

第四章 資料來源與處理

健保從 1995 年開辦至今已有 14 年，應學術界的請託，中央健保局委託國家衛生研究院以全民的資料為基礎，建立一個以學術研究為目的之資料庫，提供給專家學者從事相關醫學研究。全民健康保險研究資料庫（National Health Insurance Research Database）分為四大類：基本資料檔、系統抽樣檔、特定主題分檔以及抽樣歸人檔。

第一節 資料來源

本文研究精神分裂症用藥，使用的資料是健保資料庫裡「特定主題分檔」之中的「精神疾病住院病患歸人檔」(Psychiatric Inpatient Medical claim dataset，簡稱 PIMC)，根據國衛院⁹的說明，選取標準是從 1996 至 2001 年全民健保的住院檔中，選取就醫科別為精神科，且 ICD-9 診斷碼為 290 至 319 中任一項，或 A-Code 碼為 A210 至 A219 之任一項的住院資料，再從這些資料擷取已轉碼的身份證統一編號，共得到 91,104 個 ID，最後再用這些 ID 選取 1996~2002 的所有就醫資料，形成 PIMC 檔案。表 4.1 是本研究用到的門診及住院之醫令檔、費用檔：

表 4.1 精神疾病原始資料檔 (Original claim data)¹⁰

檔名	中文名稱	英文名稱
PSY_CD	門診處方及治療明細檔	Ambulatory care expenditures by visits
PSY_OO	門診處方醫令明細檔	Details of ambulatory care orders
PSY_DD	住院醫療費用清單明細檔	Inpatient expenditures by admissions
PSY_DO	住院醫療費用醫令清單明細檔	Details of inpatient orders

⁹張景瑞、邱淑怡、張新儀、丘政民、高淑芬、陳為堅、張憶壽（2004），全民健保精神疾病住院病患歸人檔簡介。

¹⁰全民健康保險研究資料庫網站：<http://www.nhri.org.tw/nhird/index.htm/> 資料庫內容/檔案英文譯名

另外，本文主要是針對醫生開藥行為做研究，所以像醫生、醫院的特性也是我們感興趣的部分。其中藥品主檔更是筆者選取樣本的主要依據，首先將 2002~2004 的藥品資料合併，留下藥理分類為 281600~282800 的精神治療藥，接著按照主成分挑出精神分裂症一代、二代用藥，最後刪除注射液和口服液，只留下藥錠部分，得到 Sch_drug 檔。

表 4.2 基本資料檔 (Registration files)

檔名	中文名稱	英文名稱
PER	醫事人員基本資料檔	Registry for medical personnel
DOC	專科醫師證書主檔	Registry for board-certified specialists
HOSB	醫事機構基本資料檔	Registry for contracted medical facilities
DRUG	藥品主檔	略
Sch_drug	精神分裂症用藥檔	略

第二節 串檔說明及串檔結果

全民健保資料庫龐大繁雜，在串檔的過程必須小心謹慎，稍有錯誤就可能造成之後的分析困難，本文先利用精神分裂用藥檔挑出門診、住院費用檔中，有使用精神分裂藥的資料，接著擷取其中的 ID，再利用此 ID 檔和藥檔合併門診、住院的醫令檔，最後合併醫師及醫療院所的資訊，就可得到需要的資料。串檔方式如下圖 4.1 所示。

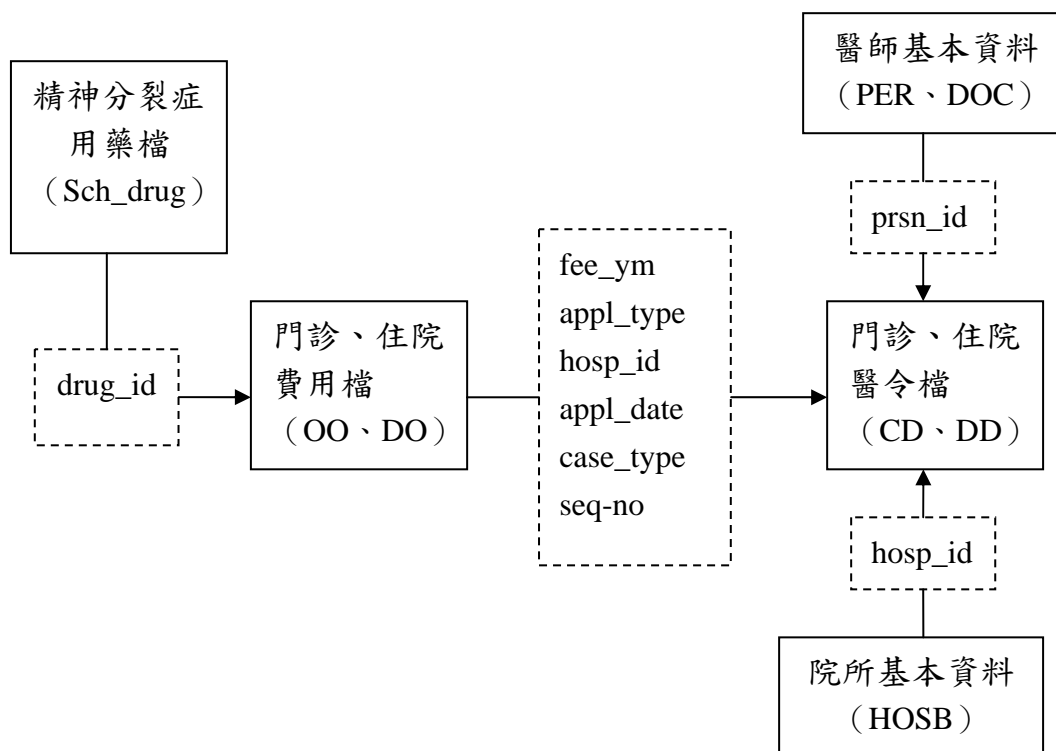


圖 4.1 串檔流程圖¹¹

串檔完成後，經過初步的作圖分析顯示，不論是門診或住診，醫生的開藥行為都蠻類似的，所以最後我們選擇門診做為主要的分析依據。此時門診中共有 4,126,201 筆資料，我們從當中選取需要的樣本及清除無法分析的資訊。

第三節 選取樣本及清檔

表 4.3 即為選取樣本後，醫師人數及病患人數的變化，以下簡單說明過程。

(一) 刪除非精神科專科醫師

在醫生的樣本選取方面，我們利用專科醫師證書主檔 (DOC) 中的「專科醫師科別」來選取，若考取的前三張證書中有精神科專科醫師 (A1300) 者，我們即判定此人為精神科專科醫師。資料中專科醫師人數有 844 人，非專科醫師有 9,331 人，開藥筆數卻只佔全部的 6.81%，本文主要研究精神科專科醫師

¹¹實線框內為所使用的資料檔，虛線框內為串檔變數，箭號代表串檔方式。

的開藥行為，故刪除非精神科專科醫師的資料。

(二) 選取診斷為精神分裂症病人，即選取 ICD-9 CODE 碼為 295-296 或 A-CODE 編碼為 A211 者，且只保留就醫科別為“13”精神科的資料
樣本數此時為 3,817,799 筆，被診斷為精神分裂症者有 3,092,222 筆（佔 80.99%），這裡刪除非被診斷為精神分裂症的 725,577 筆（19.01%）。門診檔資料中有一變數為 Func_type 代表了就醫科別，在本資料中就醫科別為精神科者佔了 99.35%，故將其它不是到精神科就診的資料刪除。

(三) 刪除樣本數不足的醫院及看診次數過少的資料

未刪除之前，醫院數共有 424 家，其中特約類別為居家照護的醫院只有一家，另外基層院所佔資料量百分比為 2.96%。在型態別方面，專科醫院有 3 家、特殊科醫院和慢性醫院都只有一家。這些醫院佔總類別過少，樣本數不足故刪除。另外有些醫生一季看診不到 10 次，由於資料過少，故刪除一季內看診小於 10 次的資料（佔總資料 0.27%）。

表 4.3 選取樣本後之病患數及醫師數變化

年度	人數	註 ¹²	(一)	(二)	(三)
1998	病患	33,786	32,152	26,095	26,058
	醫師	4,026	663	628	594
2000	病患	36,823	35,611	28,990	28,950
	醫師	3,895	738	720	677
2002	病患	36,289	35,335	29,265	29,235
	醫師	3,581	768	757	713
2004	病患	35,227	34,012	28,513	28,477
	醫師	3,728	770	754	717

¹²未刪除任何樣本之前的人數。

第四節 樣本整理

經過上述的清檔之後，我們最後可得到 3,034,720 筆資料，由於醫生是主要研究標的，我們想要瞭解這 701 位醫生在 1997~2004 年的開藥行為，於是筆者把資料改成以醫生為單位，每一年、每一季的形式。由於在一季之中，醫生的年齡、性別，醫院的特約類別及特性是不會改變的，所以這類的變數不必做更動。

相較之下，其他資料必須經過一些處理，例如：我們的被解釋變數是一季當中開新藥的比例，首先用虛擬變數標示出一代藥 ($di=0$) 及二代藥 ($di=1$)，就此我們可以算出醫生在某一季的開新藥比例。此數值介於 0 與 1 之間，若越接近 0，代表此醫生在這一季較常開舊藥，相反的，若越接近 1，代表這位醫師開出的新藥較多。同理用在解釋變數上，病人的性別也設為虛擬變數，女性為 0，男性為 1，就可算出一季之中某醫生看病人的性別比例。另外，我們也要算出每一季當中，病人的平均年齡以及藥品的平均價格。

最後一個也是最重要的變數，就是周遭醫生開藥行為所造成的影響，在此變數上處理也最麻煩且複雜，為了更清楚說明，下面用表格的方式呈現此變數的作法。

首先我們將醫事機構照年、季來排列，之後我們加總同一個醫院的醫生季開新藥比例，這個變數暫稱為「季開新藥比例加總」(q_sum)，接著算出同一家醫院的醫生數量 ($prsn_N$)，最後要算出周遭醫生開新藥比例 ($near_prop$) 對自身所造成的影響，即是把季開新藥比例加總，扣除自己的季開新藥比例，再除以醫院內除了本身的醫生數，就會得出我們所需要的變數。若用公式說明如下：

$$near_prop = (q_sum - q_prop) / (prsn_N - 1)$$

表 4.4 周遭醫師影響作法

醫事機構 代號	年度	季	醫事人員代號	季開新藥 比例	季開新藥 比例加總	同一醫院 的醫生數	周遭醫生 開藥影響
hosp_id	year	quarter	prsn_id	q_prop	q_sum	prsn_N	near_prop
131504632	1999	4	FS19224876	0.090909	0.307583	5	0.0541684
131504632	1999	4	ES07419277	0	0.307583	5	0.0768956
131504632	1999	4	RU26379772	0.062827	0.307583	5	0.0611888
131504632	1999	4	TS66388470	0.153846	0.307583	5	0.0384341
131504632	1999	4	ET96211577	0	0.307583	5	0.0768956

另外，near_prop 這個變數有可能是無此資料 (.) 的，原因是若在一家醫院只有一位精神科醫生，也就沒有所謂周遭醫生的影響了。接著也簡單說明資深醫師的比例作法。

同樣的，我們把資料照醫事機構代號、年度、季、醫生年齡、代號來排序，以表 4.5 為例，將醫師由年長排序到最年輕，分別為 1~5，代表此醫院在 1999 年第 4 季時有 5 位醫師，利用季開新藥比例來加總資深醫師的開藥比例。

表 4.5 周遭資深醫師比例作法

醫事機構 代號	年度	季	醫事人員代 號	醫師 年齡	年齡 排序	季開新 藥比例	資深醫師開 藥比例加總
hosp_id	year	quarter	prsn_id	prsn_age	old_n	q_prop	old_prop
131504632	1999	4	FS19224876	40	1	0.09091	0.0909091
131504632	1999	4	ES07419277	40	2	0	0.0909091
131504632	1999	4	RU26379772	37	3	0.06283	0.1537363
131504632	1999	4	TS66388470	35	4	0.15385	0.3075825
131504632	1999	4	ET96211577	34	5	0	0.3075825

接著為了要算出比自己年長的醫師，我們將年齡排序數減 1，即為比自己年長醫師的數量，另外把比自己資深醫師的開藥比例全部加總，成為資深醫師

開藥比例，最後再把資深醫師開藥比例除以比自己年紀大的醫生數量，可得出資深同儕影響¹³（如表 4.6 中粗體字部分）。同理，年輕同儕的醫師作法相同，只要將年齡由小到大排序即可算出年輕同儕的影響。不過這裡要注意的是這樣的作法會有缺漏值產生，例如在這醫院的第一筆及最後一筆資料會是 missing，最後只會留下中間三位醫師的資料。另外一種情況是一個醫院只有兩位以下的精神科醫師（共有 5,507 筆），則兩位醫師都會有缺漏值產生。

表 4.6 資深同儕影響

年度	季	醫事人員 代號	醫師 年齡	資深醫 師數量	季開新 藥比例	資深醫師開 藥比例	資深同儕 影響
year	quarter	prsn_id	prsn_age	old_n	q_prop	old_prop	old_affect
1999	4	FS19224876	40	0	0.09091	.	.
1999	4	ES07419277	40	1	0	0.0909091	0.0909091
1999	4	RU26379772	37	2	0.06283	0.0909091	0.0454545
1999	4	TS66388470	35	3	0.15385	0.1537363	0.0512454
1999	4	ET96211577	34	4	0	0.3075825	0.0768956

最後，若把周遭醫生影響不存在的筆數刪除(11,005 筆)，則最後會有 14,949 筆樣本，是這 701 位醫生在 190 家醫療院所，自 1997~2004 這 8 年以來每一季的資料。

我們利用這些樣本即可做出下一章的敘述統計及迴歸結果。

¹³公式說明： $old_affect = old_prop / old_n$