

113年度衛生福利資料科學中心統計軟體推廣課程

2024.05.03 (五) 09:00-16:00 @ 國立成功大學 75201電腦教室

# 健康資料研究與R軟體

## 《衛生福利資料基礎班》

劉品崧 統計諮詢暨分析師

花蓮慈濟醫院健康長壽中心

# 課程大綱

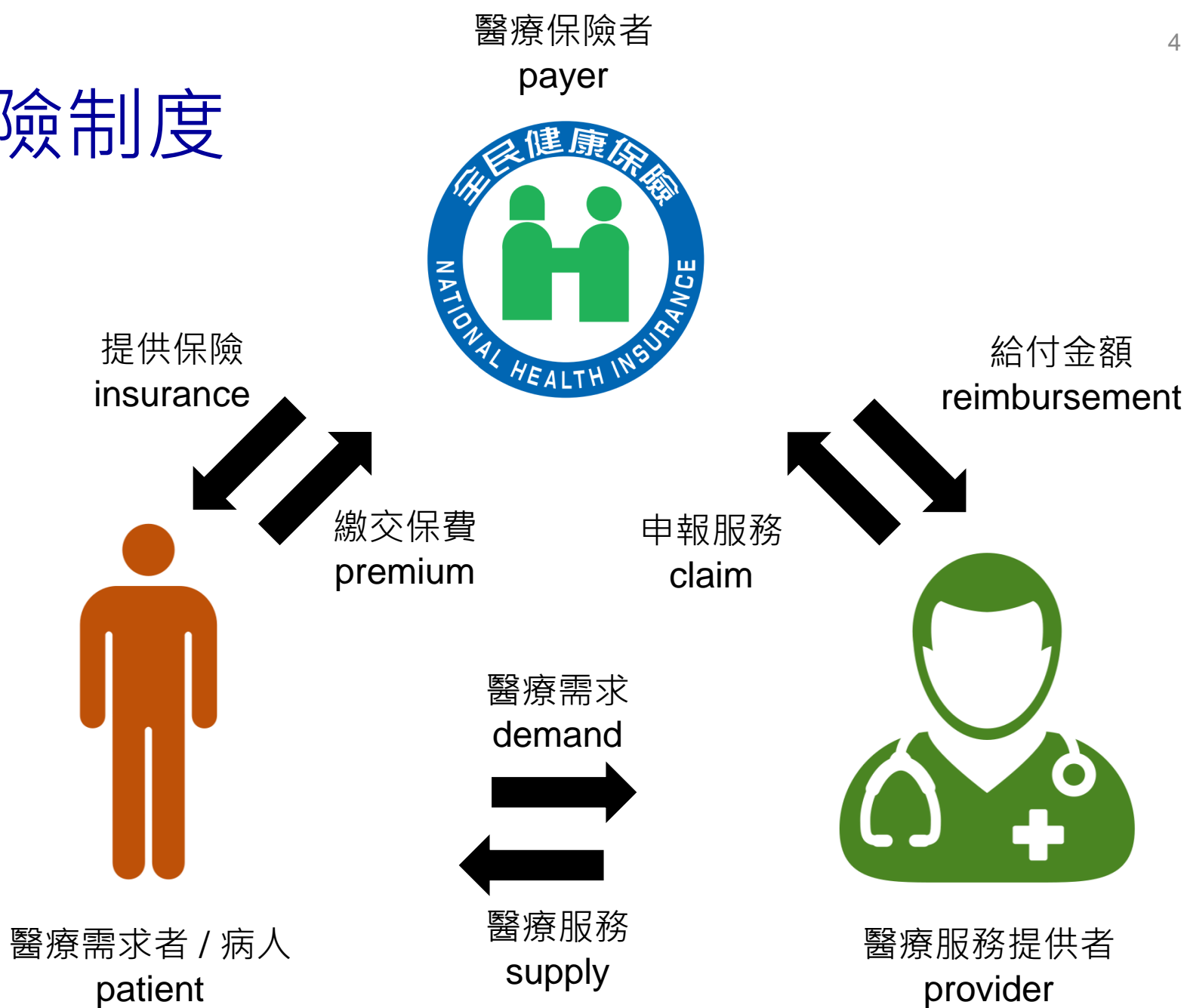
- 衛生福利資料庫
  - 資料來源
  - 管理模式
  - 資料樣態
  - 參考工具
  - 關聯架構
- 流行病學調查研究實作
  - 研究命題、研究團隊、執行計畫
  - 核心理念、工作流程、技術實踐

# 衛生福利資料庫

- 資料來源
- 管理模式
- 資料樣態
- 參考工具
- 關聯架構

# 台灣全民健康保險制度

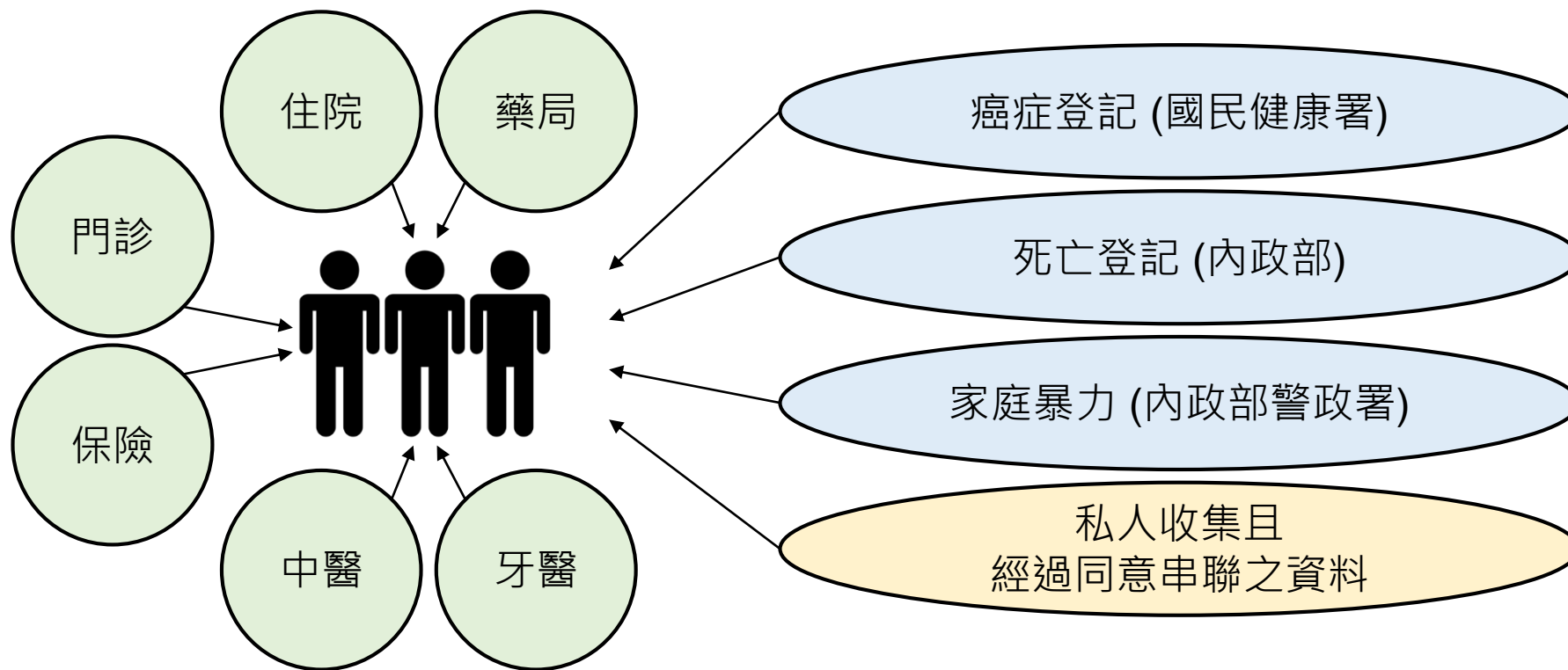
- 臺灣全民健康保險制度
  - Since 1995年3月1日
  - 99.9%人口強制納入健保
- 單一保險人給付制度
  - 西醫門診、住院、藥局
  - 中醫、牙醫
  - 預防保健、癌症篩檢



# 全民健康保險資料庫 → 衛生福利資料庫

國家衛生研究院 (NHRI) 健保資料庫工作小組  
資料年份：1995-2013  
樣本規格：100萬人口抽樣檔  
資料範圍：以健保申報資料為主

衛生福利部 (MOHW) 資料科學中心  
資料年份：1995-2022 (每年update，通常delay 2年)  
樣本規格：全人口資料檔、200萬人口抽樣檔、特殊抽樣檔  
資料範圍：公務機關之釋出資料或私人機構攜入資料









# 衛生福利部 統計處 衛生福利資料科學中心







- 資料使用環境：獨立作業區
  - 通過ISO27001資訊安全國際驗證（維護隱私與資安）
  - 不能攜帶手機、不能google、不能喝咖啡、不能討論
  - 數量有限搶先預約，僅開放AM 09 – PM 06




# 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


- 資料檔案
  - 資料夾：base
- 中文名稱及譯碼簿
  - 資料夾：codebook


 h\_nhi\_opdte2014.csv


 h\_nhi\_ipdte2014.csv


 h\_nhi\_druge2014.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



# 譯碼手冊：衛生福利部統計處

- 衛生福利部統計處「衛生福利資料科學中心專區」首頁
  - <https://dep.mohw.gov.tw/DOS/cp-5119-59201-113.html>
- 資料庫介紹
  - <https://dep.mohw.gov.tw/DOS/lp-2506-113.html>
- 全人口 / 200萬歸人檔
  - [https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS\\_dc002.html](https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS_dc002.html)
- 模擬資料庫
  - [https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS\\_dc009.html](https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS_dc009.html)

# 疾病診斷：ICD-9-CM / ICD-10-CM

- 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

<http://www.icd9data.com/2015/Volume1/default.htm>

- 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

<https://www.icd10data.com/ICD10CM/Codes>

# 藥物分類：健保用藥品項網路查詢服務

資料更新時間：113/03/27

健保目前給付 ☐ 只顯示給付中資料 ☒ 顯示歷次異動詳細資料

藥品代碼

AC57185255

務必勾選 ( 藥品更替、進退市場 )

藥品名稱(中英文)

請輸入部分中英文名稱

成分名稱

請輸入部分成分名稱

每頁顯示筆數

10筆



▼ 進階查詢

查詢

清除

藥品代碼

AC57185255

藥品名稱(英文)

Cravit IV Solution for Infusion 5mg/ml

藥品名稱(中文)

"台灣第一三共"可樂必妥靜脈輸液5毫克/毫升

成分名稱 / 成分含量

LEVOFLOXACIN (HEMIHYDRATE) 5 MG/ML

規格量

100 ML

單複方

單方

價格

1205.00

起迄

101.08.01-103.04.30

藥商

台灣第一三共股份有限公司

劑型

注射劑

藥品分類

研發廠

分類分組名稱

LEVOFLOXACIN, 注射劑, 500.00 MG

ATC代碼

J01MA12

一種成份通常有多種藥品

# 藥物分類：ATC code → 健保用藥品項查詢

- ATC code ( WHO系統 )

- 確認有無相近分類藥品

A ALIMENTARY TRACT AND METABOLISM

B BLOOD AND BLOOD FORMING ORGANS

C CARDIOVASCULAR SYSTEM

D DERMATOLOGICALS

G GENITO URINARY SYSTEM AND SEX HORMONES

H SYSTEMIC HORMONAL PREPARATIONS, EXCL. SEX

J ANTIINFECTIVES FOR SYSTEMIC USE

J01 ANTIBACTERIALS FOR SYSTEMIC USE

J01M QUINOLONE ANTIBACTERIALS

J01MA Fluoroquinolones

J01MA12 levofloxacin

- 若想擴展到J01開頭的全身性抗生素

- 健保用藥品項網路查詢服務

- 注意：此查詢方式為“包含”關鍵字都算

健保目前給付 ☐ 只顯示給付中資料 ☒ 顯示歷次異動詳細資料

ATC代碼

J01

- 健保用藥品項查詢項目檔

- 政府資料開放平台
- `R > dt[grep1("^J01", atc_code)]`
- ATC code為J01開頭的健保藥品代碼
- 藥品代碼比對醫令檔

# 全民健保醫療服務給付項目及支付標準網路查詢

## 支付標準查詢

診療項目中文

可輸入部分診療項目中文

診療項目英文

可輸入部分診療項目英文

診療項目代碼

64197C

支付價生效年

請選擇

支付價生效月

請選擇

資料範圍

全部 ( 含歷次異動 )

診療項目代碼	診療項目中文	支付點數	支付價生效起迄
64197C	肌腱切開或筋膜切開	3240	091.05.01 ~ 093.06.30
64197C	肌腱切開或筋膜切開	5038	093.07.01 ~ 108.12.31
64197C	肌腱切開或筋膜切開	6046	109.01.01 ~ 112.09.30
64197C	肌腱切開或筋膜切開	6046	112.10.01 ~ 999.12.31

務必勾選 ( 醫令更替、點值異動 )

# 健保承保檔 H\_NHI\_ENROL

- 某台灣國民身分證號S00053504
- 生於1942年 ( id\_birth\_y = 1942 ) 女性 ( id\_s = 2 ) ，今年72歲
- 以農民身分 ( id1\_unit = 31Q ) 在南投市 ( id1\_city = 3802 ) 投保
- 1月份 ( prem\_ym = 201401 ) 投保金額級距為21900元 ( id1\_amt = 21900 )
- 12月份 ( prem\_ym = 201412 ) 投保金額級距為22800元 ( id1\_amt = 22800 )

健保承保檔  
H\_NHI\_ENROL

個人身分證號	性別	出生年	保費計費年月	投保金額	被保險人身分證字號	投保單位屬性	投保單位地區代號
id	id_s	id_birth_y	prem_ym	id1_amt	id1	id1_unit	id1_city
S00053504	2	1942	201401	21900	S00053504	31Q	3801
S00053504	2	1942	201402	21900	S00053504	31Q	3801
...	...	...	...	...	...	...	...
S00053504	2	1942	201410	22800	S00053504	31Q	3801
S00053504	2	1942	201411	22800	S00053504	31Q	3801

# 健保承保檔 H\_NHI\_ENROL 譯碼簿使用方式

全民健保承保檔

(HEALTH-07：H\_NHI\_ENROL)

模擬資料庫使用手冊

一、 檔案內容說明

- (1)檔案名稱
- (2)時間間隔
- (3)資料產生方式
- (4)如何連結

檔案代號	H_NHI_ENROL	資料筆數	97,229 筆 (以 2014/12 為例)		
中文檔名	全民健保承保檔	英文檔名	Registry for Beneficiaries		
檔案大小	7.56MB (以 2014/12 為例)	欄位數	19		
屬性	現況檔	週期	年：99 年以前 月：100 年以後	譯碼簿 更新日期	2016/09/25
資料描述	包含保險對象投保類別、投保單位、投保金額等相關資料				
注意事項	<div>1. 不含在軍方單位投保之保險對象。</div> <div>2. 包含在保及停保記錄。</div> <div>3. 轉出或退保者，僅提供最後一筆異動紀錄。</div> <div>4. 無基本資料之被保險人及其眷屬投保記錄，視為異常記錄不提供。</div> <div>5. 若個人身份證字號(ID)與被保險人身份證字號(ID1)相同，則該筆資料為被保險人。若個人身份證字號(ID)不同於被保險人身份證字號(ID1)，則該筆資料為眷屬。</div> <div>6. 眷屬之身分別(ID1_TYPE)、身分屬性(ID1_IDENT)及投保金額(ID1_AMT)為空值，均以其依附之被保險人身份證字號為主。</div>				
主 鍵 與 比 對欄位	<div>1. 個人身分證字號(ID)與被保險人身份證字號(ID1)經加密處理。</div> <div>2. 個人歸戶及個人資料對照的鍵值：ID。</div>				

# 健保承保檔 H\_NHI\_ENROL 譯碼簿使用方式

- 生於1942年 ( id\_birth\_y = 1942 ) 女性 ( id\_s = 2 ) ,

## 二、 檔案格式及資料描述

全民健保承保檔 (H_NHI_ENROL)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
6	個人身分證字號性別	ID_S	Char	1	1：男性 2：女性 9：不詳。
9	出生年	ID_BIRTH_Y	Char	6	YYYY。



# 健保承保檔 H\_NHI\_ENROL 譯碼簿使用方式

- 以農民身分 ( id1\_unit = 31Q ) 投保

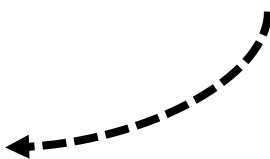
## 二、 檔案格式及資料描述

全民健保承保檔 (H_NHI_ENROL)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
4	單位屬性	ID1_UNIT	Char	3	89 年及 98 年無此欄位，詳單位屬性(ID1_UNIT)說明。

## 三、 欄位 (變項) 譯碼

### 1.單位屬性(ID1\_UNIT)

通常代碼繁多的欄位  
會另外整理成附錄  
請參考譯碼簿 p.4



保險類目	屬性代碼	被保險人
31	Q	農民

# 門診費用檔 H\_NHI\_OPDTE & 門診醫令檔H\_NHI\_OPDTO

- 今年五月 ( func\_date = 2014-05-07 ) 前往門診就醫
- 主診斷為糖尿病 ( icd9cm\_1 = 250 )
- 醫師開立藥物進行治療 ( drug\_no = A048560100 )
- 本次連續處方在門診先領了28天的藥物 ( drug\_day = 28 )

門診費用檔  
H\_NHI\_OPDTE

身分證字號	就醫日期	醫師身分證字號	國際疾病分類號一	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
id	func_date	prsn_id	icd9cm_1	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
S00053504	2014-05-07	S00221655	250	002405017	201405	00004972	1	04	29

檔案串聯用比對欄位

門診醫令檔  
H\_NHI\_OPDTO

藥品代碼	用藥天數	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
drug_no	drug_day	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
A048560100	28	002405017	201405	00004972	1	04	29

# 藥局費用檔 H\_NHI\_DRUGE & 藥局醫令檔H\_NHI\_DRUGO

- 持處方簽領了後續兩次的藥物 ( fee\_ym = 201405、201406 )

門診費用檔  
H\_NHI\_OPDTE

身分證字號	就醫日期	醫師身分證字號	醫事機構代號
id	func_date	prsn_id	hosp_id
S00053504	2014-05-07	S00221655	002405017

與原處方比對欄位

藥局費用檔  
H\_NHI\_DRUGE

身分證字號	處方日期	原處方醫師身分證字號	原處方醫事機構代號	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
id	func_date	prsn_id	r_hosp_id	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
S00053504	2014-05-07	S00221655	002405017	003533554	201406	00006337	1	2	615
S00053504	2014-05-07	S00221655	002405017	003533554	201407	00045609	1	2	777

檔案串聯用比對欄位

藥局醫令檔  
H\_NHI\_DRUGO

藥品代碼	用藥天數	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
drug_no	drug_day	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
A048560100	28	003533554	201406	00006337	1	2	615
A048560100	28	003533554	201407	00045609	1	2	777

# 住院費用檔 H\_NHI\_IPDTE & 住院醫令檔H\_NHI\_IPDTO

- 在10月 ( func\_date = 2014-10-02 ) 住院17天 ( e\_bed\_day = 17 )
- 主診斷為敗血症 ( icd9cm\_1 = 038.0 ) ，進行治療 ( order\_code )

住院費用檔  
H\_NHI\_IPDTE

身分證字號	就醫日期	國際疾病分類號一	急性病床天數	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
id	in_date	icd9cm_1	e_bed_day	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
S00053504	2014-10-02	0380	17	000056618	201411	00055219	1	1	403

檔案串聯用比對欄位

住院醫令檔  
H\_NHI\_IPDTO

醫令代碼	醫事機構代號	費用年月	申報日期	申報類別	案件分類	流水號
order_code	hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
病床費	000056618	201411	00055219	1	1	403
護理費	000056618	201411	00055219	1	1	403
電腦斷層	000056618	201411	00055219	1	1	403
筋膜切開	000056618	201411	00055219	1	1	403
抗生素	000056618	201411	00055219	1	1	403
速效胰島素	000056618	201411	00055219	1	1	403
長效胰島素	000056618	201411	00055219	1	1	403

# 死因統計檔 H\_NHI\_DEATH

- 不幸地在11月過世 ( `death_date = 2014-11-29` )

死亡登記檔

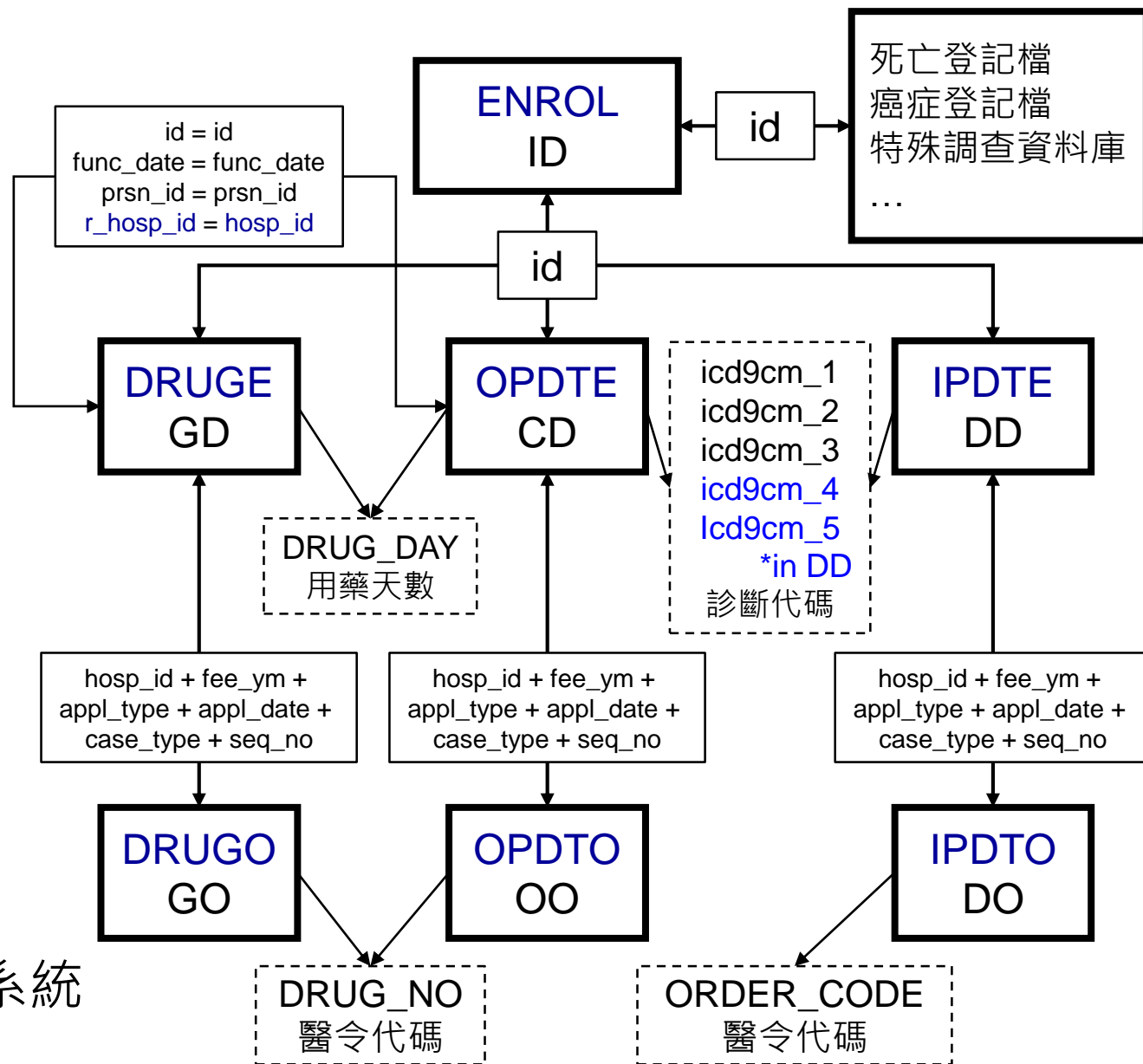
H\_OST\_DEATH

個人身分證號      死亡日期      死因分類

id	death_date	icd10
S00053504	2014-11-29	K746

# 跨資料庫關聯架構

- 本次使用檔案 ( 現今檔名 / 舊稱 )
  - 門診費用檔 ( OPDTE / CD )
  - 門診醫令檔 ( OPDTE / OO )
  - 住院費用檔 ( IPDTE / DD )
  - 住院醫令檔 ( IPDTE / DO )
  - 藥局費用檔 ( DRUGE / GD )
  - 藥局醫令檔 ( DRUGE / GO )
  - 健保承保檔 ( ENROL / ID )
  - 死因統計檔 ( DEATH / DR )
- 譯碼手冊、ICD、ATC、NHI編碼系統



# 流行病學調查研究實作

- 研究命題
- 研究團隊
- 執行計畫
- 核心理念
- 工作流程
- 技術實踐

# 臺灣高齡族群的糖尿病盛行率調查

- 研究問題

- 台灣的高齡族群當中，糖尿病的盛行率是多少？

- 預期答案

- 台灣的高齡族群當中，糖尿病的盛行率是\_\_\_\_%？



# 健康資料研究的黃金三角

- 領域專家或從業人員 → 計畫主持人 PI

- 研究問題、操作定義
- 發表產出、競爭資源

- 研究設計與統計分析

- 諮詢目的、討論邏輯
- 設計流程、執行分析

- 程式編碼及資料管理

- 符合目的、回答問題
- 調配時間、管理空間

在許多團隊當中  
通常會一起  
由統計分析師負責

# 健康資料研究的歷程

階段	PI	統計分析師
1. 議題確認	積極發想	客觀評論
2. 文獻回顧	相同：參考定義 不同：攻略方向	確認材料 可否應用
3. 研究執行計畫書	將定義、流程、共識完整文件化	
4. 執行研究計畫	親自確認	親手執行
5. 彙整研究結果	撰寫論文	編彙圖表 提供資訊
6. 投稿/發表/專利	依需求整理、修改、延伸	

# 健康資料研究的執行計畫書

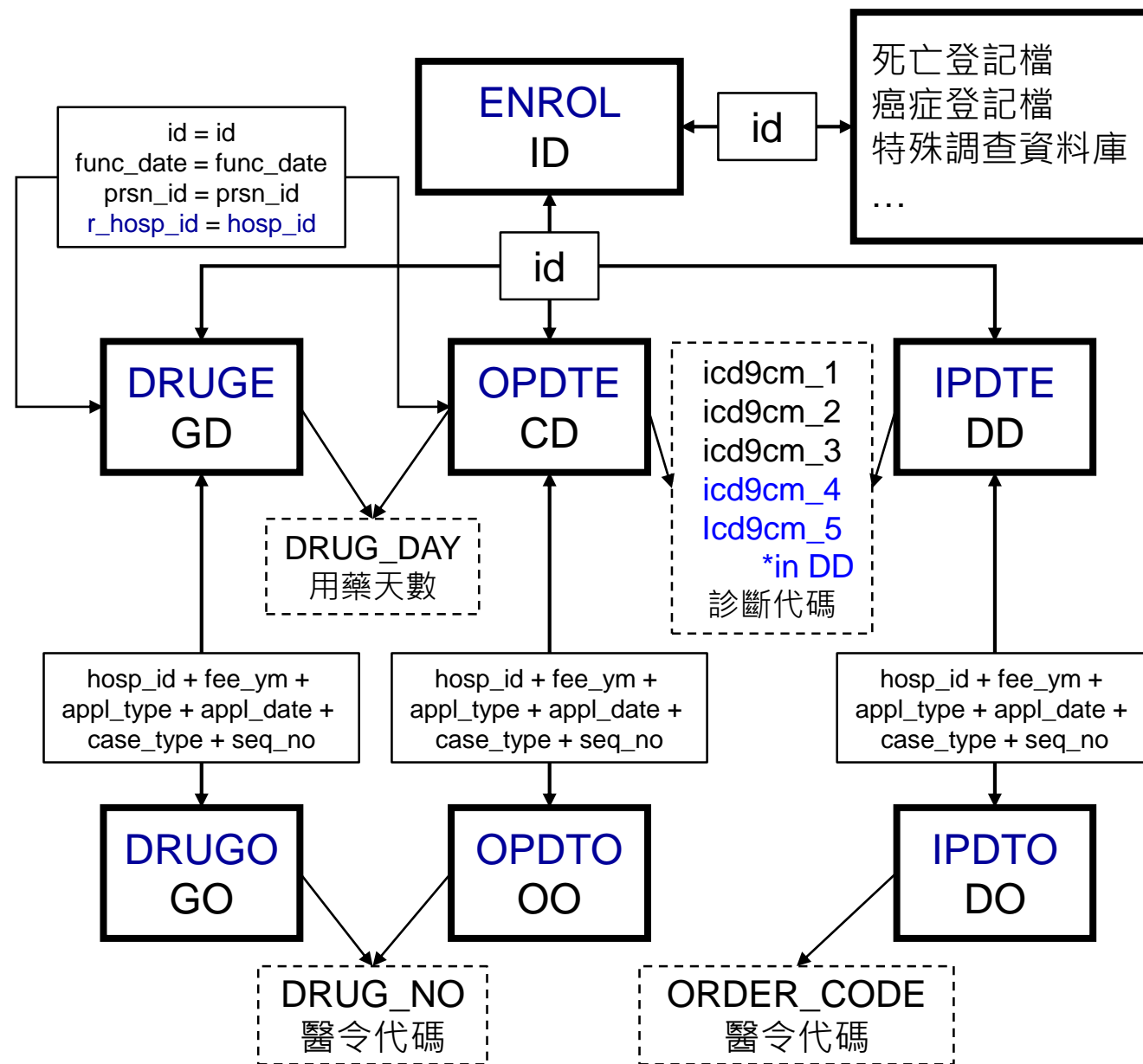
- 重要性
  - 執行計畫書把研究團隊成員的呼吸同步
  - 避免誤會、多頭馬車、錯誤結果
- 組織架構
  - 研究議題、操作定義、可行評估
  - 研究設計、統計方法、預期結果

# 操作定義

- 高齡族群
  - 資料年度 ( 2014 ) 時已經滿 (  $\geq$  ) 65歲之健保被保險人
- 糖尿病
  - 門診就醫任一個診斷碼欄位 ( icd9cm\_1 OR icd9cm\_2 OR icd9cm\_2 )  
當中出現糖尿病診斷碼
    - ICD-9-CM任一欄位為250開頭之診斷
    - 同一天有多次紀錄的話，只當作一次
  - 一整年 ( 2014 ) 當中只要出現至少2次糖尿病診斷碼，即認定為糖尿病人
  - 計算整年度歸人 ( id ) 之後的糖尿病盛行率

# 可行評估

- 健保承保檔
  - id、id\_birth\_y、id\_s
- 門診費用檔
  - id、icd9cm\_1 ~ 3



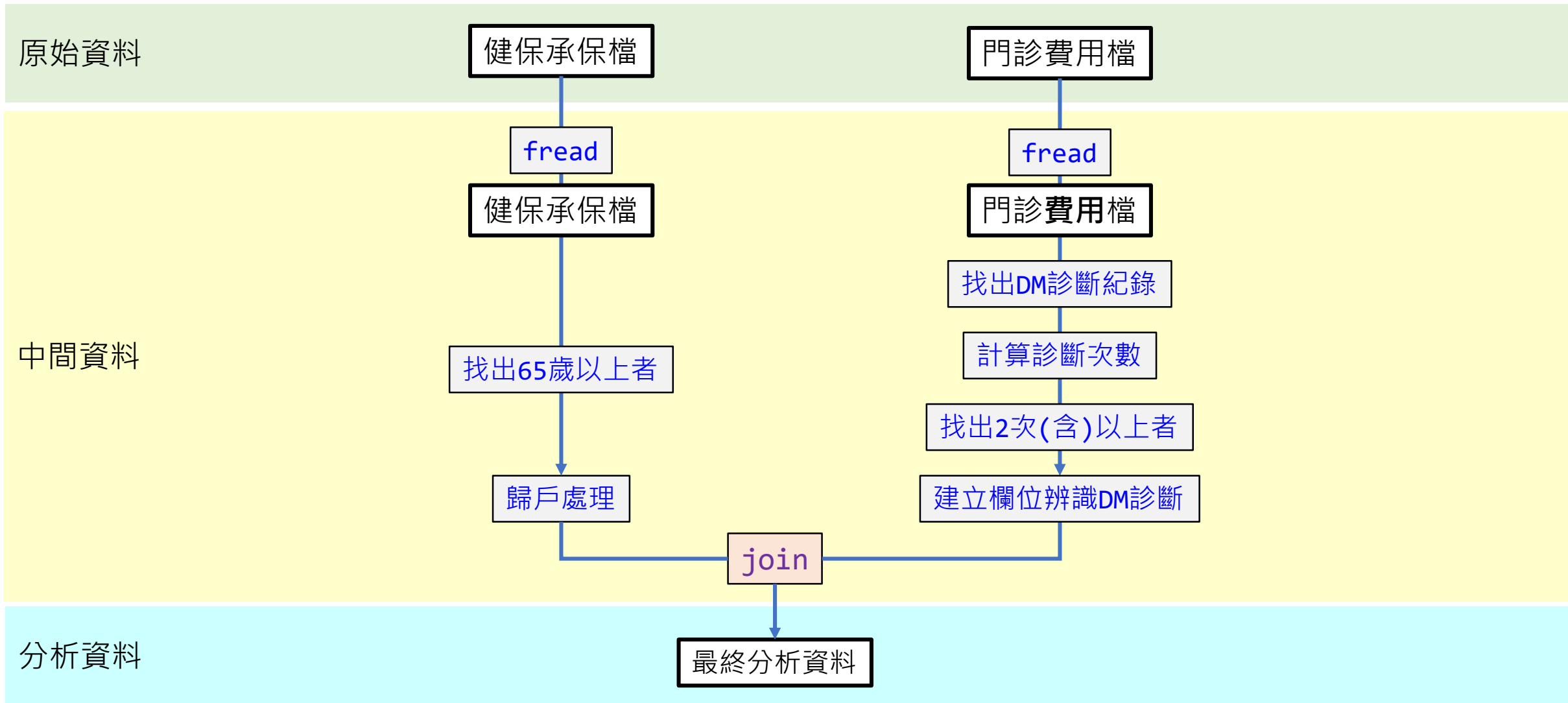
# 臺灣高齡族群的糖尿病盛行率調查

- 研究設計
  - 橫斷研究（調查2014年當下的狀況）
- 統計分析
  - 計算次數及百分比
- 預期結果（循序漸進）
  - 資料庫當中的納保人口有\_\_\_\_人？
  - 資料庫當中的納保人口有\_\_\_\_人為65歲（含）以上？
  - 資料庫當中的65歲（含）以上納保人口，糖尿病盛行率為\_\_\_\_%？

# 健康資料研究的執行計畫書組織架構重點回顧

- 研究議題
  - 一次講清楚一件事情
- 資料可行性評估與操作型定義
  - ICD codes、ATC codes、健保藥品碼、健保給付碼
- 研究設計與統計分析方法
  - 正確的學理與時間邏輯、適合的統計分析方法
- 預期研究結果架構
  - 圖表先想好，論文寫的好

# 工作流程：圖塊、關聯、流向、偽代碼





# 思考程式碼的大綱，可以一路由上往下執行

- # ~ 專案設定 -----
- # 路徑
- # 套件
  
- #### 資料管理 -----
- # ~ 高齡者 -----
- # ~ 糖尿病診斷 -----
- # ~ 年度就醫歸人 -----
- # ~ 合併族群總檔 -----
  
- #### 統計分析 -----

# Summary

- 衛生福利資料庫
  - 資料來源、管理模式、資料樣態
  - 參考工具、關聯架構
- 流行病學調查研究實作
  - 研究命題、研究團隊、執行計畫
  - 核心理念、工作流程、技術實踐

- 開放提問時間

- FB：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

