

如何使用健保資料進行經濟研究

連賢明*

全民健保自民國 84 年開辦後，至今已累積近 10 年的健保資料。如何透過已累積經驗，從中學習控制逐年高漲的健保費用是衛生當局的一個重要課題。本文從一個健保資料研究者的角度，討論健保資料庫的優點和缺點，使用原則和操作技巧，以及潛力和限制。作者首先透過和其他現有醫療資料庫的比較，使讀者瞭解健保資料庫的優缺點。由於健保資料庫檔案眾多，依據各檔案特性，作者先將健保資料庫區分為三大子資料庫：費用檔，醫令檔，和基本資料檔；其次，針對各子資料庫的整理、各子資料庫間的串結、以及健保和非健保資料連結這三部分，說明健保資料的處理原則。最後，透過一個研究實例，詳細解釋如何將處理原則和實務操作相結合，並提供一些研究方向供參考。

關鍵詞： 健保資料庫，醫療研究資料，全民健保

JEL 分類代號： I10, I19

1 前言

全民健保自民國 84 年開辦，至今已邁入第 12 年。由於健保和民衆就醫息息相關，全民健保政策的改變，諸如保險費率增加、藥價基準調整、總額制度施行、乃至卓越計畫等，每一項政策改變都引起社會大眾的諸多關心，正

*作者為國立政治大學財政系(所)助理教授。作者謝謝兩位匿名評審、許績天、劉錦添教授的評論與建議，以及韓幸紋和何慧敏兩位助理的協助。作者也感謝國家衛生研究院提供健保資料庫供研究使用，以及國家衛生研究院(NHRI-EX94-9204PP)和國科會(94-2415-H-004-013)的研究經費補助。所有文責均由作者自負。

經濟論文叢刊 (*Taiwan Economic Review*), 36:1 (2008), 115–143。
國立台灣大學經濟學系出版

反意見常常激辯不休。以最近備受爭議的醫院門住診費用比例而言,¹贊成者認為提高住診費用比例能使醫院將重心轉移到住院治療上,回歸醫院以重症治療為主的角色;²反對者認為此舉進一步壓縮了地區醫院的生存空間,也剝奪了病患就醫選擇的自由,導致醫院間的不公平競爭,間接造成醫療品質的下降。³可惜的是,由於缺乏實證基礎,在這些正反雙方論戰中,往往各執一詞,最後淪為各說各話。

為了促進健保相關研究,中央健保局自2000年起,委託國家衛生研究院發行全民健保資料庫,希望透過健保資料的發行,累積實證基礎,進而帶動健保政策的深入討論。這個資料庫累積了台灣2,300萬人口,自健保開辦以後的所有健保就醫資料,並允許研究者連結就醫病患、醫師、以及醫療院所的基本特性,不論就深度和廣度而言,和國外類似醫療資料庫相比較毫不遜色,⁴在費用上更不及國外類似資料庫價格的數十分之一。⁵令人惋惜的是,或受限於資料規模龐大,或受限於處理分析複雜,或受限於重要變數缺乏,許多研究者對使用健保資料庫猶豫不前,原先透過開放健保資料庫來深化健保政策討論的目標,並不如預期般的成功。⁶

¹健保局於2003年初實施自主管理計畫,按季與醫院協商醫療費用固定上限。2003年7月起進一步為配合醫院訂出相關指標,如門住診費用比為45:55,降低不必要用藥與重複檢查等,通稱「卓越計畫」。

²醫院總額執行委員會執行長莊逸洲認為大型醫院的門診服務需做合理限制,以確保醫療品質。詳見民生報(2004a)。

³台灣地區醫院協會理事長謝文輝指出健保局強迫地區醫院將原有的53:47的門、住院比例,強制改為44:55,導致地區醫院門診點值滑落,影響民衆就醫權益;敏盛醫院負責人楊敏盛則表示,取消卓越計畫可使所有醫療院所公平競爭。詳見民生報(2004b)。

⁴美國的Medicare資料在架構上和台灣健保資料庫相類似。Medicare資料目前發行的有兩種,一種是完整版,包涵全美65歲以上人口(約3,000萬人)的保險醫療利用記錄,同時具有轉碼後的病人代碼;另一種則是5%的抽樣檔,抽樣檔因保密原因排除許多病人和醫師相關資訊,諸如醫師和病人代碼、病人生日、和居住地點,不如完整版所含資訊較多;然而,Medicare審查相當嚴格,一般研究者通常僅能取得抽樣版。

⁵健保資料庫的住院醫療費用明細清單檔記錄台灣2,300萬人當年住院資料。這個檔的資料量大約每年兩片光碟片,每片收費新台幣500元,合計NT1000元;美國Medicare抽樣檔(涵蓋約150萬老人)的住院資料,每年收費約400美元。台灣健保資料庫收費低廉可見一般。

⁶若以國衛院的研究成果繳交清單來計算,92和93年度使用健保資料共有69和66篇論文,94則減為51篇論文。有關健保資料研究成果繳交清單可見於網頁(http://www.nhri.org.tw/nhird/talk_05.htm)。

筆者自 2002 年開始接觸健保資料庫，在過去 4 年間，一方面為健保資料庫的處理龐雜，頭痛不已；另一方面，也對健保資料庫的豐富詳盡，暗自心喜。筆者深信在妥善使用下，全民健保資料庫不但具有學術發表的潛力，也將成為政策規劃的基石。在這個前提下，筆者將自己個人經驗和讀者分享，拋磚引玉，期望更多先進能貢獻他們的寶貴心得，吸引更多學者投入這個領域的研究。

本文各節大綱如下。第 2 節介紹台灣現有醫療資料庫，希望透過其他醫療資料庫介紹，一方面使讀者瞭解現有資料庫，一方面也透過不同資料庫的比較，更清楚知道健保資料的優缺點；第 3 節先針對健保資料檔案的特性，將健保資料區分為三大子資料庫：費用檔，醫令檔，和基本資料檔，並針對各子資料庫的整理、各子資料庫間的串結，以及非健保和健保資料如何連結等三部分，詳細說明健保資料的處理原則和技巧；第 4 節透過一個研究實例，使讀者能進一步將前述健保資料的使用原則和實務操作相結合；最後一節總結，並提供一些研究方向供讀者參考。

2 台灣現有醫療資料庫

在介紹全民健保資料庫前，先介紹台灣一些現有的醫療資料庫。由於相關資料庫眾多，一一列舉有實際困難。我們因此採下列標準選取資料庫，第一，資料庫為具有公信力之機構所蒐集；第二，樣本具有代表性；第三，資料庫為長期收集，非單次或單年資料；第四，資料有公開發行或申請管道。在這些標準下，未公開發行或申請困難資料（如癌症登記檔、出生登記檔、和死亡登記檔），或是針對某單一議題所作單次問卷（如台灣地區青少年性知識調查），不列在介紹範圍。

為了使讀者能清楚區分資料庫特性，我們採取和 Evans et al. (2000) 類似作法，將相關醫療資料庫區分為兩類：一類著重在需求面（就醫病患），另一類則著重在供給面（醫療院所）。表 1 列出和供給者相關的資料庫。目前台灣有關醫療院所資料庫並不多，除健保資料以外，較完整的就只有衛生署的醫療院所現況調查及醫療院所服務量調查。⁷ 前者是以所有醫療院所

⁷衛生署每年根據這兩份資料，整理成醫療機構現況及醫療服務量統計摘要。該統計包含有醫事機構概況、醫院家數、病床數、醫事人員數、設備數，以及服務量等資訊。這份統

為樣本,紀錄1988年以來每年現存醫療院所地址、特性(如權屬、評鑑、型態別)、各類病床、醫事人員數目,以及各項醫療設備數目(如昂貴儀器);後者則針對1992年以來,各醫院的服務量(如門診人數、住院天數、生產、剖腹產人數等)、各類病床使用天數,以及昂貴儀器使用數做調查。由於這兩項調查均保留院所原始代碼,且資料調查時間橫跨健保前後,是國內目前少數能分析實施健保政策效果的重要資料。

相較於醫療院所現況調查和服務量調查,健保基本資料檔所含院所資料明顯薄弱。由表1可知,院所基本資料檔僅紀錄健保相關門住診服務總量,並未涵蓋所有細項服務量(如剖腹產人數),也未包括非健保給付服務,更未針對院所所有各項設備、人力、以及昂貴儀器做紀錄。⁸然而,院所基本資料檔的最大優勢在於這些院所基本資料能與其他健保子資料庫相連結,並透過其他子資料檔串連出各院所當年度的各類細項服務量,甚至是連結到任一單次醫療服務。這個特點在下面討論中更能清楚說明。

表2列出和消費者相關資料庫。由表2清楚可知,這些資料庫不論是就樣本規模、抽樣單位,乃至於涵蓋的問項都有顯著差異。由於本文重點在介紹健保資料庫,我們僅透過和其他資料庫比較,使讀者了解健保資料庫優缺點,對於非健保資料庫,限於篇幅關係,並不一一贅述。⁹透過表2,我們可以清楚知道健保資料庫的幾個不足之處。第一,健保資料是以個人為樣本單位,所以不包括家戶組成的相關資料;第二,健保資料僅有保險給付相關資訊,不包括非保險給付之保健支出(如牙齒矯正或心臟支架);第三,健保資料僅有健保就診診斷資料,不包括投保人自覺健康資訊,亦無家族成員疾病史;最後,也是最重要的,健保資料對病患特性僅紀錄病患性別及生日,並不包括重要社經(如教育或所得)和生命變數(如死亡日期)。最後這個缺點對病患特性的掌握造成相當嚴重的限制。

計資料的網址為: <http://www.doh.gov.tw/statistic/index.htm>。

⁸健保院所基本資料檔中有包含醫事人力基本資料檔;然而,該檔並無法用於計算單一院所所使用的醫事人力。

⁹有關表2非健保相關醫療資料庫,絕大部分可透過中研院的調查研究中心取得調查資料以及資料說明(包括抽樣設計、樣本特性、相關變數)。該中心網址為(<http://www.sinica.edu.tw/as/survey>)。最新的國民健康調查,老人保健調查,家庭與生育力調查則需透過國民健康局取得,國民健康局網址為(<http://www.bhp.doh.gov.tw/BHP/do/www/themeParkView?themeParkId=538>)。

表 1: 醫療相關資料庫特性 (供給面)

資料名稱	資料型態	抽樣單位	起迄年份	抽樣間隔	地點	評鑑 ^a	權屬 ^b	人力 ^c	病床 ^d	昂貴儀器 ^e	服務量 ^f
健保院所基本資料檔*	縱橫	特約院所	1997-2004	1年	○	○	○	×	○	×	○
醫療院所現況調查	縱橫	全體院所	1988-2005	1年	○	○	○	○	○	○	×
醫療院所服務量調查	縱橫	全體醫院	1992-2005	1年	○	○	○	×	×	×	○

* 健保院所基本資料檔包括醫事機構基本資料檔、病床主檔、副檔資料檔、診療科別明細檔、門診費用申請總表主檔, 以及住院費用申請總表主檔。

a 健保院所基本資料檔評鑑等級大略區分醫學中心、區域醫院、地區醫院, 以及基層診所四類; 醫療院所現況及服務量調查較詳細, 除了前四類之外精神專科醫院和教學醫院等分類。可惜的是, 現況及服務量調查所用標準歷年不齊一, 造成跨年比較上的困難。

b 健保院所基本資料檔權屬別採細分類, 研究者需自行將其整合為大分類 (如市立或署立合併為公立醫院, 宗教或非宗教合併為法人醫院等); 醫療及服務量調查則直接區分大類。

c 醫療院所現況及服務量調查在人力部份區分為專業醫師、執業醫事人員, 以及未具醫事專門職業證書工作者三大類, 其下再細分各類和其人數, 以再細分為醫師、中醫師, 以及牙醫師等。

d 健保院所基本資料檔 (病床主檔) 記錄了每個病床號碼和其使用目的 (如急性病床、洗腎病床等); 醫療院所現況及服務量調查僅記錄一同類的病及性病床, 各類特殊床等) 以及其各類病床數目。兩個資料分類類似但有差異。

e 醫療院所現況調查對特定醫療技術檢查、檢驗儀器 (如磁振造影機, 電腦斷層掃描儀等) 皆有統計其登記數目。然而, 這個數目為院所向主管機關? 和院所實際數目有一定差距。

f 健保院所基本資料檔僅登記門診、住診次數, 門、住診金額, 以及血液透析服務量; 醫療院所服務量調查除上述服務量以外, 尚有生產等服務量人次, 儀器之使用人次。

表 2: 醫療相關資料庫特性 (需求面)

資料名稱	資料型態	起迄年份	抽樣間隔	抽樣單位/個數	社經變數	醫療費用 ^c	健康狀態 ^f
全民健康保險資料庫	縱橫	1996-2004	1年	個人/2,300 萬人	無	健保 (分項)	診斷
主計處家庭收支調查 ^a	橫斷	1968-2004	1年	家戶/1.4 萬戶	教育、婚姻狀態 所得、家戶組成	全部 (分項)	無
國民健康訪問調查 ^b	橫斷	2001-2005	4年	家戶/0.7 萬戶	教育、婚姻狀態 所得、家戶組成	全部 (分項)	自覺、診斷、 疾病史
國民營養調查 ^c	橫斷	1980-2002	5年	家戶/1 萬人	教育、婚姻狀態 所得、家戶組成	無	檢查、疾病史
華人家庭動態資料庫	縱橫	1999-2003	1年	家戶/0.1 萬戶	教育、婚姻狀態 所得、家戶組成	全部	自覺
老人保健與生活調查 ^d	縱橫	1989-2003	3-4年	個人 (老人)/ 0.5 萬人	教育、婚姻狀態 所得、家戶組成	全部	自覺、疾病史
台灣家庭與生育力調查 ^b	橫斷	1965-2004	2-7年	個人 (婦女)/ 0.2 萬-1.2 萬人	教育、婚姻狀態 所得、家戶組成	無	自覺

資料出處：非健保醫療資料庫，絕大部分可透過中研院的調查研究中心取得調查資料，以及資料的抽樣設計、樣本特性、相關變數等說明。中心網址為 (<http://www.sinica.edu.tw/as/survey>)。最新的國民健康調查，老人保健調查，家庭與生育力調查則需透過國民健康局取得，國健局的網址為 (<http://www.bhpdoh.gov.tw/BHP/DO/www/themeParkView?themeParkId=538>)。

a 主計處家庭收支調查始於民國 53 年，每兩年調查 1 次，但從 59 年起，改為每年調查 1 次。

b 2001 年國民健康訪問調查同時串聯前後 3 年受訪者健保利用資料，所以同時具有疾病診斷和疾病史資訊。健康訪查自 2005 年起給為以個人為單位抽樣。國民營養調查於 1999 年後，修改為國民健康狀況變遷調查。

d 老人保健與生活問題調查於 1996 年之後，擴大調查對象，修改為中老年保健與生涯規劃調查。

e 全部是指包括健保和非健保所有醫療費用；分項是指調查除整體醫療費用外，另外再細分門診、住診、預防保健等項目。

f 診斷是指資料列出就醫疾病分類代碼，疾病史是指資料記載個人是否罹患常見疾病，自覺是指個人自覺健康狀態，檢查是指調查中附帶進行簡單檢查。

表 2 也相對顯示出健保資料庫的優點。健保資料庫涵蓋了台灣 2,300 萬人，從 1996 年至 2004 年間所有健保使用資料。¹⁰ 這些資料不論是在樣本規模或涵蓋時間，即使相較於國外類似資料庫，其完整性可說是相當少見。更重要的是，健保資料保留了院所，醫師，以及病人的一致代碼，可將歷年醫療使用記錄，轉換為以病患、醫院或醫師為單位的縱貫資料。並且可透過這些一致代碼，進一步將病患、醫師、以及醫療院所的基本資料相結合。有關這些資料相互連結的問題，我們留待後文介紹。

3 全民健保資料庫

本節主要說明健保資料庫的結構、各子資料庫的整理、各子資料庫間的連結，以及如何透過連結外部資料來彌補一些重要變數的缺乏。有關全民健保資料庫的申請資格、所需費用、以及使用規定，請參見國家衛生研究院所設置的全民健保資料庫網站 (www.nhri.org.tw/nhird) 或盧瑞芬·文羽萃 (2003) 的介紹。

3.1 健保資料結構

全民健保資料庫是由中央健保局於 1997 年委託國家衛生研究院，依據醫療機構向健保局申報資料，在保障民眾隱私以及資料安全的前提下，自 2000 年起每年年底對外發行前一年的健保申報資料。自健保實施以來，除了 1995 年全部和 1996 年部份資料未公布外，其餘均已對外發行，目前有 1997 到 2004 年，共八個年度的健保資料。

由於健保資料庫的原始目的是醫療費用申報，要清楚了解健保資料，首先須瞭解醫療院所申報流程。在一個診療過程中，醫師會根據病患病徵，做出檢查，診斷，以及處置，這其中的每個項目都依據「全民健保給付標準」與予給付。¹¹ 舉例來說，某病患因頭痛及發燒就診，醫師針對病徵，判斷病

¹⁰根據健保局公布資料，截至 95 年 6 月底止健保總投保人數（不包含軍人）共有 22,374,752 人，實質納保率達 99% 以上。資料詳見 (http://www.nhi.gov.tw/webdata/AttachFiles/Attach_8219_2_135_業務執行報告.pdf)。

¹¹健保支付標準根據處置，藥品，以及特殊材料區分為三大類。這三類的支付標準可在健保局網站上找到，網址為 http://www.nhi.gov.tw/webdata/webdata.asp?menu=1&menu_id=26&webdata_id=870。

患可能罹患感冒，爲了瞭解感冒是否併發肺炎，醫師可能要求病患做 X 光檢查，一旦排除併發肺炎的可能，醫師給予病患點滴注射和 3 天份藥品（如抗生素、退燒藥、胃藥）。這些檢查（X 光攝影）、診斷（上呼吸道感染），以及處置（點滴注射，藥品給予），在支付標準上均有相對應的醫令，每一醫令又有相對應的單價，加總後再扣除病人的自付額，即是單次的健保醫療門診申報金額。

在了解醫療費用申報的流程後，我們開始說明健保資料的結構。每一單項處置、檢查，以及診療等醫令，有關住院部份被置於「住院醫療費用醫令清單明細檔」，門診部分被置於「門診處方醫令明細檔」（這兩個檔以下簡稱醫令檔）；由於單次就診（特別是住院）可能涵蓋數十，甚至數百筆的醫令，爲了方便起見，健保局針對門診或住院，將各項醫令分類加總，各項醫療資訊彙整，形成以門診爲單位的「門診處方及治療明細檔」，和以住院爲單位的「住院醫療費用清單明細檔」（這兩個檔以下簡稱費用檔）。這兩個費用檔記載了就診相關的人（就診病患與醫師）、事（就診疾病）、時（就診時間）、地（就診院所）以及錢（就診費用），可說是健保資料庫的核心。最後，爲了使研究者能針對病患、醫師、以及醫療院所基本特性有所掌握，健保資料庫另行針對病患、醫師，以及院所各自發行不同的「基本資料檔」。

針對上述說明，圖 1 列出全民健保資料的三大子資料庫：費用檔、醫令檔、以及基本資料檔。依照國衛院的檔案分類和英文縮寫名稱，費用檔包括了門診處方及治療明細檔（CD）和住院醫療費用清單明細檔（DD）；醫令檔包括了門診處方醫令明細檔（OO）與住院醫療費用醫令清單明細檔（DO）；基本資料檔包括和醫事人力相關的醫事人員基本資料檔（PER）與專科醫師證書主檔（DOC），和醫院相關的醫事機構基本資料檔（HOSB），醫事機構副檔資料檔（HOSX），醫事機構診療科別明細檔（DETA），醫事機構病床主檔（BED），門診費用申請總表主檔（CT），與住院費用申請總表主檔（DT），還有和病人相關的承保資料檔（ID）和重大傷病證明明細檔（HV）。¹² 這三大子資料庫互有特色，可視研究目的相互結合。若研究重心在於門診醫療費用，可單純使用門診費用檔；若想了解門診處置是否合宜，

¹²嚴格來說，健保資料庫上尚包括特約藥局處方及調劑明細檔（GD），特約藥局醫令檔（GO），與藥品主檔（DRUG）。但由於台灣醫藥並沒有明確分業，前面兩個檔案在健保資料庫的價值並不是很大，最後一個檔案和健保支付標準的藥品支付標準大半雷同。

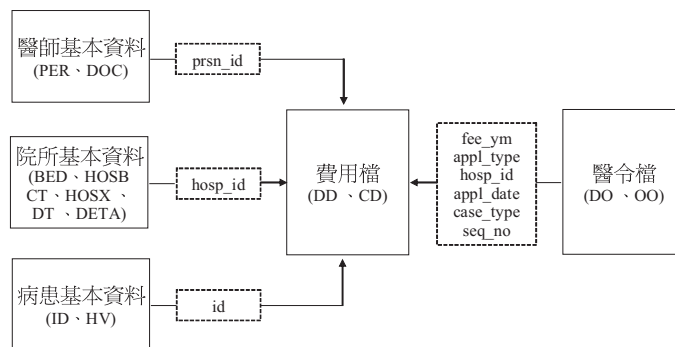


圖 1: 全民健保資料的三大子資料庫

則需串連門診醫令檔；想進一步瞭解不同院所或醫師的處置差異，則需串連醫療院所或醫事人員的基本資料檔。舉例來說，若研究者想比較新進和資深外科醫師的手術成果，則需使用住院費用檔選出相關手術、串連住院醫令檔找出處置、並串連醫事人員基本檔來區分新進和資深外科醫師。

透過這三個子資料庫，理論上研究者可根據自身需要，選取適當樣本進行分析。但由於健保資料龐大，自行選取資料有實務操作上的困難。¹³ 也因為如此，國衛院針對不同研究需要，事先選取樣本，將健保資料維持在一個易於操作的規模內。目前公開發行的資料中，有以門診或住院次數為抽樣單位的「系統抽樣檔」；有以病患為抽樣單位的「抽樣歸人檔」；及以不同疾病、病患身分，以及醫療層級為選取基準的「特定主題分檔」。此外，對於未公開發行資料，使用者可根據自身研究領域，自行申請不同類型檔案。¹⁴

3.2 子資料庫的處理

3.2.1 費用檔

整個健保資料庫的核心在於費用檔。它紀錄了就診病患、醫師、疾病、時

¹³以2004年的檔案規模來說，健保資料檔共達 225GB，其中門診費用檔為 67GB，門診醫令檔為 141GB，住院費用檔和醫令檔分別為 2GB 及 15GB。

¹⁴若所需資料不在公開發行資料中，研究者可將樣本選取標準列出，透過國衛院特殊申請獲取資料。有關特殊申請的規定和費用請參閱下列網址 (http://www.nhri.org.tw/nhird/apply_04.htm)。

間、院所，以及費用的相關資訊。由於健保資料庫目的在於健保費用收付，故上述就診資訊的準確性相當高，研究者應可放心使用。為了使研究者能多了解些費用檔的處理技巧，我們下面依就診的人、事、時、地、和錢，一一介紹。

由於支付的關係，健保最詳盡，也最準確的應是錢的方面。費用檔不但明列了就診的總費用，還記錄了根據不同檢查、診療，處置等項目，各分類項目下的個別費用，¹⁵ 假使研究的目的是在於計算門診或住院支出，只需選取總費用即可，對於個別項目不需理會。然而，在計算健保費用時有幾點須留意：第一，健保資料所紀錄為院所申報費用，而非健保局最後核定給付，之間可能有3-5%左右的差距；¹⁶ 第二，所有申報費用均未經過物價平減，在做跨年比較時，應先將其物價平減，比較上才較客觀；最後，全民健保於1998年開始針對不同醫療服務陸續實施總額制度，¹⁷ 在總額制度之下，計算醫療費用須再乘以當季對應點值，才是正確的醫療費用。

在時間上，費用檔包括了幾個和日期相關的變數：就診日期、申報日期（門住診費用檔均有），出院日期與急慢性病床天數（住診費用檔特有）。申報日期是健保局支付使用，通常可以忽略。透過其他日期變數，研究者可以掌握就診時間與住院天數，並藉以此計算每個院所在一定時間內的服務量。有一點須提醒的是少數的住院資料並沒有紀錄出院日期，造成以出入院日期計算住院天數有所誤差，建議研究者以急性和慢性病床天數相加，所得住院天數數目較為精確。

在事的方面，費用檔記錄了病患就診原因，亦即疾病分類。研究者可透過病患疾病分類的幫助，選擇適當的研究樣本。歷年健保資料使用兩套疾病分類標準：國際疾病分類第九版的診斷代碼 (Diagnostic Codes of Inter-

¹⁵在門診上面，費用檔列出了藥品費、診療費，以及病患部分負擔費用；在住院方面，列出了診察費、病房費、膳食費、檢查費、放線性診療費、治療處置費、手術費、復健費、透析費、血液費、麻醉費、材料費、藥費、藥事服務費、注射費、精神科治療費、嬰兒費，以及病患部分負擔費用。一般而言，這些個別項目費用加總即為單次就診的總費用。然而，這個規則不適用於採固定給付的論病計酬項目（如剖腹產、人工關節置換）。

¹⁶健保局對於最初申報金額和最後核付金額的差距比例稱為核檢率。核檢率常因不同醫療服務項目，或不同健保分局轄區而有差距，一般而言，核檢率低於3%。

¹⁷健保局於1998年7月起針對牙醫，2000年7月起對中醫，2001年7月起針對西醫基層（俗稱小總額）制度，2002年7月起針對醫院（俗稱大總額）實施總額制度。

national Classification of Diseases Ninth Revision, 簡稱 ICD9 Diagnostic Codes) 和簡易疾病診斷碼 (A-Code)。¹⁸ ICD9 診斷代碼為國際通用疾病診斷碼, 分類較嚴謹; A-Code 為勞保時所使用疾病診斷碼, 分類較寬鬆。除早期門診費用紀錄仍沿用勞保 A-Code 標準外, 2000年起的門診費用檔和所有住院費用檔均採 ICD9 這個國際標準。提醒研究者在使用數年門診費用檔時, 留意這兩類疾病分類標準的差距, 避免因不同疾病分類標準造成選樣上的偏誤。

整個費用檔最特別的在於就診病患、醫師, 以及院所的代碼。費用檔記錄了每個病患、醫師, 以及院所的一致代碼, 這個代碼使研究者得以透過歷年健保資料, 建構一個病患的完整醫療史, 或是醫院、醫師的醫療處置史。以目前資料庫所涵蓋 8 年的健保歷史資料, 對從事醫療相關研究者有莫大助益。更特別的是, 這些代碼能和基本資料檔相串連, 使研究者掌握就診醫師、院所, 和病患的個別特性。

除了上述就診資訊外, 費用檔還包括了許多其他變數, 諸如住院費用檔的國際疾病分類處置代碼和給付類別,¹⁹ 和門診費用檔的特定治療項目等變數。這些變數大多反映病患就診時接受處置, 若能妥善運用能減少串連醫令檔困難。可惜的是, 由於這些變數在健保局規定上屬於選填欄位, 導致有相當比例並未如實記載, 直接使用可能會有問題。²⁰ 舉例來說, 門診費用檔中的特定治療項目, 雖將洗腎治療的腹膜透析和血液透析作進一步區分, 但紀錄上約有五分之二為空白。²¹ 建議研究者在使用這些變數前, 最好

¹⁸ICD 的疾病分類標準相當繁瑣複雜, 對於非醫學背景尤其艱奧。健保局網站上將一些常見疾病整理, 列出這些疾病的 ICD9 和 A-code, 對醫療研究者是個相當好的參考資料。疾病分類的網站地址為 (http://www.nhi.gov.tw/webdata/webdata.asp?menu=3&menu_id=56&webdata_id=1008)。

¹⁹處置代碼和診斷代碼均是國際疾病分類代碼的一部份。不同於診斷代碼記錄就診原因, 處置代碼記錄醫師處置。然而, 由於健保醫令檔已詳細記錄醫療處置, 且處置代碼在資料記錄上有相當缺失, 這個變數在健保資料的實用性不高。

²⁰門住診醫療服務點數的申報總表、點數清單、醫令清單的媒體格式中, 列出哪些變數為健保局規定醫療院所申報費用時的必填欄位, 通常這類變數準確性較高。有關媒體格式可參照健保局網頁 (http://www.nhi.gov.tw/webdata/webdata.asp?menu=3&menu_id=222&webdata_id=1175)。

²¹許續天·連賢明 (2006) 針對特定治療項目中有關血液透析和腹膜透析的記載, 利用抽樣歸人檔的門診醫令檔驗證這些紀錄的準確性, 有興趣的讀者可自行參閱該文附錄。

透過其他途徑確定其準確性。

3.2.2 醫令檔

相較於費用檔，醫令檔所紀錄的變數相對簡單。除了串連費用檔所須變數外，²² 就只有醫令類別、醫令代碼、醫令單價、醫令數量，以及總金額等變數。由於醫令代碼本身為一連串數字，研究者須透過「全民健保支付標準」來解讀任一特定醫令所對應的藥品，檢查和處置，和特殊材料。²³ 舉例而言，若想了解那些住院治療使用電腦斷層掃描 (Computed Tomography, 簡稱 CT) 和磁共振造影 (Magnetic Resonance Imaging, 簡稱 MRI)，研究者須先找出這些項目的醫令代碼，再利用住院醫令檔搜尋這些相關醫令；最後，將這些相關醫令串連費用檔找出所對應的住院費用紀錄。

雖說醫令檔變數較少，實務處理上並不輕鬆。原因有三：第一，「全民健保支付標準」相當龐大，沒有使用醫令檔實務經驗者，很容易在數千項檢查和處置，與數萬項藥品和特殊器材醫令中迷失；第二，即便醫療服務相同，但由於健保支付規定，不同層級供給者可能給予不同給付，也因此使用不同醫令；²⁴ 第三，即便醫療服務和層級供給者相同，健保局可能在不同年間使用不同醫令。²⁵ 筆者建議研究者使用健保局的快捷搜尋，透過關鍵字找尋相關醫令，以避免可能疏漏。²⁶

3.2.3 基本資料檔

整個健保資料庫最龐雜的應屬基本資料檔。由圖 1 可知，基本資料檔共包括了六個和醫療院所相關，兩個和就診醫師相關，以及兩個和就診病人相

²²包括費用年月、申報類別、醫事機構代碼、申報日期、案件分類，以及流水號等六個變數。

²³「全民健保支付標準」中處置和檢查醫令代碼為5碼，藥品和特殊材料分別為10及12碼。這些支付標準可由健保局網頁下載 (http://www.nhi.gov.tw/inquire/query2.asp?menu=3&menu_id=58)。

²⁴舉例來說，一般病床住院(天)的醫令在醫學中心為03001K(點數512)，區域醫院為03002A(點數456)，地區醫院和基層診所為03003B(點數395)。

²⁵舉例來說，磁共振造影(有顯影劑)在85年7月和93年6月間醫令為33085A(點數13000)，但自93年7月起增為33085A和33085B(點數11500)。

²⁶健保局快捷搜尋網頁網址為 (http://www.nhi.gov.tw/inquire/query1.asp?menu=3&menu_id=58)。

關的十個檔案。雖說就診院所相關的基本資料檔最多，筆者經驗是實用的僅有院所基本資料主檔 (HOSB) 以及病床檔 (BED) 兩個檔案，讀者可以自己驗證這個經驗的可靠性。²⁷

除了檔案眾多外，處理基本資料檔有兩個常遭遇的困難。首先，基於這些基本資料檔案對健保費用收付並無直接助益，健保局也因此不「刻意」維護資料的正確性，造成某些檔案有先天缺陷。這個現象在病床檔上最明顯，除了2004年外，病床檔歷年 (1997–2003) 資料從未更新，直接造成醫療院所的病床相關資訊偏誤。^{28,29}

另一個問題在於後天資料的編排繁雜。不同於費用檔或醫令檔，基本資料檔的編排採用「流水帳」方式記事，新舊資料雜陳；更麻煩的是，這些基本資料檔的儲存方式相當特殊。相較於費用檔或醫令檔顯示單次就診資訊，基本資料檔記錄的是某項變數所含資訊的有效期間。為了清楚顯示這個問題，圖2列出從院所基本主檔找出的一個實例。由這個例子可知，院所基本資料主檔所記載的並不是該院所於某時間下的特約類別，而是該類別的有效起迄日期。這個編排方式不但徒增整理難度，更不利於與費用檔的串連。

為了方便串連，筆者建議將基本資料檔從流水帳式、有效期間的編排，改為以代碼和時間為基礎的格式 (見圖2)。這個方式不但方便日後串連，更能有效降低串連可能錯誤。就以圖2這個例子來說，該院所的特約類別在4年中經過三次變動，若不先處理就逕自串連，很可能會連結到錯誤資訊，造成分析上的錯誤。由於院所和病患基本資料多半會隨時間改變 (如病床數增減或投保薪資變動)，建議處理為以代碼 (如院所或病患) 和時間 (如年) 為基礎的格式。相反的，醫師相關資料多半並不隨時間改變 (如醫師生日、專科醫師執照)，可將其轉成以醫師為單位格式即可。

²⁷和院所基本資料相關的其他四個檔案所含資訊多半可透過串連費用檔取得。以門、住診費用申請總表主檔 (CT, DT) 為例，這兩個檔案紀錄各院所門、住診服務總量，相關資訊在串連費用檔後也可取得。

²⁸健保資料庫中的病床主檔資料規模大小自1997年到2003年均相同。

²⁹文羽萃·江東亮 (2002) 將健保資料庫的病床檔中各院所的病床數與衛生署統計的醫事機構基料檔病床數目相較，發現其中有約20%的差異 (健保病床檔所得病床數為97,310床，衛生署所得病床數為121,483床)。

醫事機構代號	特約類別	特約類別起日	特約類別迄日	型態別	型態別起日	型態別迄日
12xxxxxxx	地區醫院	1997/7/1	1999/6/30	綜合醫院	1997/7/1	1999/6/30
12xxxxxxx	地區醫院	1998/9/1	2000/8/31	綜合醫院	1998/7/1	2000/6/30
12xxxxxxx	地區醫院	2000/7/1	2002/6/30	綜合醫院	2000/7/1	2002/6/30

↓

年	醫事機構代號	特約類別	評鑑等級
1997	1269850428	地區醫院	綜合醫院
1998	1269850428	區域醫院	綜合醫院
1999	1269850428	區域醫院	綜合醫院
2000	1269850428	地區醫院	綜合醫院

圖 2: 基本資料檔的轉換原則

3.3 各子資料間的串連

健保相關研究要豐富, 各子資料間的串連是一個不可或缺條件。有關各子資料庫串連所需變數, 已於健保資料庫譯碼簿中的「各檔案間串檔變項說明」圖中詳列, 本文不另討論, 僅就串檔過程提出兩點個人建議。首先, 在串連基本資料檔時, 研究者除依「各檔案間串檔變項說明」, 使用正確串連變數連結檔案外, 還須留意所串連資料是否「合時」。舉例而言, 當串連院所基本資料時, 除院所代碼外, 還需根據病患的就醫年月, 才能串連就醫院所的有效資訊, 這也是筆者建議將基本資料檔轉換格式的主因。然而, 由於資料本身問題, 筆者經驗是有極少數(約 1-2%) 就診資料無法正確串連, 由於有問題的樣本並不多, 應不會對整體分析造成影響。

其次, 在各子資料相互串連時, 研究者須妥善考慮串連後的資料規模, 最好於資料串連前先行處理, 避免串連後的資料過大。這項考量在串連醫令檔時特別重要。由於醫令檔規模龐大, 且單次就診通常包含許多醫令, 直接將醫令檔和費用檔串連無謂增加電腦負荷, 並造成日後分析困難, 故筆者建議僅保存分析所需醫令, 以降低串連後的資料規模。

3.4 串連非健保資料

由於健保資料的先天限制, 一些重要研究變數, 諸如院所所在鄉鎮的市場

狀況, 就診病患的社經變數, 以及病患治療後的健康狀態等均付之闕如。這些變數的缺乏不但限制了健保研究的潛力, 也使得研究者對健保資料的使用卻步。誠然, 這類缺陷多屬先天, 並非後天努力可完全彌補。然而, 筆者經驗是可透過串連其他外部資料來解決部分困難。為了說明上的方便, 我們將外部資料分為兩類, 一類和就診院所相關, 另一類和就診病患相關。

3.4.1 就診院所

在醫療研究中, 特別是醫療經濟領域, 醫療供給者的行為是否受到鄰近醫療需求或市場結構的影響, 是一個重要的研究議題。在院所基本資料主檔中, 僅記錄就診院所所在鄉鎮, 對院所所在地的醫療需求和市場結構的相關資訊相當缺乏。此時, 研究者可藉由一些外部資料的串連來彌補這些資訊。舉例來說, 研究者可將醫療院所所在鄉鎮和經緯度資料連結, 藉由鄉鎮經緯度, 計算該院所在一定距離內的院所、病床, 以及醫師數目, 藉以衡量該院所面對的市場結構; 在醫療需求上, 可透過內政部公布的《中華民國台閩地區人口統計年報》取得台灣各鄉鎮的戶籍人口總數, 各年齡或性別分組下的人口數和就業數, 這些數字反映各鄉鎮的人口結構, 可用於衡量該地區的醫療需求。值得注意的是, 有相當比例居民其實際居住鄉鎮和戶籍登記鄉鎮有差異, 單純使用戶籍人口資料可能會造成偏誤,³⁰ 建議讀者併用1990和2000年的戶口普查資料, 對居住和戶籍人口做校正。³¹

3.4.2 就診病患

研究者最迫切需要的外部資料, 莫過於就診病患的社經與生命資料。³² 由於健保資料中的病患身分代碼均已轉碼, 除非透過健保局的協助, 無法串

³⁰石曜堂等 (2003) 在國民健康調查訪查中發現約有17%的居民實際的居住鄉鎮和戶籍登記的鄉鎮不同。

³¹有關戶口普查資料的申請網址為 (<http://www.moi.gov.tw/stat/index.asp>); 《中華民國台閩地區人口統計年報》的相關網址為 (<http://www.dgbas.gov.tw/lp.asp?ctNode=3273&CtUnit=384&BaseDSD=7>)。

³²在國家衛生研究院每年為健保資料庫所舉行研討會上, 如何串連病患的生命資料, 一直是一個最常被提問的問題。

連外部資料。³³ 然而，筆者經驗是有許多重要病患社經資訊「混跡」於承保資料檔，有待有心人的發掘。舉例而言，就業人口和非就業人口的投保類別不同，研究者可藉由病患的投保類別推論其就業狀況；同時，承保資料檔中紀錄了投保人的投保金額，對已就業的投保人來說，特別是中低階的勞工而言，投保金額應和薪資所得關係密切；再者，對某些類別的投保人來說，投保單位所在的鄉鎮必須和戶籍所在的鄉鎮一致，透過這個關係，可以確認投保人的居住鄉鎮；最後，有些醫療服務是不可間斷的（如洗腎），透過病患的醫療利用情況，研究者可以「間接」推斷出病患的存活狀況。若能一步步的抽絲剝繭，雖說未能實際串連病患的外部資料，卻能解決許多研究上面對的挑戰。

4 實例操作

4.1 研究主題

經過前述原則說明後，許多讀者可能仍似懂非懂。本節透過筆者的一個研究實例 (Lien et al., 2006)，逐步說明如何將前述原則轉換為實際操作。這個研究主題，在於了解市場競爭對醫療費用和品質上的影響。這個議題在國外有許多討論，所得到的實證結果卻不一致。許多研究指出市場競爭，導致醫院從事「醫武競賽」(Medical Arm Race)，增加無謂的醫療費用，卻未能改善醫療品質 (Robinson and Luft, 1985, 1987)；相反的，也有研究指出市場競爭導致醫療院所的「良性」發展，降低醫療費用或增進醫療品質 (Dranove et al., 1992, 1993; Town and Vistnes, 2001)。

為了釐清市場競爭對醫療費用和品質的影響，此研究選擇1996–2002年間的中風住院病人，比較在不同競爭程度市場下的醫療費用和品質。在品質上，利用病人中風後1個月內和1年內的死亡率，作為短期與長期品質指標；在費用上，利用中風病人的當次住院費用和中風後1年內的門住診費用，作為短期與長期費用指標。此外，為了衡量「醫武競賽」的影響，進一步將中風病人的住院費用檔串連住院醫令檔，以瞭解有多少病例使用電腦斷層掃描和磁振造影等昂貴儀器，和這些檢查的費用。

³³蕭慶倫 (2003) 透過政府機構協助，串連戶籍檔以取得家戶組成和成員教育程度、串連所得稅核定與申報檔以取得家戶所得概況，以及串連死亡檔以取得生命資料。

除品質和費用指標外, 研究的另一個重點為市場競爭程度的衡量。在一般市場競爭相關文獻中, 最常見指標是賀氏指標 (Hirschman-Herfindahl index, 簡稱 HHI), 它為產品市場中各競爭者市佔率的平方總和。通常 HHI 值越高代表市場集中度越高, 競爭越弱化, 反之則競爭越激烈; 然而, 醫療服務相當複雜而特殊, 清楚界定某院所市場疆域並不容易。基於重症治療, 特別是中風治療講求時效性, 病患通常被送到一定距離內的醫療院所。基於此, 研究採用變動半徑法 (Variable-Radius Measure), 以院所涵蓋該院所 75% 中風病患的最大就醫距離, 作為疆域來界定市場。³⁴ 距離內醫院視為相關競爭者, 距離外則非相關競爭者, 再以市場中競爭院所的中風病患數, 來計算院所市佔率以及 HHI 指標。

圖3 列出該研究所使用資料。由圖 3 可知, 這個研究使用了六個健保資料檔和三個外部資料檔。六個健保資料檔分別是門診費用檔 (CD), 住院費用檔 (DD), 住院醫令檔 (DO), 院所相關的院所基本資料主檔 (HOSB), 病床檔 (BED), 以及承保資料檔 (ID)。研究首先透過費用檔搜尋新進中風病人和計算治療費用; 透過醫令檔串連 CT 和 MRI 紀錄; 透過病床檔和院所基本主檔串連病床數目和醫院特性; 並透過承保檔推估病患居住鄉鎮。三個外部資料檔包括以經緯度檔以計算各病患就醫距離; 以人口檔以計算各鄉鎮和一定距離內的人口結構; 最後, 以死亡檔衡量醫療品質。為了使讀者能清楚資料建構的細節, 將一步步過程在下面介紹。

4.2 研究步驟

4.2.1 透過住院費用檔, 選擇中風病人

研究的第一步是透過住院費用檔中的疾病診斷碼確定中風病患。中風病患依其病可分為兩類: 栓塞性中風 (ICD9 診斷代碼: 前三碼為 434; A-Code 為 A292 及 A293) 和出血性中風 (ICD9 診斷代碼: 前三碼為 430、431; A-Code 為 A291)。為了降低病患疾病嚴重差異程度, 樣本選擇新中風病人, 亦即病人於住院前 1 年內未曾有中風住院紀錄者。

³⁴除了變動半徑法, 還有行政區劃分法, 與固定半徑法 (Fixed-Radius Measure), 有關常用的市場劃分方式討論, 請見 Phibbs and Robinson (1993) 和 Gresenz et al. (2004)。Kessler and McClellan (2000) 認為這些方法都產生誤差, 建議採用估計機率來做為市場競爭指標。

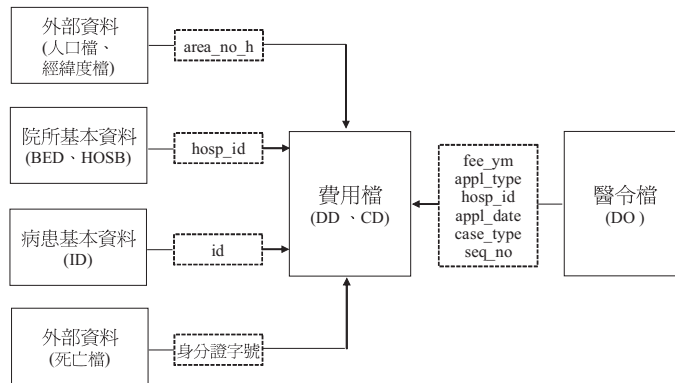


圖 3: 市場競爭對醫療費用和品質的影響 (以中風病人為例)

4.2.2 串聯院所資料主檔和病床檔, 取得醫院特性

在醫療照護中, 供給者扮演了相當重要角色。為了取得醫院特性 (如所在鄉鎮, 院所評鑑, 院所權屬) 和院所規模 (病床數), 將費用檔串聯病床檔和院所基本資料主檔這兩個資料檔。若先前已將這兩個資料檔處理為以院所和年月為基礎的格式, 此時僅需透過院所代碼和病患入院時間 (年、月) 即可串聯正確醫院特性。

4.2.3 並聯門診費用檔, 計算醫療費用

由於研究涵蓋時間為總額制度實施前, 醫療費用計算僅需經過物價平減, 不需再依總額點值調整。短期費用為當次住院費用, 通常即為當次住院所有費用;³⁵ 長期費用為自中風後1年內的門住診費用, 除了當次住院費用外, 還須將中風病人的住院日期與住院費用、門診費用檔相並聯, 才能計算中風1年內的醫療費用。

4.2.4 串聯住院醫令檔, 計算昂貴儀器費用

要計算住院 MRI 和 CT 使用和其費用, 須先透過健保局的快捷搜尋找尋

³⁵比較麻煩的是少數病患可能於就醫中轉院, 或單一住院中產生兩筆住院記錄, 造成計算上的困難。研究者可依病患住院日期和出院日期作整理, 若是單一住院有兩筆紀錄, 則可將其費用合併; 若分處於兩個院所, 則以花費較多的院所為主。

MRI 和 CT 的醫令代碼 (MRI 為 33084A、33084B、33085A、33085B; CT 為 33067B–33072B、33098B、36021C、37043C、37044C)。在住院醫令檔中搜尋這些代碼醫令, 並刪除其餘醫令, 然後再將這些醫令和費用檔相結合, 以找出住院相關費用。這樣可將串檔後的資料規模控制在可能範圍內。

4.2.5 串聯承保檔, 取得居住 (工作) 鄉鎮

整個研究最複雜的地方在於就醫距離的計算。計算就醫距離須同時知道病患的居住 (工作) 鄉鎮和醫院的所在鄉鎮。院所所在鄉鎮已於串連院所基本資料主檔取得, 比較麻煩的是病患的居住 (工作) 鄉鎮。由於健保資料並沒有直接記載病患居住鄉鎮, 僅能透過承保檔取得投保鄉鎮、投保類別、和稱謂代號這些投保資訊。雖說承保資料檔記載投保地點, 但所謂投保地點指的是投保人的投保鄉鎮, 並非被保人的居住鄉鎮。若被保人和投保人並未居住在同一鄉鎮, 則投保資訊不見得有用; 即便投保人和被保人共同居住, 基於健保投保的若干規定, 投保人的投保鄉鎮也不見得就是實際居住 (工作) 鄉鎮。

研究中採用三個條件來確定病患的居住 (工作) 鄉鎮。第一, 選擇被保人身份為投保人自己或配偶。這個條件可以確保投保人和被保人居住在同一鄉鎮; 第二, 選擇投保類別為私營企業受雇者, 地區人口, 以及農漁民。這些投保人需透過工作單位, 以及戶籍所在鄉鎮的戶政事務所或農漁會承保, 可以確保投保人的投保鄉鎮和工作或戶籍鄉鎮相符; 然而, 有相當比例的投保人並未居住於戶籍所在的鄉鎮, 為了確定被保人的投保鄉鎮即為病患的居住 (工作) 鄉鎮, 我們根據投保鄉鎮來計算就醫距離, 並將樣本限制為就醫距離低於 50 公里病患。一般而言, 中風後需儘速送到最近的醫院治療。以台灣現有的院所密度而言, 絕大部分病患的就醫距離均應遠低於 50 公里。這個條件可減少錯估病患居住鄉鎮所導致的偏誤。

4.2.6 串聯經緯度檔, 計算就醫距離和 HHI 指標

透過經緯度資料, 我們可將病患居住 (工作) 鄉鎮和就醫院所所在鄉鎮連結, 以計算每一中風病患就醫距離。在獲得了所有病患的就醫距離後, 則

以該院所其涵蓋 75% 中風病患的就醫距離範圍做為市場疆域，進一步計算市場內各院所的市佔率，和該院所 HHI 指標。

4.2.7 串聯人口檔，取得鄉鎮人口數

依據內政部戶籍資料，再輔以戶口普查資料，可計算出每一鄉鎮的人口結構，以用於衡量當地醫療需求指標。由於院所患者可能來自鄰近鄉鎮，分析時使用該院所一定距離下 (10 公里) 的鄰近居民人口結構，作為衡量指標。

4.2.8 串聯死亡檔，取得品質指標

要計算醫療品質，須將住院費用檔和死亡檔相串連，以計算病患 1 個月和 1 年的死亡率。由於健保費用檔中的病患代碼已轉碼，一般研究者無自行串聯。此部分串聯是透過健保局協助進行。

4.3 敘述統計

為了使讀者能瞭解樣本選擇的影響，表 3 紀錄了 1997 到 2001 年間不同樣本條件下的病患和醫院數目，以及病患和醫院特徵。1996 年樣本因無法確定新進中風病患而排除 (過去 1 年未曾中風)；相同的，2002 年病患也因未能計算 1 年內的死亡率而排除 (未來 1 年是否死亡)。

由表 3 清楚可知，目前台灣 1 年新進中風病患約為 3 到 4 萬人，這個數目隨著時間增加而呈現上升趨勢。平均而言，中風病患以男性為多，年齡約為 67.4 歲，其中 1/4 為出血性中風，其餘為栓塞性中風。病患中風後超過 2/3 是送到大型醫院醫治 (病床數 300 床以上)，僅有 10% 病患送到小型醫院 (病床數 100 床以下)。也因為如此，限制醫院層級至少為地區醫院，且該醫院於該年有 10 個以上中風病例，所排除醫院比例遠大於病患比例。以醫院數目來說，約有 20% 醫院因這個限制被排除，但僅約 5% 的病患受影響。比較大的衝擊在於確定病患的居住 (工作) 鄉鎮，約四成的病患在這個條件下被排除；剩下的樣本中，有 10% 因不符合就醫距離的條件遭排除。由於不同樣本擷取條件對樣本數目造成相當限制，最後的樣本數僅約原先一半。然而，在醫院和病患特性上，並未見到有大幅波動情況。

表 3: 中風研究樣本分析 (表中為樣本數目)

	醫院數			
	新中風病人具有完整醫院和病人資訊	... 於地區醫院以上, 且 1 年有 10 個中風病例以上	... 且有可靠居住鄉鎮資訊	... 且就醫距離為 50 公里內
1997	277	219	218	217
1998	268	208	208	208
1999	263	210	210	210
2000	265	216	216	216
2001	257	210	210	210

	病患數			
	新中風病人具有完整醫院和病人資訊	... 於地區醫院以上, 且 1 年有 10 個中風病例以上	... 且有可靠居住鄉鎮資訊	... 且就醫距離為 50 公里內
1997	32,610	32,278	17,333	15,214
1998	34,277	33,921	18,128	15,939
1999	35,794	35,484	48,879	16,554
2000	38,173	37,901	20,392	18,069
2001	40,097	39,743	22,107	19,567

	就診病患和醫院特徵			
	新進中風病患具有完整醫院和病人資訊	... 於地區醫院以上, 且 1 年有 10 個中風病例以上	... 且有可靠居住鄉鎮資訊	... 且就醫距離為 50 公里內
病人特徵				
出血性中風	25.7%	25.7%	25.1%	24.4%
男性	58.0%	58.0%	61.5%	61.9%
年紀	67.4	67.4	68.0	68.1
醫院特徵				
病床數 (0-100)	11.6%	11.0%	11.5%	12.1%
病床數 (100-300)	22.2%	22.3%	22.5%	24.5%
病床數 (300+)	66.1%	66.7%	65.0%	63.4%
私立醫院	28.1%	27.7%	29.4%	30.7%
法人醫院	47.2%	47.5%	45.8%	44.5%
公立醫院	24.7%	24.8%	24.7%	24.8%

表 4: 中風病人市場集中度衡量

	1997	1998	1999	2000	2001	1997- 2001
HHI (距離半徑: 75 百分比)						
中位數	0.160	0.163	0.170	0.174	0.199	21.1%
平均值	0.218	0.233	0.235	0.236	0.253	15.8%
	(0.183)	(0.176)	(0.187)	(0.183)	(0.189)	
HHI (三等均分)						
低集中度	66.94%	62.45%	58.05%	54.38%	49.61%	-25.9%
中集中度	19.32%	20.45%	26.01%	29.39%	32.99%	70.8%
高集中度	13.74%	17.10%	15.94%	16.23%	17.40%	26.6%

括號內為標準差。

表4記錄就診病患所在醫院的 HHI 指標。如前所述，這個指標是以容納75%病患的最大就醫距離作為該醫院市場疆界為計算基礎。就表4來看，HHI 指標不論是中位數或平均數，都出現逐年上揚的趨勢，反映出台灣醫療市場朝大型化和集中化的發展趨勢。為了進一步分析這些集中度指標變化情形，表4將歷年 HHI 分配 (1997-2001) 進一步區分為三組：低度集中、中度集中，以及高度集中。從這個劃分可知，中風病患於低度集中市場接受治療的比例正逐年下降，取而代之是中度以及高度集中市場。

表5將醫療費用與品質指標依不同集中度作區分。平均而言，集中度和病患醫療費用成反比：集中度愈高，就醫病患較低的短期費用與長期費用，但短期費用在不同集中度的差距反較長期來的明顯。在品質上面，長短期的趨勢並不一致，低集中度市場中病患有較高的短期死亡率，卻有較低的長期死亡率。這說明在醫療品質上，市場集中度的影響可能有長短期上的差別。

表5同時將病患和醫院特性依低度、中度和高度集中做區分。相較於位於市場集中度低的醫院，坐落於市場集中度高的醫院多半病床規模較小，評鑑等級較低 (多屬地區醫院)，顯示這些市場缺乏大型醫院。這個趨勢反映在鄰近人口數上更明顯：坐落於市場集中度高的醫院，有7成其鄰近人口在20萬以下；相反的，坐落於市場集中度低的醫院，其鄰近人口數有6成在20萬人以上，顯示出大型醫院偏好以中型或大型都市作為主要市場。

表 5: 中風病人依市場集中度區分之敘述統計

	市場集中度		
	高	中	低
支出及死亡率			
1 個月死亡率	0.102	0.107	0.123
1 年死亡率	0.234	0.230	0.227
中風住院支出 ^a	37,342 (58,794)	45,596 (68,330)	62,304 (83,853)
中風 1 年內支出	134,420 (210,764)	147,790 (230,940)	163,995 (234,732)
醫院特性			
權屬別			
私立醫院	36.5%	37.1%	26.5%
法人醫院	38.4%	32.4%	51.6%
公立醫院	25.1%	30.5%	21.9%
特約類別			
地區醫院	66.9%	45.0%	11.5%
區域醫院	29.1%	50.6%	38.7%
醫學中心	4.1%	4.4%	49.8%
規模			
病床數 (0-100)	24.4%	16.9%	6.6%
病床數 (100-300)	51.1%	35.1%	12.4%
病床數 (300+)	24.5%	48.0%	81.0%
病人特性			
鄰近人口 (/0000, 10km)			
(20-)	69.0%	49.5%	40.0%
(20-50)	15.2%	27.6%	26.5%
(50-200)	10.8%	15.0%	18.0%
(200+)	5.0%	7.9%	15.4%
人口特性和健康			
男性	61.1%	61.7%	62.2%
年齡 (< 50)	7.1%	7.6%	11.1%
年齡 (50-60)	10.2%	11.7%	15.0%
年齡 (60-70)	29.6%	29.5%	31.3%

續接下頁

承接上頁

	市場集中度		
	高	中	低
年齡 (70–80)	38.2%	37.4%	31.8%
年齡 (> 80)	14.9%	13.8%	10.8%
出血性中風	18.5%	20.3%	27.8%
前一年住院	0.5825 (1.160)	0.5757 (1.168)	0.4898 (0.977)
前一年住院支出 (‘0000)	1.8221 (7.982)	1.8314 (6.581)	1.6945 (5.999)
前一年門診支出 (‘0000)	2.6762 (5.396)	2.9269 (6.017)	3.0623 (6.307)
觀察值	49,282	22,271	13,790

括號內為標準誤。

a 支出為新台幣計，並經過消費者物價指數的平減 (2001年=100)。

在醫療相關研究上，如何有效控制病患健康狀況一直是一個研究難題。除了一般常用的性別和年紀外，很少有變數能反映病患的健康。此時，健保資料庫中使用一致病患代碼的優勢就突顯出來了。研究者可利用病患「過去」醫療利用，作為疾病嚴重程度的一個指標。表5底部列出病患過去1年內門診和住院的費用，以及曾經住院的次數；這些指標反映了病患過去的醫療利用，可幫助衡量病患中風前的疾病嚴重度。由表5可知，就診於市場集中度高的病患通常有較低的門診支出，以及較高的住院次數。這結果可能反應了醫療資源的可近性。病患鄰近位於醫療市場集中度低的地區，較容易利用醫療資源，因而有較高門診支出。而這些門診治療，有可能降低了住院機率。

5 討論與建議

自全民健保實施以來，如何控制逐年高漲的健保費用一直是衛生當局的最大難題。在高民衆滿意度的光環下，隱藏的是健保財務不斷挖東牆補西牆的困窘。也因此，健保局一而再、再而三的學習國外費用控制手段，期望能

透過這些措施，堵住千瘡百孔的健保漏洞，有效減緩財務失血情況。然而，各國醫療利用的歷史傳統大異其趣，保險制度也迥然不同，如何妥善運用實際經驗，設計可行的健保政策，以拉近理論和實際差距，成為衛生當局的一大挑戰。

本文嘗試從一個健保資料研究者的角度，介紹全民健保資料庫的優點和缺點，使用原則和操作技巧，以及可能和限制。由於健保資料規模龐大，檔案眾多，一一介紹常落入「見樹不見林」的問題。筆者根據各檔案特性，將健保資料區分為三大子資料：費用檔、醫令檔、以及基本資料檔；由於各子資料庫所涵蓋資料類別大不相同，處理技巧也有所差異。依據個人經驗，說明整理各子資料的處理原則，以及各子資料間如何相互連結來進行資料互補；再者，為了彌補健保資料的缺陷，筆者解釋如何利用外部資料來補充不足。最後，為使讀者瞭解處理原則和實際操作的差異，筆者提供一個研究實例，使研究者能按圖索驥，一步步組織成一個適合的研究資料。

一個筆者常被問到問題是這般複雜的健保資料是否值得長期投入。誠然，使用健保資料庫的進入障礙相當高，研究者通常需要花費許多時間和心力去試驗和熟悉。倘若不具備相當研究潛力，的確很難說服研究者長期投入。一般而言，研究可貴在於「創新」，而健保資料的可貴就它有一些獨特性，可回答一些使用國外資料庫所難以回答問題。首先，健保資料庫將醫療處置完整紀錄在醫令檔內，研究者可據此分析醫院或醫師採用不同治療的影響因子，以及這些不同治療的成效。這個優點在藥品研究上特別明顯。相較於國外許多藥品研究受限於精確的用藥資料，醫令檔提供了各種藥品名稱，藥品成分，給藥天數，以及藥品價格。研究者可利用這個優勢從事新舊藥間的替代行為比較 (Lichtenberg, 1996)，或是評估新藥的治療效果 (Duggan, 2005; Lichtenberg, 2001)。

第二，健保資料庫記錄了完整就診醫師資訊。相對於國外資料通常對醫師資訊嚴格保密，健保資料庫是少數具有能力研究醫師行為的資料。就以醫師的學習效果為例，學習效果是否存在一直是文獻上一個爭論焦點 (Birkmeyer et al., 1999; Epstein, 2002; Gowrisankaran et al., 2006; Ho, 2002; Luft et al., 1979, 1987)；其次，假使這些效果存在，這些效果是來自於學校教學 (school effect)，同儕切磋 (peer effect)，或從做中學 (learning by

doing), 文獻上亦沒有明確的答案 (Mary et al., 2006)。健保資料具有回答這些問題的潛力。

第三, 在過去10年中, 爲了控制健保費用的成長, 健保局推行一系列的支付制度變革。這些制度變革, 有些是參照國外政策 (如論病計酬), 有些是土洋參半 (如總額制度), 有些則是台灣特有 (如卓越計畫)。瞭解這些變革的政策效果, 不但能提供其他國家參考, 更是可作爲衛生當局修訂政策的重要依據。而歷年累積的健保資料正是釐清這些政策效果的最重要基礎。

除了上述三個方向外, 還有許多重要議題可藉由健保資料來分析。由於受限於個人研究領域, 無法針對所有可能議題一一介紹。在文章的最後, 筆者強調本文僅就個人所知部分, 作一番整理,³⁶ 若文中有錯誤或疏漏, 冀望各位先進見諒。

參考文獻

- 文羽萃·江東亮 (2002), “全民健康保險學術資料庫基本檔的應用經驗”, 《台灣衛誌》, 21(2), 150-155。
- 民生報 (2004a), “卓越計畫 是什麼碗糕?”, 《民生報》, 2004年10月16日。
- (2004b), “自主管理 卓越計畫 保障醫院收入”, 《民生報》, 2004年7月9日。
- 石曜堂·洪永泰·張新儀·劉仁沛·林惠生·張明正·張鳳琴·熊昭·吳聖良 (2003), “「國民健康訪問調查」之調查設計、內容、執行方式與樣本人口特性”, 《台灣衛誌》, 22, 419-430。
- 許績天·連賢明 (2006), “賺得越少, 洗得越多? — 台灣血液透析治療的誘發性需求探討”, 《經濟論文叢刊》, 已接受。
- 盧瑞芬·文羽萃 (2003), “臺灣醫療服務產業相關資料庫的比較分析”, 《調查與研究》, 14, 113-132。
- 蕭慶倫 (2003), “健康保險實施成效評估暨醫療服務品質測量政策分析”, 技術報告DOH90-NH-017, 中央健保局報告。

³⁶國衛院健保資料庫網頁上有針對各使用者問題作題庫整理, 有其他問題可參閱。有關問題網頁爲 (http://www.nhri.org.tw/nhird/talk_02.php)。

- Birkmeyer, J. D., Siewers, A. E., Finlayson, E. V., Stukel, T. A., Lucas, F. L., Batista, I., Welch, H. G., and Wennberg, D. E. (1999), "Hospital volume and surgical mortality in the United States", *New England Journal of Medicine*, 346, 1128–1137.
- Dranove, D., Shanley, M., and Simon, C. (1992), "Is hospital competition wasteful?", *RAND Journal of Economics*, 23(2), 247–262.
- Dranove, D., Shanley, M., and White, W. D. (1993), "Price and concentration in hospital markets: The switch from patient-driven to payer-driven competition", *Journal of Law and Economics*, 36(1), 179–204.
- Duggan, M. (2005), "Do new prescription drugs pay for themselves? The case of second-generation antipsychotics", *Journal of Health Economics*, 24(1), 1–31.
- Epstein, A. (2002), "Volume and outcome – it is time to move ahead", *The New England Journal of Medicine*, 346, 1161–1164.
- Evans, W. N., Levy, H., and Simon, K. I. (2000), "Data watch: Research data in health economics", *The Journal of Economic Perspectives*, 14(4), 203–216.
- Gowrisankaran, G., Ho, Vivian, and Town, R. J. (2006), "Causality and the volume-outcome relationship in surgery", Mimeo.
- Gresenz, C. R., Rogowski, J., and Escarce, J. J. (2004), "Updated variable-radius measures of hospital competition", *Health Services Research*, 39(2), 417–430.
- Ho, Vivian (2002), "Learning and the evolution of medical technologies: The diffusion of coronary angioplasty", *The Journal of Health Economics*, 21(5), 873–885.
- Kessler, D. P. and McClellan, M. B. (2000), "Is hospital competition socially wasteful?", *Quarterly Journal of Economics*, 115(2), 577–615.
- Lichtenberg, F. R. (1996), "Do new (and better) drugs keep people out of hospitals?", *American Economic Review Papers and Proceedings*, 384–388.
- (2001), "Are the benefits of newer drugs worth their cost? Evidence from the 1996 MEPS", *Health Affairs*, 20(5), 241–225.
- Lien, Chou, and Liu (2006), "The impact of hospital competition on treatment cost and quality of care: Evidence from stroke treatment in Taiwan", Mimeo.
- Luft, H. S., Bunker, J. P., and Enthoven, A. C. (1979), "Should operations be regionalized? The empirical relation between surgical volumes and mortality", *The New England Journal of Medicine*, 301, 1364–1369.

- Luft, H. S., Hunt, S. S., and Maerki, S. C. (1987), "The volume-outcome relationship: Practice makes perfect or selective referral patterns?", *Health Services Research*, 22, 157–182.
- Mary, A. B., Fournier, G. M., and Prasad, K. (2006), "Geographic variations in a model of physician treatment choice with social interactions", Mimeo.
- Phibbs, C. S. and Robinson, J. C. (1993), "A variable-radius measure of local hospital market structure", *Health Service Research*, 28(3), 313–324.
- Robinson, J. C. and Luft, H. S. (1985), "The impact of hospital market structure on patient volume, average length of stay, and the cost of care", *Journal of Health Economics*, 4(4), 333–356.
- (1987), "Competition and the cost of hospital care, 1972 to 1982", *JAMA*, 257(23), 3241–3245.
- Town, R. and Vistnes, G. (2001), "Hospital competition in HMO networks", *Journal of Health Economics*, 20(4), 3–53.

投稿日期: 2006年9月29日, 接受日期: 2006年12月30日

How to Use National Health Insurance Data to Conduct Health Economics Research

Hsien-Ming Lien

Assistant Professor, Department of Public Finance, National Cheng-Chi University

Since the inception of National Health Insurance in 1995, Taiwan has accumulated a decade of National Health Insurance Data (NHID). This data can assist the health authority to learn to help control the rising health spending. From a researcher's perspective, this paper discusses advantages and disadvantages, possibilities and restrictions, as well as principles and skills for using NHID. I first show the strength and weakness of NHID by comparing NHID with other important health data sets in Taiwan. Because NHID consists of many data files, I regroup these files into three sub-databases based on each file's specific role: expenditures, medical orders, and basic characteristics of patients, providers and physicians. Then I describe the rules concerning how to analyze each sub-database, how to merge files from two different sub-databases, and how to connect NHID with other non-health data. Finally, I use a research example to explain in detail the process of data construction, and discuss directions of future research.

Keywords: Taiwan national health insurance data, Health research data,
Universal health insurance

JEL classification: I10, I19