- 門診醫令檔 (OPDTO)
- 住院費用檔 (IPDTE)
- 住院醫令檔 (IPDTO)
- 健保承保檔 (ENROL)
- 死因統計檔 (DEATH)

- 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：DM用藥

- DM用藥
 - ATC code：A10開頭
- DM用藥 - insulin
 - ATC code：A10A開頭
- 從健保藥品檔當中抓取藥物清單
- 在從門診、藥局檔篩選
- 利用給藥天數計算要用區間

操作定義：低血糖急症

- 低血糖：任一診斷碼欄位被下hypoglycemia的診斷
 - ICD-9-CM：251.0, 251.1, 251.2
- 診斷來源
 - 住院檔
 - 急診(case_type = "02" | func_type = "22")

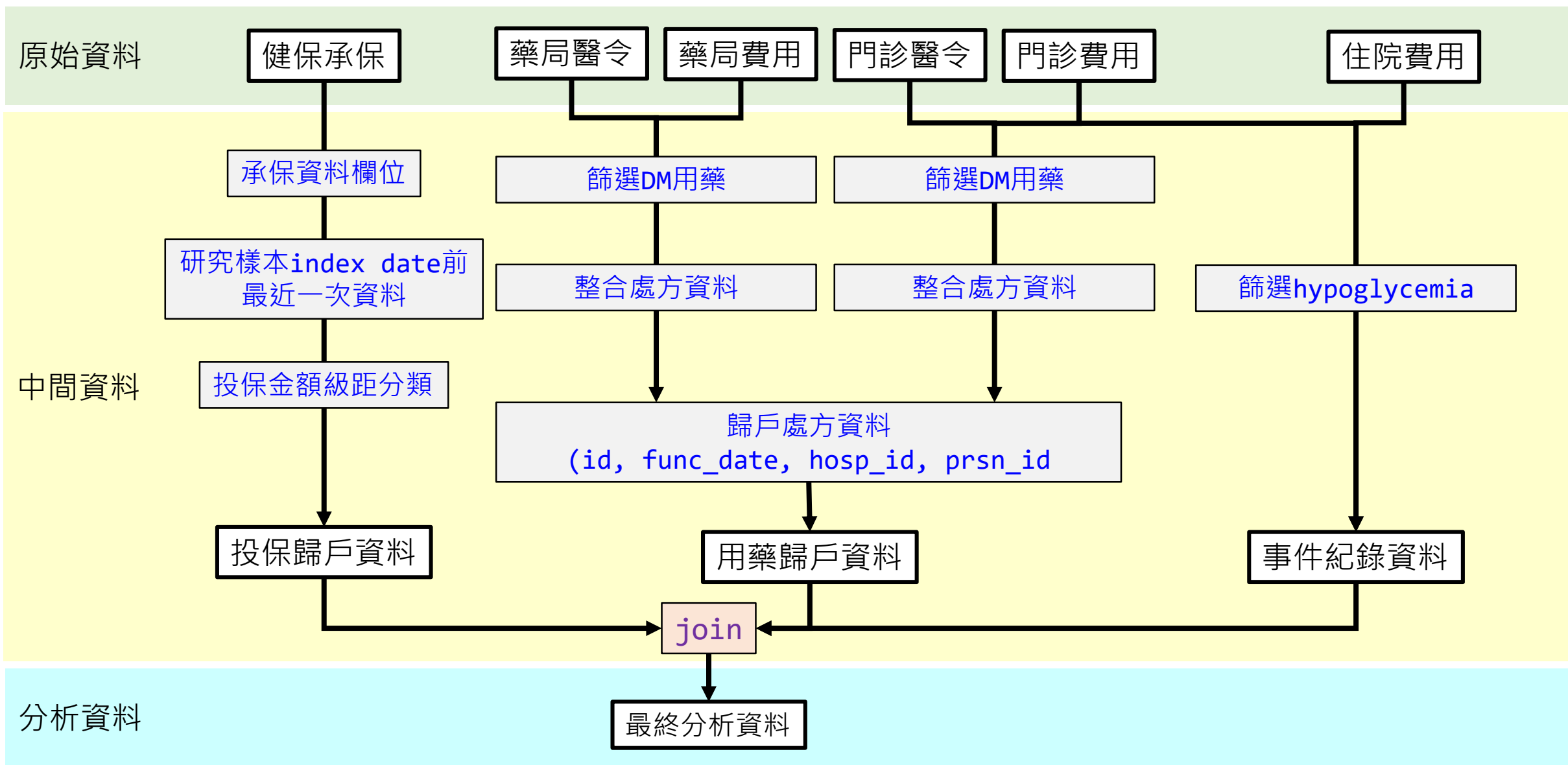
操作定義：相關因子探討(控制變項)

- 性別：承保檔的性別(**id_s**)欄位，分為男、女
- 年齡
 - 指標日期(**index_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** < NTD 45,000
 - NTD 45,000 \leq **投保金額**

操作定義：相關因子探討(疾病史 ICD-9-CM)

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

工作心流



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

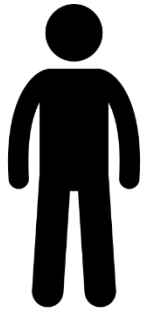
- 劉品崧

- 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 Gas Khoun Lay
from the Noun Project



Created by Lars Meier toberens
from the Noun Project