

健康資料管理與研究實務

衛生福利資料的研究設計與資料管理

《病例對照研究 Case-control study》

劉品崧 統計諮詢分析師 / 組長

花蓮慈濟醫院高齡暨社區醫學部

112年度資料管理與研究實務（下半年）

• 課程列表

日期	時間	地點	主題	軟體
10/06(五)	13：30 - 16：30	臺北醫學大學信義校區	統計軟體R與SAS在資料管理與統計分析之應用	SAS+R
10/16(一)	09：00 - 12：00	臺北醫學大學雙和校區	衛福資料庫之研究設計與統計分析：病例對照研究	SAS
10/16(一)	13：30 - 16：30	臺北醫學大學雙和校區	衛福資料庫之研究設計與統計分析：病例對照研究	R
10/20(五)	09：00 - 12：00	國家衛生研究院（苗栗）	衛福資料庫之研究設計與統計分析：世代追蹤研究	SAS
10/20(五)	13：30 - 16：30	國家衛生研究院（苗栗）	衛福資料庫之研究設計與統計分析：世代追蹤研究	R
10/28(六)	09：00 - 16：30	慈濟大學（花蓮）	衛福資料庫之研究設計與統計分析：病例對照研究	SAS
11/03(五)	09：00 - 12：00	高雄醫學大學	統計軟體R與SAS在資料管理與統計分析之應用	SAS+R
11/06(一)	09：00 - 12：00	國立成功大學	衛福資料庫之研究設計與統計分析：世代追蹤研究	SAS
11/06(一)	13：30 - 16：30	國立成功大學	衛福資料庫之研究設計與統計分析：世代追蹤研究	R

112年度資料管理與研究實務（下半年）

- 課程設計理念
 - 思考研究設計、實際資料管理、完成統計分析
- 學習目標重點
 - 追求邏輯貫通、分享實戰經驗
- 課前具備基礎
 - 軟體操作（R / SAS）、流行病學、研究設計、生物統計

課程注意事項

- 兩個承諾
 - 每50分鐘休息10分鐘，讓各位intake / output
 - 過程當中隨時可以打斷我，問題留給我，收穫你帶走
- 兩個不可以
 - 課程練習資料為模擬資料檔，不可以直接用於實際研究用途
 - 操作定義僅供教學演練使用，不可以直接用於實際研究用途

課程大綱

- 病例對照研究
- 研究範例說明
- 實作練習時間

病例对照研究 Case-control study

- 利用發生觀察事件的病例 (case)

Population + event occur

- 比較沒有發生事件的對照 (control)

Population + event absence

- 兩組之間的暴露狀況 (exposure)

Interested history

- 探討過去暴露與事件發生的相關性 (association)

Statistical analysis

肺炎住院病人的入院前CCI與入院後90天死亡之相關性

- 利用發生觀察事件的病例（ case ）

肺炎住院病人在入院後90天內死亡

- 比較沒有發生事件的對照（ control ）

肺炎住院病人在入院後90天內存活

- 兩組之間的暴露狀況（ exposure ）

入院之前的CCI指標分數

- 探討過去暴露與事件發生的相關性（ association ）

Logistic regression

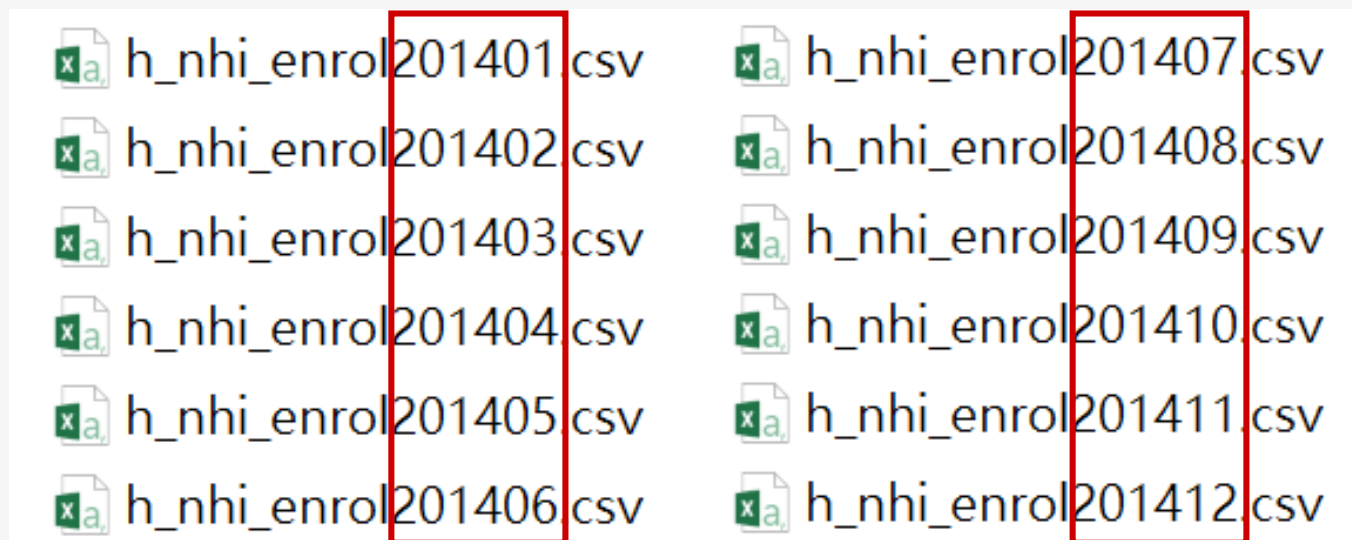
研究材料：衛生福利資料庫模擬資料檔

- 10萬人模擬資料檔，資料年度為2014年1月～12月
- 主要用途
 - 認識資料輪廓
 - 瞭解資料關聯
 - 樣本數量估算
 - 測試程式結果
- 今日主軸五大常用資料庫
 - 門診、住院、藥局、承保、死因

研究材料：衛生福利資料庫模擬資料檔

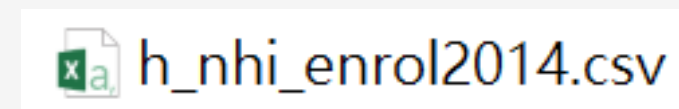
- 原始型態

- 不同月份資料分開儲存
- 實際模樣
- 需要迴圈處理



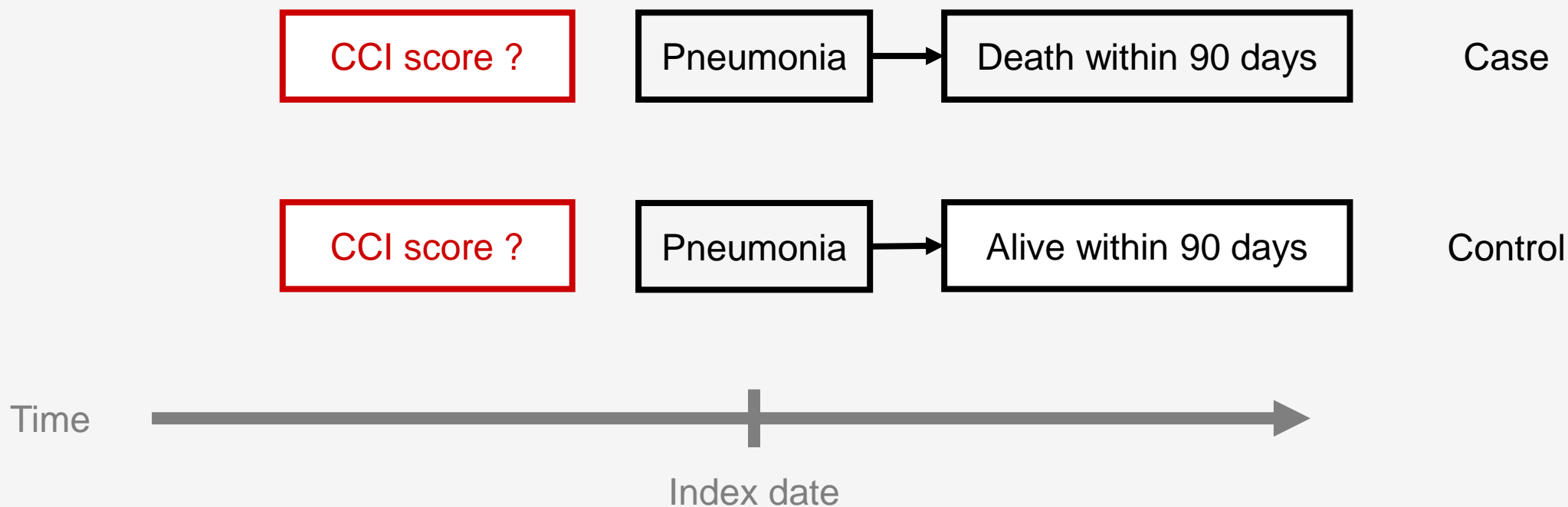
- 堆疊型態 (本日使用)

- 整個年度一起儲存
- 教學使用
- 不要執著技術層面



將核心理念進行圖像化

- 肺炎住院病人的入院前CCI與入院後90天死亡之相關性

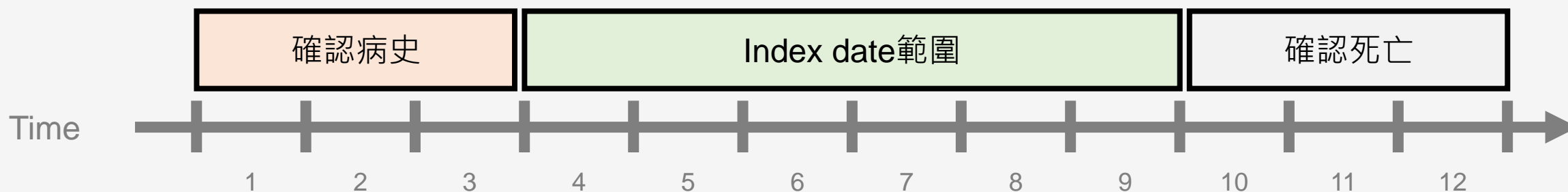


將核心理念進行拆解區塊工作項目

1. 找出研究樣本
2. 連結死亡登記，確認住院後死亡事件的發生時間（是否在90天內）
3. 撈取健保承保資料，確認研究樣本基本資料
4. 撈取健保承保資料，將投保金額級距進行分類
5. 觀察過去3個月內的就醫紀錄，計算CCI score
6. 把整理好的資料儲存下來，並且合併資料
7. 統計分析

研究樣本：肺炎住院病人

- 住院費用檔有研究樣本診斷碼
 - 任一診斷碼欄位：icd9cm_1、icd9cm_2、icd9cm_3、icd9cm_4、icd9cm_5
 - ICD-9-CM：480 — 486
- 以日期排序後取出每人在2014年的首次住院
- 首次住院的日期僅限於4 — 9月
 - 過去至少有3個月可以找出病史，未來至少有3個月可以確認死亡



終點事件：入院後90天內死亡

- 從全國的死亡登記檔找出研究樣本的資料
 - 身分證號：**id**
 - 死亡日期：**d_date**
- 將 d_date (文字型態) 轉換為日期型態
- 與 index_date 比較是否在90天內

基本資料：健保投保的出生年與性別

- 從全國的健保承保檔找出研究樣本的資料
 - 基本資料：`id, id_s, id_birth_y`
 - 投保月份：`prem_ym`
- 年齡：出生年度 - 住院年度，不可以 <0 歲或是 >100 歲
- 性別編碼只能是1（男性）和2（女性）
- 取投保月份最早的一筆

社經狀況：健保投保的保費

- 從全國的健保承保檔找出研究樣本的資料
 - 投保級距：**id1_amt**
 - 投保月份：**prem_ym**
 - 原始格式為yyyymm，把投保日假定為每月1日，轉換為日期格式 yyyymmdd
- 找出距離住院日期（不含）前最近一次的投保紀錄
- 投保金額級距切分
 - (1) 15,840以下
 - (2) 15,840 — 29,999
 - (3) 30,000 — 44,999
 - (4) 45,000以上

Charlson comorbidity index, CCI (1 / 4)

- 由17類共病症所建立的一個指標 (index) 可用於預測未來死亡風險

TABLE 1. ICD-9-CM and ICD-10 Coding Algorithms for Charlson Comorbidities

Comorbidities	ICD-10	Enhanced ICD-9-CM
Myocardial infarction	I21.x, I22.x, I25.2	410.x, 412.x
Congestive heart failure	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5–I42.9, I43.x, I50.x, P29.0	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4–425.9, 428.x
Peripheral vascular disease	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9	093.0, 437.3, 440.x, 441.x, 443.1–443.9, 47.1, 557.1, 557.9, V43.4
Cerebrovascular disease	G45.x, G46.x, H34.0, I60.x–I69.x	362.34, 430.x–438.x
Dementia	F00.x–F03.x, F05.1, G30.x, G31.1	290.x, 294.1, 331.2
Chronic pulmonary disease	I27.8, I27.9, J40.x–J47.x, J60.x–J67.x, J68.4, J70.1, J70.3	416.8, 416.9, 490.x–505.x, 506.4, 508.1, 508.8
Rheumatic disease	M05.x, M06.x, M31.5, M32.x–M34.x, M35.1, M35.3, M36.0	446.5, 710.0–710.4, 714.0–714.2, 714.8, 725.x
Peptic ulcer disease	K25.x–K28.x	531.x–534.x
Mild liver disease	B18.x, K70.0–K70.3, K70.9, K71.3–K71.5, K71.7, K73.x, K74.x, K76.0, K76.2–K76.4, K76.8, K76.9, Z94.4	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570.x, 571.x, 573.3, 573.4, 573.8, 573.9, V42.7
Diabetes without chronic complication	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9	250.0–250.3, 250.8, 250.9
Diabetes with chronic complication	E10.2–E10.5, E10.7, E11.2–E11.5, E11.7, E12.2–E12.5, E12.7, E13.2–E13.5, E13.7, E14.2–E14.5, E14.7	250.4–250.7

TABLE 1. ICD-9-CM and ICD-10 Coding Algorithms for Charlson Comorbidities

Comorbidities	ICD-10	Enhanced ICD-9-CM
Hemiplegia or paraplegia	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0–G83.4, G83.9	334.1, 342.x, 343.x, 344.0–344.6, 344.9
Renal disease	I12.0, I13.1, N03.2–N03.7, N05.2–N05.7, N18.x, N19.x, N25.0, Z49.0–Z49.2, Z94.0, Z99.2	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582.x, 583.0–583.7, 585.x, 586.x, 588.0, V42.0, V45.1, V56.x
Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin	C00.x–C26.x, C30.x–C34.x, C37.x–C41.x, C43.x, C45.x–C58.x, C60.x–C76.x, C81.x–C85.x, C88.x, C90.x–C97.x	140.x–172.x, 174.x–195.8, 200.x–208.x, 238.6
Moderate or severe liver disease	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7	456.0–456.2, 572.2–572.8
Metastatic solid tumor	C77.x–C80.x	196.x–199.x
AIDS/HIV	B20.x–B22.x, B24.x	042.x–044.x

Charlson comorbidity index, CCI (2 / 4)

- 中文回顧與簡介

回顧與探討次級資料適用之共病測量方法

A Review of claims-based Comorbidity Measures

朱育增(Yu-Tseng Chu) ； 吳尚琪(Shiao-Chi Wu)

台灣公共衛生雜誌 ； 29卷1期 (2010 / 02 / 01) ， P8 - 21

繁體中文 DOI : 10.6288/TJPH2010-29-01-02 

- ICD-9-CM與
ICD-10-CM比較

- Enhanced ICD-9-CM
- ICD-10-CM

➤ Med Care. 2005 Nov;43(11):1130-9. doi: 10.1097/01.mlr.0000182534.19832.83.

Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data

Hude Quan ¹, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, William A Ghali

Affiliations + expand

PMID: 16224307 DOI: 10.1097/01.mlr.0000182534.19832.83

Charlson comorbidity index, CCI (3 / 4)

- 使用正規表達式尋找疾病診斷編碼
- Cerebrovascular disease
 - 原始描述 362.34, 430.x – 438.x
 - 去除.x 362.34, 430 – 438
 - 連接連續數字 362.34, 43[0-8]
 - 模式開頭標記 ^362.34 ^43[0-8]
 - 連接所有模式 ^36234|^43[0-8]

(pattern)

Charlson comorbidity index, CCI (4 / 4)

- 從全國的門診費用檔和住院費用檔找出研究樣本的資料
 - 只看住院日 (不含) 前過去90天的醫療紀錄
- 疾病史認定標準
 - 診斷碼在門診出現至少2次或診斷碼在住院出現至少1次
 - 任何一個診斷碼的紀錄都看 (icd9cm_1 ~ icd9cm_5)
 - 門診紀錄的權重 = 1分，住院紀錄的權重 = 2分
 - 同一類型的診斷，一天只算一次
- 加權計算總分

資訊合併

- 將整理好的眾多資料依照病人身分證號（id）進行合併
 - 研究族群：pop
 - 基本資料：idfile
 - 健保承保：nhi
 - CCI分數：cci
- 如果有遺漏值，記得要補 0 或是對應的虛擬值 !!!
 - 純粹遺漏未補的話整筆觀察值會被排除於統計分析之外

樣本納入與排除條件處理

- 排除條件
 - 基本資料不齊全者
 - 死亡日期早於住院日
 - 住院時年齡 <0 歲 或是 >100 歲

描述性統計分析：Table 1

- 以90天內是否死亡（ death_90day ）為分組
- 比較兩組的特性差異
 - 性別、年齡、投保金額類別、CCI
- 使用Standardized mean differences （ SMD ）量化差異的大小
 - 通常使用 $SMD > 0.1$ 判斷兩組之間有差異存在
 - SMD在大數據時代較不容易因為N值擴大而產生統計顯著

描述性統計分析：Table 1

Variable	Mortality result with 90 day				SMD
	Deceased		Alive		
	N = 94		N = 406		
	N	%	N	%	
Age*	77.6	12.4	48.2	33.8	1.155
Sex					
Male	54	57.5	234	57.6	0.004
Female	40	42.6	172	42.4	0.004
NHI premium range					
Dependent	36	38.3	112	27.6	0.229
15,840–29,999	40	42.6	189	46.6	0.081
30,000–44,999	13	13.8	52	12.8	0.030
Above 45,000	5	5.3	53	13.1	0.270
CCI score*	3.04	2.99	1.17	1.60	0.780
0	16	17.0	193	47.5	0.691
1–2	35	37.2	151	37.2	0.001
Above 3	43	45.7	62	15.3	0.701

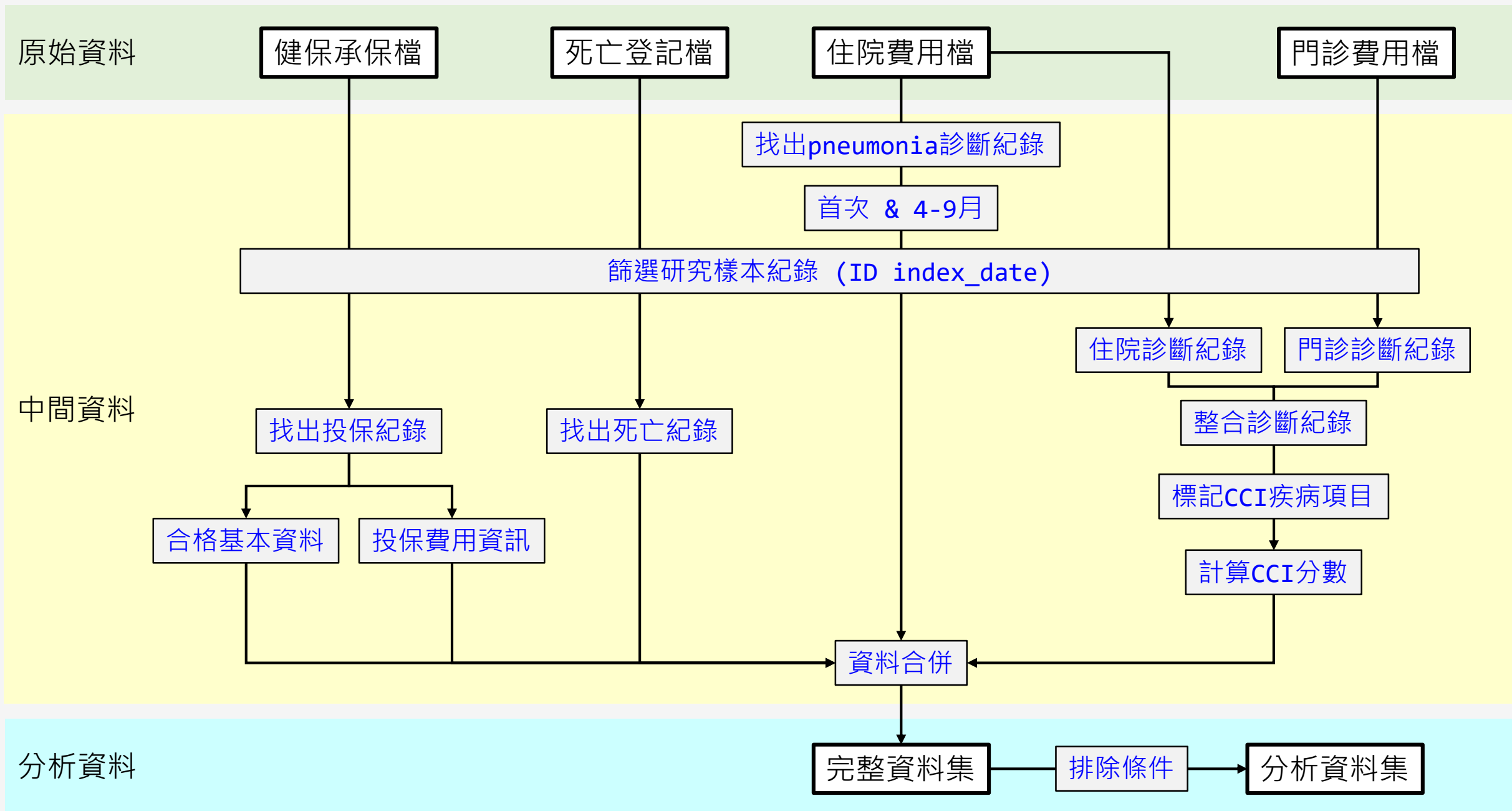
* Expressed in mean and standard deviation.

推論性統計分析：Logistic regression

- Model
 - Crude (univariable) model
 - $Y = 90\text{天內死亡}$
 - $X = \text{CCI score}$
 - Adjusted (multivariable) model
 - $Y = 90\text{天內死亡}$
 - $X = \text{CCI score} + \text{Age} + \text{Male} + \text{NHI}$
- Odds ratio & 95% confidence intervals of **CCI score**

推論性統計分析：Logistic regression

Variable	Crude model				Adjusted model			
	OR	95% CI		<i>p</i>	aOR	95% CI		<i>p</i>
CCI score	1.45	1.30	1.62	<.001	1.32	1.18	1.48	<.001
Age					1.04	1.03	1.06	<.001
Male					0.96	0.58	1.61	0.886
NHI premium range								
Dependent					1.00	(Reference)		
15,840—29,999					0.91	0.51	1.60	0.736
30,000—44,999					1.56	0.69	3.51	0.284
Above 45,000					0.59	0.19	1.78	0.346



統計軟體實際操作練習

- 檔案資料夾
 - 原始模擬資料：pathraw
 - 健保承保檔：h_nhi_enrol2014
 - 住院費用檔：h_nhi_ipdte2014
 - 門診費用檔：h_nhi_opdte2014
 - 死亡登記檔：h_ost_death2014
 - 完成整理資料：pathuse
- R程式碼以UTF-8編碼儲存，先調整RStudio設定，開啟才不會是亂碼
- 將程式碼挖空的部分填入正確的內容

Summary

- 核心理念
- 工作心流
- 技術實踐
- 知識獲取問ChatGPT
- 系統訓練找小劉老師
- 開放提問時間
- 劉品崧
- Peter Pin-Sung Liu
- psliu520@gmail.com
- <https://github.com/PSLiu/>



109年度R基礎課程-劉品崧老師

