

健保資料庫與就業，薪資及家庭結構*

許績天

長庚大學醫管學系（所）

連賢明

國立政治大學財政學系（所）

全民健保自民國 84 年開辦後，至今已累積近十年的健保資料，雖說該資料庫具有完整醫療利用資料，卻缺乏保險人有關就業、所得、居住地、家庭結構等重要社經變數，進而限制了健保資料的應用和價值。本文利用主計處 2003-2005 年家庭收支調查樣本，將其串連其受訪者家庭健保承保資料，試圖透過家庭調查豐富的社經變數，來推論如何從承保檔建立保險人就業，薪資，居住鄉鎮、乃至家庭結構等社經變數。根據比較結果，我們發現（1）除職業工會外，絕大多數公保和勞保投保人的就業狀態和收支調查的訪查相同（2）若使用公勞保就業準確樣本，健保投保薪資可作為薪資代替變數（3）以投保鄉鎮來論，地區人口和農民有九成居住在投保鄉鎮（4）未成年兒童若經由父母投保，高達九成五和父或母同住。最後，我們提供一個研究實例，詳細解釋如何透過承保檔，來分析罹患乳癌對病患就業，以及薪資的影響。

關鍵字: 健保資料庫，家庭收支調查，就業，薪資

JEL 分類代號： I10，I19

* 許績天感謝國科會於研究期間的計畫補助（NSC 96-2416-H-182-007-MY2）；連賢明感謝 XXXXX 於研究期間的計畫補助；作者同時感謝行政院衛生署統計室提供相關資料庫使得本研究得以順利進行。所有文責均由作者自負。

1. 前言

全民健康保險開辦以後，2000 年中央健保局委託國家衛生研究院發行全民健康保險資料庫，至今已有九年的時間，期間累積了相當豐富的「醫療」相關資訊，由於資料含括所有醫療使用者與醫療提供者，歷年因健保而產生的所有醫療費用紀錄，因此該資料庫儼然成為世界上一個含括非常豐富且詳盡的縱貫性

(longitudinal) 醫療資料之一¹。此特性亦使該資料庫在學術研究上所扮演的角色越來越重要，事實上歷年利用全民健保資料庫所進行的研究及期刊發表數量亦日漸豐富²，且應用領域更含括實證醫學，公共衛生，流行病學，以及醫療政策，以及醫療經濟行為等領域之研究。

以社會科學的研究方法而言，不同社會經濟特性之健康及醫療利用始終為研究關心重點之一 (Feinstein, 1993; Marmot and Wilkison, 1999)。然而健保資料庫卻甚少在此議題的研究扮演著重要角色。究其原因，主要在於健保資料庫的設計乃以醫療提供者向健保局申報醫療費用為主，對於醫療使用者之經濟特性資訊不足，使得研究者即使有心想要透過健保資料庫來瞭解台灣不同社會經濟特性下健康與醫療利用情形，亦困難重重。根據連賢明 (2008)，健保資料庫的設計具有三個不足之處，其一，該資料以個別醫療照護使用者為單位，因此不包含家戶組成之相關資訊；其二，該資料僅有健保就診診斷資料，不包含自覺健康資訊，亦無家族成員疾病史。其三，健保資料僅包含醫療利用者之性別與年齡，其他重要的社會經濟特性 (如是否就業，與所得等)，乃至生命變數 (如是否死亡) 皆無直接記載³。因此，健保資料庫雖然已經累積了多年的資訊，但在許多社會科學所關注的議題上卻難有助益。

對於健保資料庫的不足，通常研究者一者可能放棄使用，轉而利用其他調查資料進行相關分析，如使用國民營養健康調查，國民健康調查等；或透過某些樣

¹ 由於資料庫詳細記錄所有健保醫療提供者與使用者之就醫記錄，而資料中使用相同病患，醫師與醫院代碼，因此研究者可以依據研究需要，而將資料整理成以醫療使用者 (病患) 為主之歷年就醫紀錄之縱貫性資料；亦可整理成以醫療提供者為主之歷年醫療院所相關行為之縱貫性資料。關於如何以健保資料庫進行研究，請參見連賢明 (2008)。

² 參見國家衛生研究院所記載之研究成果清單 (http://w3.nhri.org.tw/nhird/talk_07.htm)。根據統計，以健保資料庫進行研究之研究成果中，發表於期刊者，2004 年則有 11 篇，2006 年 22 篇，至 2008 年則有 46 篇，由 2002 年算起，至今總共有 110 篇。其中研究領域分散再如醫藥 (31 篇)，流病 (36 篇)，經濟與衛生政策相關 (43 篇)。

³ 有關健保資料庫與台灣其他相關健康資料庫之比較，請見連賢明 (2008)。

本取樣方法，在健保資料庫所包含的資訊中，嘗試「猜測」醫療利用者之相關經濟特性變數，如 Chouxxxx ()，YYYY (X) 等。就後者而言，不同研究者往往根據其使用資料庫經驗及學理判斷等，反推相關的經濟特性變數。這樣的作法雖然就其使用的樣本而言具有一定的可信度，然而其「猜測」是否正確卻始終無人可以確認。本文利用健保資料庫與主計處之家庭收支調查資料進行比對，說明利用健保資料庫以確認醫療利用者之相關社會經濟特性（包含就業與就業薪資），居住區域之認定方法與條件，同時確認醫療利用者（未成年兒童）之家庭依附關係的確認方法及條件。希冀透過兩資料庫的比對，以瞭解當研究者欲直接使用健保資料庫以「猜測」醫療使用者之相關經濟特性變數時，其可行之條件與正確性為何。

根據我們的比較分析，本文發現：（1）一般較穩定的工作，包括扣除地方公職人員之公保（11A，11B，11D，11E，12F）被保險者，參加勞保之公營事業機構工員（12G），民營事業機構受雇者（12H）、中央非事業機關、國立大專學校，省（市）以下政府機關、學校，以及私立學校之工員（13I，13J，13K）等，健保資料庫確實在其就業狀態之判斷正確性較高。然其中只有私立大專與中小學教職員（11D 與 11E），不論在就業狀態與無就業狀態之正確性皆高達 97% 以上。

（2）以就業狀態較為正確的投保類別之投保者來看，健保的投保薪資某個程度上確實可以作為代替其所得之變數。整體而言，投保薪資普遍低於調查所得，然而差距大多低於 2 標準差以內（96% 以上）。以及（3）以投保者之投保機構所在地來反推投保者之居住鄉鎮市區，僅有地區人口（62），農民（31Q），與地方公職人員（11C）有較高（95% 以上）之正確性。（4）未成年兒童其投保人為父或母時，其與父或母同時居住一起之比率高達 97% 以上，如果僅看 12-15 歲之透過父或母投保之兒童，其與父或母同居比率更高達 99.66%。然而，這些未成年兒童與其雙親同時居住一處之比率僅有 84% 左右；（5）與雙親同居之未成年兒童，當家庭中有兩至三名兒童時，其所有兒童以單一投保人投保者比率為 97% 以上，即使有三名以上，其採單一投保人投保之比率亦高達九成。而這些單一投保者，通常皆為父母之薪水較低者（比率高達 95% 以上）。

健保資料庫在各社會科學領域之研究已經有非常多年的歷史，也累積許多發表文獻，雖然許多文獻，嘗試透過各種樣本選樣邏輯與方法，以獲得醫療利用者之相關社會經濟變數，然而這些文章之方法皆未有明確與嚴謹的確認。本文為國

內第一篇有系統且嚴謹的比對利用健保資料庫以獲取相關經濟社會變數之方法之正確性之文獻。本文各節大綱如下。第二節說明用以進行比對分析之健保資料庫與家庭收支調查，其整合方法與整合後資料說明。第三節，則分別針對投保者之就業（與無就業）狀態，薪資，與居住地點，以及未成年兒童之家庭依附關係進行兩資料庫比較。第四節透過一個研究實例，使讀者能進一步將前述確認不同社會經濟特性之方法與實務操作結合。最後一節總結，並提供一些研究方向供讀者參考。