

110年衛生福利資料科學中心

R軟體推廣課程《進階篇》

共病指數CCI的資料處理及應用

講師：劉品崧 統計分析師

花蓮慈濟醫院

課程大綱

- 前言
- CCI簡介
- 衛生福利資料庫簡介
- 計算CCI因子實務流程及R語言應用指令解說
- 討論

故事是這樣開始的

- 科主任：「A醫師，我看young V需要多寫點paper，把這個data拿去統計一下來投稿吧」



pneumonia.csv

- 肺炎(pneumonia)患者預後研究
 - 研究設計為世代追蹤研究，觀察期間為2014年1 ~ 12月
 - 納入6月的新入院的pneumonia(ICD-9-CM: 480-486)患者為研究對象
 - 從住院起始日開始follow-up到12月31日為止
 - 想知道性別對於180天mortality是否會有相關
 - 收集1~5月的病史資訊作為baseline covariates

研究設計與資料收集之時間軸

Baseline characteristics

- 年齡、性別
- 健保投保金額級距

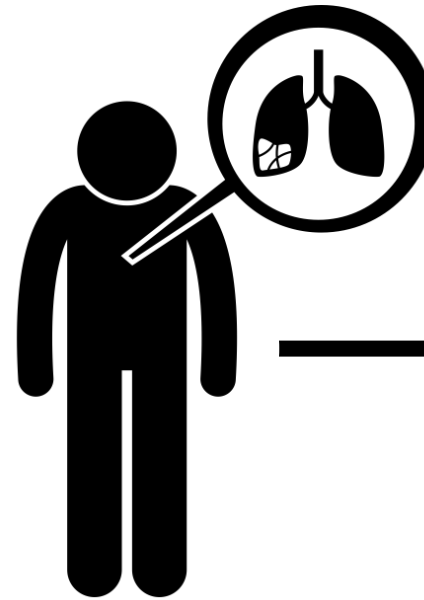
Medical history

- 高血壓、糖尿病、高血脂
- 是否曾因為肺炎住院

Follow-up outcome

- 180天內死亡
- 180天內因肺炎重複再入院

Jan, 01
2014



Dec, 31
2014

Hospital admission

- 醫學中心(Y/N)
- 診斷碼*5

研究結果

- 統計分析方法
 - Logistic regression
- 結論：在肺炎住院病患中，**男性與180天內死亡風險有較高的相關**，
 $aOR = 5.92$, $p \text{ value} = 0.0421$

Table 1. Associated factors of pneumonia 180-days mortality

Variables	aOR	95% CI		P value
Age	1.05	1.01	1.11	0.0583
Male	5.92	1.21	40.07	0.0421
NHI Premium range				
Dependent	1.00	(Ref.)		
NTD 15,840-30,000	1.30	0.25	7.30	0.7539
NTD 30,000 and above	0.76	0.09	5.58	0.7919
Medical center admission	0.74	0.12	3.95	0.7260
Disease history				
Pneumonia	14.49	3.49	84.45	0.0008
HTN	1.73	0.32	10.36	0.5282
DM	0.19	0.02	1.31	0.1195
Hyperlipidemia	4.28	0.37	46.23	0.2208

But，投稿的時候被Reviewer要求

- You should calculate the CCI score and adjust in model.



CC0簡介

- 緣起
- 版本與回顧文獻
- 今日課程參考版本

Charlson comorbidity index, CCI

- CCI是一個綜合性的共病指標
 - 遠起於預測住院病患的死亡風險
 - 包含17種疾病，各自有不同的權重
- 在回溯性病歷/資料當中，研究分組通常為非隨機，基線特質 (baseline characteristics)的差異可能同時影響分組與結果，是需要校正的因子(confounder)

Charlson comorbidity index, CCI (cont.)

- 有許多種不同的版本

Comparative Study

➤ J Chronic Dis. 1987;40(5):373-83. doi: 10.1016/0021-9681(87)90171-8.

A new method of classifying prognostic comorbidity in longitudinal studies: development and validation

M E Charlson, P Pompei, K L Ales, C R MacKenzie

被引用 38059 次

➤ Methods Inf Med. 1993 Nov;32(5):382-7.

Risk adjustment in outcome assessment: the Charlson comorbidity index

W D'Hoore¹, C Sicotte, C Tilquin

被引用 544 次

➤ J Clin Epidemiol. 1992 Jun;45(6):613-9. doi: 10.1016/0895-4356(92)90133-8.

Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases

R A Deyo¹, D C Cherkin, M A Ciol

被引用 9242 次

➤ Med Care. 1998 Jan;36(1):8-27. doi: 10.1097/00005650-199801000-00004.

Comorbidity measures for use with administrative data

A Elixhauser¹, C Steiner, D R Harris, R M Coffey

被引用 7161 次

國內有關CCI的相關回顧文獻參考

- 華藝線上圖書館  airiti Library 華藝線上圖書館

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 

共病；次級資料；行政申報資料；comorbidity；administrative data；claim data

探討共病測量方法於健保次級資料之應用

Assessing Measures of Comorbidity Using National Health Insurance Databases

朱育增(Yu-Tseng Chu)；吳尚琪(Shiao-Chi Wu)；李玉春(Yu-Chun Lee)；賴美淑(Mei-Shu Lai)；譚醒朝(Sing-Chew Tam)

台灣公共衛生雜誌；29卷3期(2010/06/01)，P191-200

繁體中文 DOI: 10.6288/TJPH2010-29-03-01 

共病；CCI；次級資料；行政申報資料；comorbidity；CCI Charlson Comorbidity Index；administrative data；claim data

今日課程參考版本

- Quan *et al*(2005).
 - ICD-9 & ICD-10
- 本課程僅作為練習範例，
不對各位的研究信效度
做保證，請選擇適合自
己研究族群的共病指標，
以及對programming過程
自行負責。

➤ [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](#)

[Full text links](#)

[Cite](#)

Abstract

Objectives: Implementation of the International Statistical Classification of Disease and Related Health Problems, 10th Revision (ICD-10) coding system presents challenges for using administrative data. Recognizing this, we conducted a multistep process to develop ICD-10 coding algorithms to define Charlson and Elixhauser comorbidities in administrative data and assess the performance of the resulting algorithms.

被引用 6605 次

CCI的因子與對應疾病診斷碼

Comorbidities	Enhanced ICD-9-CM
Myocardial infarction	410.x, 412.x
Congestive heart failure	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	093.0, 437.3, 440.x, 441.x, 443.1–443.9, 47.1, 557.1, 557.9, V43.4
Cerebrovascular disease	362.34, 430.x–438.x
Dementia	290.x, 294.1, 331.2
Chronic pulmonary disease	416.8, 416.9, 490.x–505.x, 506.4, 508.1, 508.8
Rheumatic disease	446.5, 710.0–710.4, 714.0–714.2, 714.8, 725.x
Peptic ulcer disease	531.x–534.x
Mild liver disease	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	250.0–250.3, 250.8, 250.9
Diabetes with chronic complication	250.4–250.7

Comorbidities	Enhanced ICD-9-CM
Hemiplegia or paraplegia	334.1, 342.x, 343.x, 344.0–344.6, 344.9
Renal disease	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	140.x–172.x, 174.x–195.8, 200.x–208.x, 238.6
Moderate or severe liver disease	456.0–456.2, 572.2–572.8
Metastatic solid tumor	196.x–199.x
AIDS/HIV	042.x–044.x

(Quan et al., 2005)

以現有研究設計再補上一塊重要的拼圖

Baseline characteristics

- 年齡、性別
- 健保投保金額級距

Medical history

- 高血壓、糖尿病、高血脂
- 是否曾因為肺炎住院

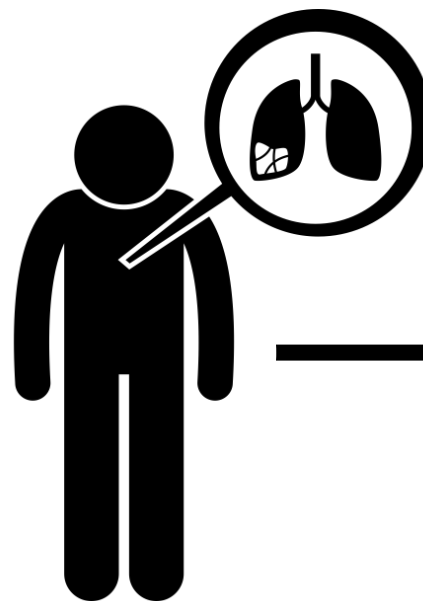
Follow-up outcome

- 180天內死亡
- 180天內因肺炎重複再入院

Jan, 01
2014

CCI

- MI、CHF、COPD ... etc



Dec, 31
2014

Hospital admission

- 醫學中心(Y/N)
- 診斷碼*5

在開始計算CCI之前
讓我們打開R
看一下資料的長相！

衛福資料庫簡介

- 概論
- 門診檔(費用/醫令)
- 住院檔(費用/醫令)

衛生福利模擬資料檔

- 虛擬10萬人資料
 - 今日使用資料庫：健保門診檔、健保承保檔
- 優點：很棒的練習材料
 - 未來有可能進入資科中心分析資料
 - 助於研究樣本數估計、決定實際資料購買策略
- 缺點：部分重要欄位無法使用/轉譯
 - func_date/in_date：就診日期，無法轉換成實際日期

門診檔費用/醫令檔

門診費用檔(CD檔/OPDTE)

id	id_s	age	func_date	func_type	prsn_id	icd9cm_1	icd9cm_2	icd9cm_3	t_dot
{%{{{/-/]}"/{&	1	62歲	21842	12	:]:*_>\$~~[*(<	43400	40200	2720	1232

譯碼簿(codebook)的使用：門診檔費用/醫令檔

全民健保處方及治療明細檔_門急診—西醫、中醫及牙醫

(H_NHI_OPDTE)

模擬資料庫使用手冊

一、檔案內容說明

檔案代號	H_NHI_OPDTE	資料筆數	1. 114,910 筆 (西醫) 2. 12,438 筆 (牙醫) 3. 14,591 筆 (中醫) (以 2014/12 為例)
中文檔名	全民健保處方及治療明細檔	英文檔名	AMBULATORY CARE EXPENDITURES BY VISITS
檔案大小	1. 30.9MB (西醫) 2. 3.43MB (牙醫) 3. 4MB (中醫) (以 2014/12 為例)	欄位數	51
屬性	月份歷史檔	週期	月
資料描述	1. 原則上一次就醫 (同一醫院同一天同一科別一筆資料) 一筆資料。 2. 內容包含病人當次就醫之疾病、醫院及費用等相關資料。 3. 資料年度包含 2014 年		
注意事項	1. 計算人次時，補報資料不應重覆計算。 2. 計算費用時，補報資料應合併。		
主鍵與比對欄位	1. 個人身分證字號、醫師身分證字號、藥師身分證字號及醫療機構代碼經加密處理。 2. 本檔與門急診醫令明細檔 (H_NHI_OPDTE) 比對之唯一鍵值為就醫代碼：FEE_YM+APPL_TYPE+HOSP_ID+APPL_DATE+CASE_TYPE+SEQ_NO。 3. 個人歸戶及個人資料比對的鍵值：ID。 4. 醫事機構歸戶與資料比對的鍵值：HOSP_ID。		

(1)檔案名稱

(2)時間間隔

(3)資料產生方式

(4)如何連結

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診

(HEALTH-04: H_NHI_OPDTE)

模擬資料庫使用手冊

一、檔案內容說明

(1)檔案名稱				
檔案代號	H_NHI_OPDTE	資料筆數	127,491 筆(西醫) 34,545 筆(牙醫) 108,607 筆(中醫) (以 2014/12 為例)	
中文檔名	全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診	英文檔名	Details of Ambulatory Care Orders	
檔案大小	18.6MB(西醫) 5.12MB(牙醫) 15.8MB(中醫) (以 2014/12 為例)	欄位數	21	
(2)時間間隔				
屬性	月份歷史檔	週期	月	譯碼簿 更新日期 2016/09/25
資料描述	1. 為一次門診多筆醫令資料。 2. 相同之 FEE_YM + APPL_TYPE + APPL_DATE + CASE_TYPE + SEQ_NO + HOSP_ID 視為同一次就醫之醫令。 3. 每一筆資料包括醫令碼、數量及點數等。			
注意事項	同一次門診之各筆醫令點數欄位(TOTAL_DOT)加總等於門診明細檔(H_NHI_OPDTE)之合計點數(T_DOT)。			
主鍵與比對欄位	1. 醫療機構代碼(HOSP_ID)經加密處理。 2. 本檔無病人身份證字號，以 FEE_YM+APPL_TYPE+APPL_DATE+CASE_TYPE+SEQ_NO+HOSP_ID 為鍵值與門診明細檔(H_NHI_OPDTE)對照。 3. 醫療機構歸戶及醫療機構資料對照的鍵值：HOSP_ID。			
(4)如何連				

(3)資料產生方式

(4)如何連結

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療明細檔 (H_NHI_OPDTE)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	身分證字號	ID	Char	12	經加密處理。國民身分證統一編號，或外籍居留證號碼（如無居留證號碼請填護照號碼），左靠不足補空白
38	性別	ID_S	Char	1	1：男性 2：女性 9：不詳
16	就醫年齡	AGE	Char	20	以就醫日期及生日計算。 「0-14天」、「15-28天」、「29天-未滿一歲」、1歲、2歲、...、84歲、85歲以上
14	就醫科別	FUNC_TYPE	Char	2	詳見欄位就醫科別(FUNC_TYPE)譯碼說明

3.就醫科別 (FUNC_TYPE)

科別	代碼
眼科	10
皮膚科	11
神經科	12
精神科	13
復健科	14

通常代碼繁多的欄位
會另外整理成附錄
請參考譯碼簿 p.13

1. 一名62歲男性(`id_s == "1"`)在某日前往醫院神經科(`func_type == "12"`)就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(`icd9cm_1 == "43400"`)，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(`total_dot`)加總，等於費用檔的總申報點數(`t_dot`)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療明細檔 (H_NHI_OPDTE)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
20	國際疾病分類號一	ICD9CM_1	Char	5	左靠不足補空白， 詳見欄位國際疾病分類號 (ICD9CM_1~ICD9CM_3)譯碼說明
21	國際疾病分類號二	ICD9CM_2	Char	5	
22	國際疾病分類號三	ICD9CM_3	Char	5	

CEREBROVASCULAR DISEASE (430-438)

434 Occlusion of cerebral arteries

The following fifth-digit subclassification is for use with category 434:

0 without mention of cerebral infarction

1 with cerebral infarction

Use additional code, if applicable, to identify status post administration of tPA (rtPA) in a different facility within the last 24 hours prior to admission to current facility (V45.88)

434.0 Cerebral thrombosis

[0-1]

Thrombosis of cerebral arteries

434.1 Cerebral embolism

[0-1]

434.9 Cerebral artery occlusion, unspecified

[0-1]

(icd9.chrisendres.com)

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDTO)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
7	醫令類別	ORDER_TYPE	Char	1	詳醫令類別(ORDER_TYPE) 譯碼說明。

2. 醫令類別(ORDER_TYPE)

醫令類別代碼	說明
0	診察費
1	用藥明細
2	診療明細
3	特殊材料
4	不得另計價之藥品、檢驗(查)或診療項目
5	EPO 注射
6	HCT 檢驗
7	代檢(95.07 刪除)
8	器官捐贈
9	藥事服務費

請參考譯碼簿 p.7

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDTO)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
8	藥品(項目)代號	DRUG_NO	Char	12	參考中央健保局網頁公告之「全民健保藥品」、「醫療費用支付標準碼」或特殊材料代碼。

醫療服務給付項目及支付標準網路查詢服務

資料更新日期：1100701 共2856817人



衛生福利部中央健康保險署

查詢條件

中文項目名稱：	<input type="text" value="請輸入中文項目名稱"/>
英文項目名稱：	<input type="text" value="請輸入英文項目名稱"/>
診療項目代碼：	<input type="text" value="05209A"/>
支付價生效日期：	<input type="text" value="▼"/> <input type="text" value="▼"/> 月



項目代碼	診療項目	支付點數	參考起迄日
05209A	門診藥事服務費 - 慢性病處方給藥二十八天以上	83	110.03.01 ~ 迄今

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDТО)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
8	藥品(項目)代號	DRUG_NO	Char	12	參考中央健保局網頁公告之「全民健保藥品」、「醫療費用支付標準碼」或特殊材料代碼。

健保用藥品項查詢

藥品名稱(中文)：	<input type="text"/>
藥品名稱(英文)：	<input type="text"/>
藥品代碼：	<input type="text" value="AC43862100"/>
劑型：	<input type="text"/>

藥品代碼	藥品名稱 (英文)	藥品名稱 (中文)	成分
AC43862100	COFARIN TAB 1MG "GENTLE"	“政德”可化凝錠 1 毫克	WARFARIN SODIUM

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

住院費用檔(DD檔/IPDTE)

住院費用/醫令檔

id	id_s	in_age	in_date	func_type	icd9cm_1	icd_op_code1	e_bed_day	s_bed_day
_%~/)%=!^>\	2	53歲	55998	BD	5400	4700	2	0

hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
2852371	201401	6250	2	5	688

住院醫令檔(DO檔/IPDTO)

order_type	order_code	order_dot	rate_type
1	AB29836277	25	1
1	AC46544210	333	1
1	...		
2	96020C	6659	1.7
2	74004B	18681	1.83
2	...		
3	SCV03625LL2C	2940	1.05
3	CDD0530502P2	285	1.05
3	...		

1. 一名53歲女性(`id_s == "2"`)
在某日因腹痛前往醫院家就診，經外科醫師
(`func_type == "BD"`)診斷
為闌尾炎，收住院開刀治療。
2. 手術當中施行麻醉(`96020C`)
並切除了病患闌尾(`74004B`)。
3. 住院日總共為2天
(`e_bed_day + s_bed_day = 2`)，期間有施予抗生素
(`AC46544210`)。

實際衛福資料庫疾病診斷碼版本：以中風為例

• ICD-9-CM，2015 (含)以前



Home > 2015 ICD-9-CM Diagnosis Codes > Diseases Of The Circulatory System 390-459 >

Cerebrovascular Disease 430-438 >

- [430](#) Subarachnoid hemorrhage
- [431](#) Intracerebral hemorrhage
- [432](#) Other and unspecified intracranial hemorrhage
- [433](#) Occlusion and stenosis of precerebral arteries
- [434](#) Occlusion of cerebral arteries
- [435](#) Transient cerebral ischemia
- [436](#) Acute, but ill-defined, cerebrovascular disease
- [437](#) Other and ill-defined cerebrovascular disease
- [438](#) Late effects of cerebrovascular disease

• ICD-10-CM，2016(含)以後



ICD-10-CM Codes > I00-I99 Cerebrovascular diseases > Cerebrovascular diseases I60-I69

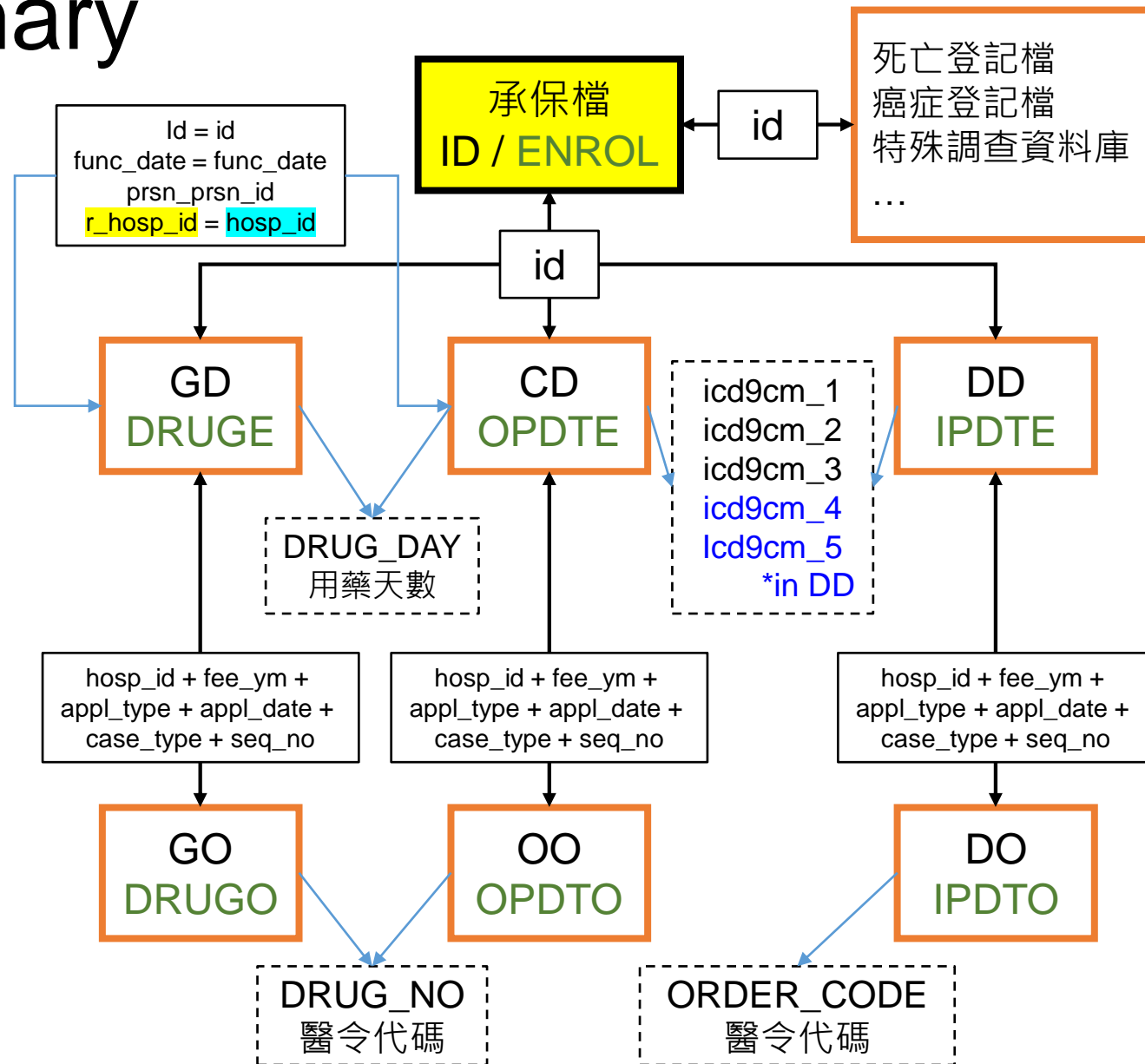
Cerebrovascular diseases I60-I69

Codes

- [I60](#) Nontraumatic subarachnoid hemorrhage
- [I61](#) Nontraumatic intracerebral hemorrhage
- [I62](#) Other and unspecified nontraumatic intracranial hemorrhage
- [I63](#) Cerebral infarction
- [I65](#) Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
- [I66](#) Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
- [I67](#) Other cerebrovascular diseases
- [I68](#) Cerebrovascular disorders in diseases classified elsewhere
- [I69](#) Sequelae of cerebrovascular disease

衛生福利資料庫summary

- 本次使用檔案 (現今檔名/國衛院舊稱)
 - 門診費用檔(OPDTE / CD) , 一月三檔
 - 門診醫令檔(OPDTO / OO) , 一月三檔
 - 檔名結尾：西醫10、牙醫20、中醫30
 - 住院費用檔(IPDTE / DD) , 一年一檔
 - 住院醫令檔(IPDTO / DO) , 一月一檔
 - 健保承保檔(ENROL / ID) , 一月一檔
 - 死因統計檔(DEATH / DR) , 一年一檔
- 資料庫手冊
 - 確認你要的資料需要使用什麼欄位
 - 確認跨資料庫之間要怎麼連接
 - 身分證號、時間、空間、串聯欄位
 - 瞭解資料庫的能與不能



計算CCI因子實務流程 及R語言應用指令解說

- 流程
- 應用指令

資料處理

統計分析

迴圈程式

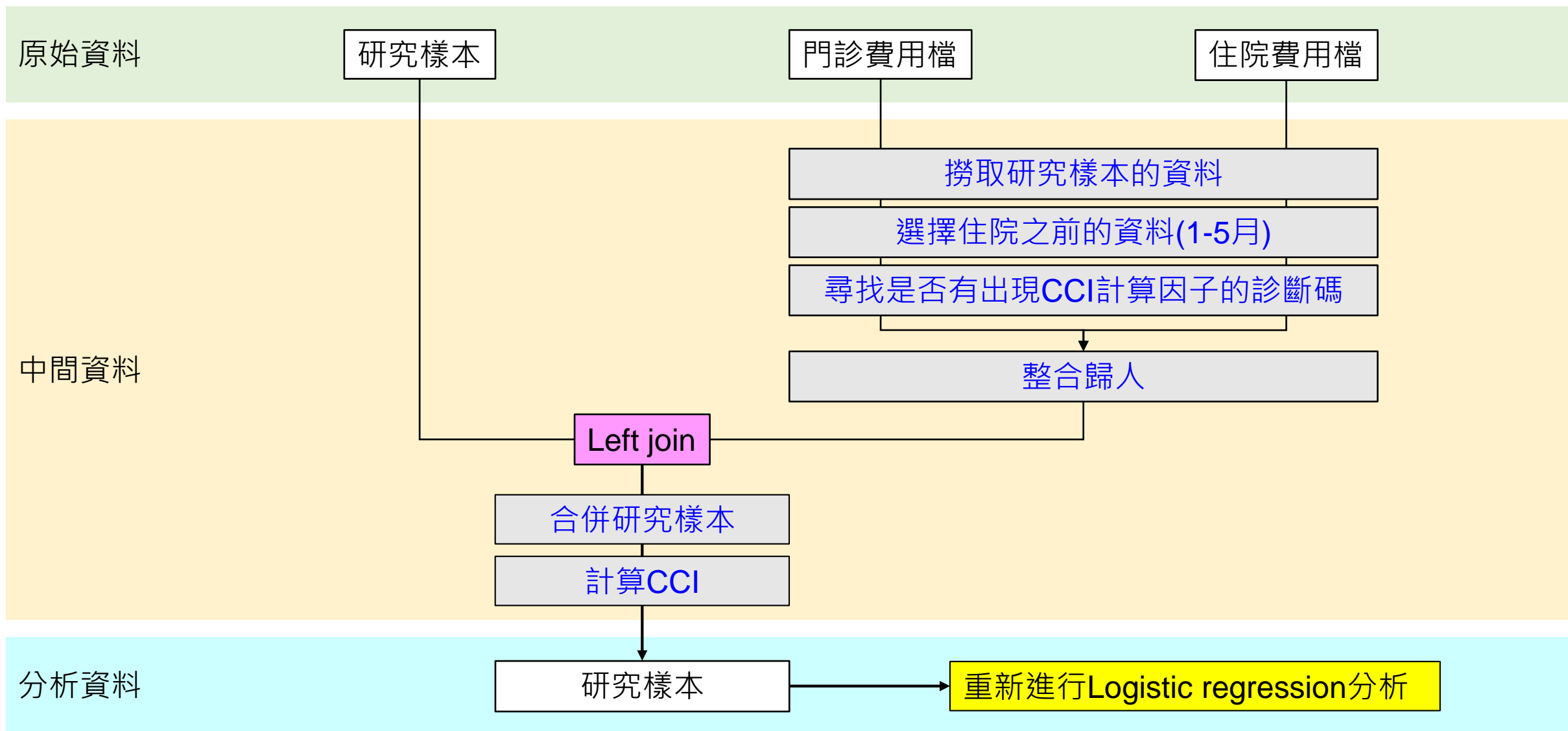
資料

處理

合併

分析

工作心流



資料處理：fread {data.table}

- 讀取.csv檔案
- `fread("h_nhi_ipdte103.csv", colClasses = "character")`



需要讀取的檔名



讀取進來後都先轉為文字形式

資料處理：dummy_cols {fastDummies}

- 快速製作虛擬變項(dummies)
 - 將類別欄位的屬性，都獨立成一個variable去表達樣本的屬於哪一類
- `dummy_cols(dt, select_columns = c("sex", "incgp"))`

↑
要製作虛擬變項的資料物件

↑
要製作虛擬變項的欄位名稱

ID	sex	sex_1	sex_2
X07	1	1	0
Y56	2	0	1

資料處理：melt {data.table}

- 橫向資料轉置為直向資料
- melt(

id	func_date	icd9cm_1	icd9cm_2	icd9cm_3
G56	2021/8/19	434	402	272

id	func_date	variable	icd
G56	2021/8/19	icd9cm_1	434
G56	2021/8/19	icd9cm_2	402
G56	2021/8/19	icd9cm_3	272

資料

data = temp,

軸心

id.vars = c("id", "func_date"),

變數

measure.vars = c("icd9cm_1", "icd9cm_2", "icd9cm_3"),

輸出

value.name = "icd"

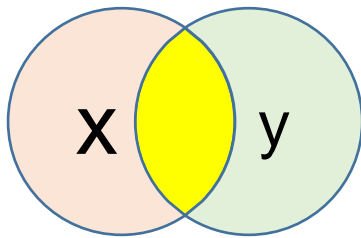
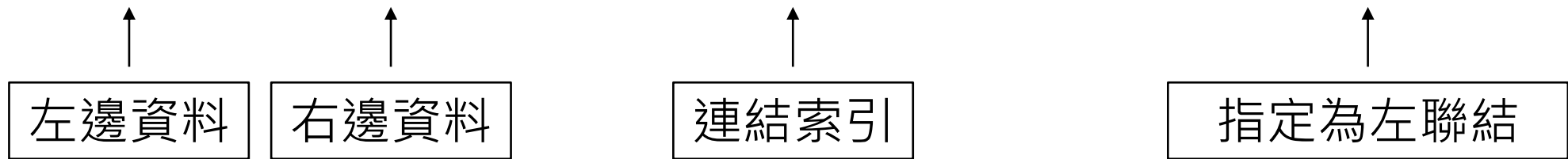
)

資料處理：grep1 {base}

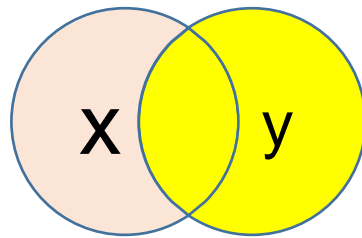
- 搜尋文字
- `temp[grepl(pattern = "^41[0-2]", x = icd), mi := 1]`
- pattern內是正規表達式
 - `diagnosis <- "4102"`
 - `substr(diagnosis, 1, 3) %in% c("410", "411", "412")`
 - `grepl("^410|^411|^412", diagnosis)`
 - `grepl("^41[0-2]", diagnosis)`，文字為41開頭且後面接0-2的數字

資料處理：merge {data.table}

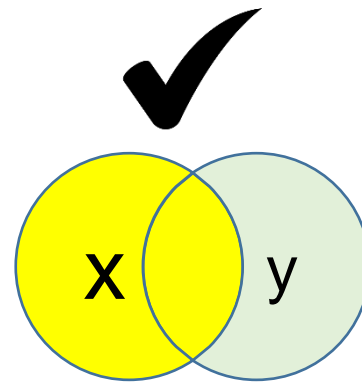
- 資料水平合併，此範例為左聯結
- `merge(x = dt, y = dt_cci, by = c("id"), all.x = TRUE)`



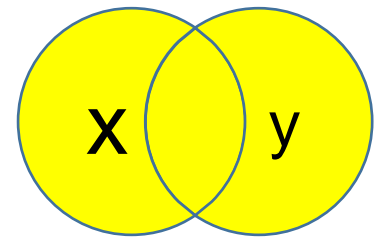
交集



右聯結



左聯結



聯集

資料處理：data.table {data.table}

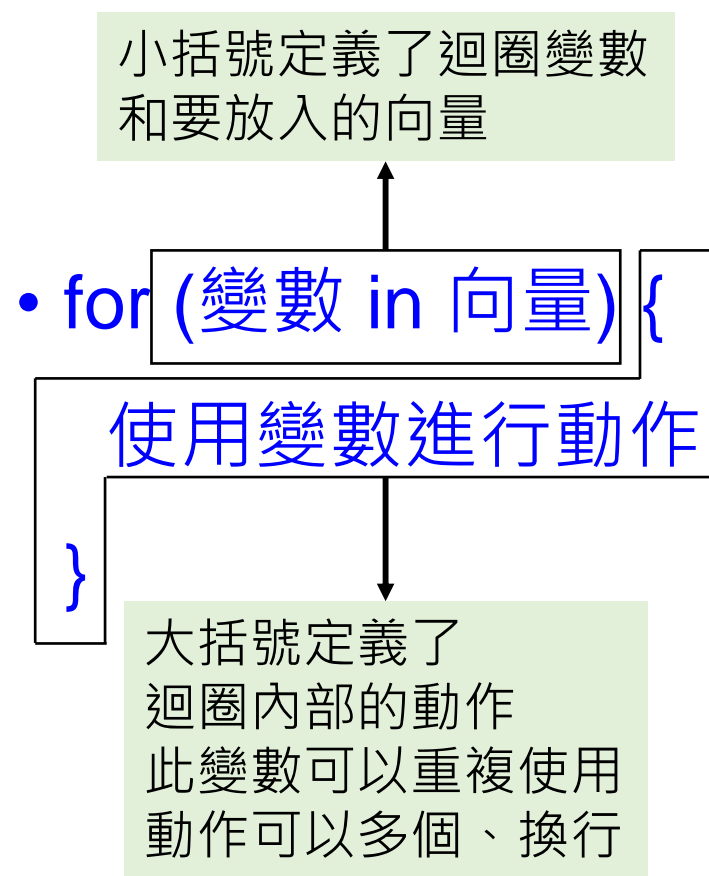
- 在次資料(sub-data.table)當中將許多變數做統一的處理
- dt_cci_opd[,
 lapply(.SD, sum, na.rm = T),
 .SDcols = c("mi", "chf", "pvd" ...),
 by = .(id)]
- 依據ID區分sub-data.table，將mi, chf等變數進行加總(sum)

統計分析：glm {stats}

- 一般化線性模型
- ```
model_2 <- glm(
 formula = death ~ age + sex_1 + ... ,
 data = dt_2,
 family = "binomial")
```
- 迴歸模型公式： $\text{outcome} \sim \text{predictor}_1 + \text{predictor}_2 \dots$
- Family會依據outcome類型不同而調整
  - gaussian、binomial, poisson、Gamma etc.

# 迴圈程式：for迴圈

- 基本架構



- 舉例

```
• for (i in a1) {
 print(i)
}
```

- a1向量元素會逐一放入迴圈中
- 成為變數 i 在迴圈內部操作
- 就這麼簡單～

# 迴圈搭配清單使用

- 清單A：想撈取的檔名，長度為5(1 ~ 5月共5個檔案)
- 清單B：處理過檔案的儲存容器，長度同清單A

| 清單A             |               |                                                                                             | 清單B                 |
|-----------------|---------------|---------------------------------------------------------------------------------------------|---------------------|
| <u>Variable</u> | <u>Vector</u> | <u>Main function</u>                                                                        | <u>Store output</u> |
| i = 1           | cd10301.csv   | <pre># 讀取資料 temp &lt;- fread(input[[i]])  # 經過一番中間處理  # 輸出物件 output[[i]] &lt;- result</pre> | ...                 |
| i = 2           | cd10302.csv   |                                                                                             | ...                 |
| i = 3           | cd10303.csv   |                                                                                             | ...                 |
| i = 4           | cd10302.csv   |                                                                                             | ...                 |
| i = 5           | cd10303.csv   |                                                                                             | ...                 |



```
for (i in 1:length(list_a)) {
 temp <- fread(list_a[[i]])
 result <- my_process(temp)
 list_b[[i]] <- result
}
```

# 討論

- 修正後的結果
- Programming實務考量

# 修正後的研究結果

- 統計分析方法
  - Logistic regression
- 結論：在肺炎住院病患中，加入CCI作為adjust confounder，結果發現男性與180天內死亡風險有較高的相關，aOR = 5.39, p value = 0.0446

Table 1. Associated factors of pneumonia 180-days mortality

| Variables                | aOR   | 95% CI |       | P value |
|--------------------------|-------|--------|-------|---------|
| Age                      | 1.04  | 0.99   | 1.10  | 0.1360  |
| Male                     | 5.39  | 1.14   | 34.56 | 0.0479  |
| NHI Premium range        |       |        |       |         |
| Dependent                | 1.00  | (Ref.) |       |         |
| NTD 15,840-30,000        | 1.17  | 0.21   | 6.66  | 0.8584  |
| NTD 30,000 and above     | 0.61  | 0.06   | 4.88  | 0.6558  |
| Medical center admission | 0.82  | 0.13   | 4.54  | 0.8255  |
| CCI                      | 1.28  | 0.90   | 1.87  | 0.1770  |
| Disease history          |       |        |       |         |
| Pneumonia                | 13.29 | 3.08   | 81.37 | 0.0015  |
| HTN                      | 2.08  | 0.37   | 14.03 | 0.4185  |
| DM                       | 0.16  | 0.01   | 1.19  | 0.1016  |
| Hyperlipidemia           | 3.25  | 0.27   | 37.25 | 0.3368  |

# Programming實務考量(1)：診斷碼轉置？

- 如果**不轉置**時的資料結構

| id  | func_date | icd_1 | icd_2 | icd_3 |
|-----|-----------|-------|-------|-------|
| G56 | 2021/8/19 | 434   | 402   | 272   |

- `dt[grepl("^41[0-2]", icd_1)|grepl("^41[0-2]", icd_2)|grepl("^41[0-2]", icd_3)]`

- 如果**有轉置**時的資料結構

- `dt[grepl("^41[0-2]", icd)]`

| id  | func_date | variable | icd |
|-----|-----------|----------|-----|
| G56 | 2021/8/19 | icd9cm_1 | 434 |
| G56 | 2021/8/19 | icd9cm_2 | 402 |
| G56 | 2021/8/19 | icd9cm_3 | 272 |

- 如果需要處理的資料不只有診斷碼，需要重新架構
  - 申報類別、日期、醫師ID etc.



# Programming實務考量(2)：去重複 & 歸人？

- 去重複(unique)
  - 避免重覆計算造成統計錯誤
  - 會低估較果，傾向null
  - 但此時結果仍有顯著的話，真實效果理論上會更強
- 歸人
  - 統整個人在一段觀察時間內的資訊
  - 配合統計軟體的需要去整理data
    - 若是repeated-measurement類型，亦有wide或long-format格式需要特別處理

# 課程討論 & Final remark

- Summary

- 研究設計決定資料處理

不要盲目撈data打code

- 找尋適合自己研究族群的  
comorbidity index

- 劉品崧

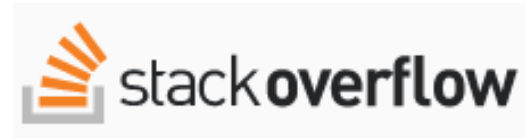
- **Email** : psliu520@gmail.com
  - **PubMed** : Peter Pin-Sung Liu

- See more

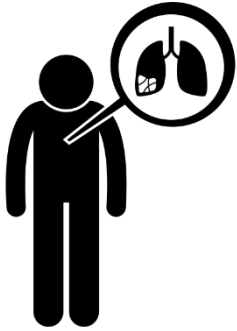
- Stack Overflow

- STHDA

- Build by A. Kassambara  
creator of survminer



# 使用圖片版權來源



Created by Gan Khoon Lay  
from Noun Project



Created by iconcheese  
from Noun Project



Created by Adrien Coquet  
from Noun Project