

如何使用健保資料推估社經變數*

連賢明**

國立政治大學財政學系教授

近年來使用健保資料庫相關研究日益增加；然而，由於健保資料缺乏病患社經變數，阻礙健保資料在社會學科應用。在主管機關協助下，本文以 2005 年家庭收支調查家戶為樣本，串聯健保承保資料，利用家庭收支調查所記錄豐富社經變數，驗證使用承保資料推估病患戶籍縣市、全職就業、薪資所得以及兒童家庭結構等變數準確性。比較結果發現：(1)使用投保類別為地區人口，農民，與地方公職人員，且投保人和被保險人相同者，其投保單位縣市和戶籍所在縣市相同者達九成五；(2)使用投保人和被保險人相同樣本，在公、勞保投保人的就業準確度相當高，平均達九成以上；(3)使用就業準確樣本推估薪資所得，公保投保者其投保薪資為實際薪資 9 成，勞保民營受僱者投保薪資最低 1/3 投保人，其投保薪資較調查薪資低估 3 成，最高 1/3 投保人投保薪資高約 6%；(4)十五歲以下兒童投保人為父或母時，與投保人同居比率高達 93%；(5)使用承保資料推估十五歲以下手足數目，準確度達九成以上。最後，我們提供一個研究案例，利用承保資料來分析乳癌對就業婦女存活、就業和薪資的衝擊。

關鍵字：健保資料、就業、薪資、家庭結構

* 作者感謝衛生署統計室黃旭民主任、李品青專員、徐俊強專員和林雅惠小姐在家庭收支調查和健保承保檔資料串聯的協助，助理許鈴宜在資料處理的優異表現；也謝謝三位評審人的建議、韓幸紋、許績天和陳麗光老師，以及中研院調查研究中心演講參與者的評論；所有文責均由作者自負。

** 國立政治大學財政系（所）教授，email: hmlie@nccu.edu.tw；作者感謝國科會於研究期間的計畫補助（NSC 97-2410-H-004-009-MY2）

收稿日期：99 年 5 月 21 日；接受刊登日期：99 年 9 月 16 日

壹、前言

自全民健保開辦後，在保險收付的需要下，中央健康保險局（以下簡稱健保局）累積了可觀的醫療利用資料。基於健保政策分析和學術研究需要，在不傷害個人隱私的前提下，健保局於 2000 年起委由國家衛生研究院，將歷年健保資料轉換成全民健康保險資料庫。該資料庫能有效將病患醫療利用連結醫院、醫師和病患基本資料，¹ 滿足不同領域在醫療研究上的需求。近年來使用該資料庫所進行研究數量與日俱增，² 研究領域也亦趨廣泛，涵蓋實證醫學、公共衛生、流行病學和醫療管理等領域。

相較於這些領域，健保資料庫較少應用於社會科學領域。這固然是因為社會學科中從事醫療相關的學者較少；但更重要的是，在以保險收支為基礎所建構的健保資料庫中，有關醫療使用者（即病患）的變數收集相當有限。這個變數的不足主要反映在幾個方面，第一，該資料庫以個人為單位，無法得知病患家戶組成；第二，該資料庫僅記錄病患就醫地點，無法得知實際居住地；最後，也是最重要的，該資料庫在個人特性上僅紀錄病患性別與年齡，其他重要社經特性（如教育、就業、所得等）皆無清楚記載（詳見連賢明，2008）。這些社經變數缺乏，導致使用健保資料在社會學科的門檻提高，進而阻礙該資料庫的廣泛使用。

在這些先天限制下，許多研究者改而使用調查資料（如國民營養健康調查、國民健康調查等）來取得個人社經變數。這個方法雖然彌補了健保資料的先天限制，卻同時放棄了健保資料庫中龐大且精確的醫療利用記錄。但由

1 健保資料庫保存詳細健保就醫記錄，且資料中病患、醫師與醫院的代碼一致，研究者可基於個人研究需要，將資料整理成以醫療使用者（病患）或醫療提供者（醫師或醫院）為主的縱貫性資料。有關如何使用健保資料庫進行研究，請參見連賢明（2008）。

2 根據國衛院統計，以健保資料庫進行研究之研究成果發表於期刊者，由 2002 年逐年增加（2004 年 11 篇，2006 年 22 篇，2008 年 46 篇），至今共有 110 篇。其中醫藥領域（31 篇），流行病學（36 篇），經濟與衛生政策（43 篇），但最後一類多為醫管領域，較少為社會學科相關學科。相關資料請參見國家衛生研究院（2008）健保資料研究成果清單（http://w3.nhri.org.tw/nhird/talk_07.htm）。

於疾病發生率普遍不高，使用這些資料庫很難分析單一疾病。舉例來說，根據 2007 年台灣癌症登記檔，全台共有 75,769 人罹患癌症，換算疾病粗發生率為每十萬名中有 330 名。假使研究者使用 2005 年國民健康調查（樣本數 30,680），則僅能找出約百名癌症新增患者；若嘗試進一步侷限於乳癌，則僅能找出約十多名乳癌患者（乳癌粗發生率約十萬分之三十），使用其他資料庫對罹病樣本大小影響，可見一般。

為有效使用健保資料，近年來一些研究者開始利用健保承保檔所記錄變數來「推估」病患個人變數。這個方法主要是利用健保資料庫中龐大投保人口，以「切香腸」方式先行剔除有問題樣本，再從可用樣本的承保資料來推論病患生命變數、社經變數、家戶組成乃至居住地點。如許績天和連賢明（2007）使用洗腎病患的醫療利用記錄，推論洗腎病患的存活時間；Lien, Chou, and Liu（2008; 2009）使用心臟病和中風病患手術開刀後一年內的退保紀錄推論這些病患的術後死亡狀態；Lien, Chou, and Liu（2008）使用地區人口和農保投保單位鄉鎮，來推論居住地鄉鎮；魏郁純（2006）使用公保，勞保病患的投保記錄，推論這些病患的就業狀態；韓幸紋、連賢明（2008）更利用兒童絕大部分透過父母投保，且投保於單一保險人（父或母）名下，來推論家戶中子女數目和順序。這些「推論」雖立基於合理行為假設，也和一般現實狀況大致相符，但由於缺乏實際資料佐證，推論和現實差距始終無法確認，容易導致研究爭議。

在主管機關協助下，本文以 2005 年主計處家庭收支調查樣本（以下簡稱家支調查），串聯健保資料庫中的承保資料檔（以下簡稱承保檔）取得這些受訪者當年度承保資料。由於家庭收支調查記錄豐富的個人社經變數、家戶組成乃至居住地點資訊，這兩個資料庫的結合，使我們能透過實際資料來討論透過承保檔取得這些社經變數的可能性。如此一來，不但能化解研究者對這些推論的質疑，更藉由實際資料佐證這些變數準確性，大幅增加使用健保資料進行社會學科研究的可能性。

根據這兩個資料庫比較結果，本文發現：(1)以投保單位縣市來推論投保人戶籍縣市，在地區人口，農民與地方公職人員投保類別中準確性高達 95%；(2)由於健保規定就業者須不可依附他人投保，選用公、勞保投保人中

被保險人和投保人代碼相同樣本，所認定就業狀態準確度相當高，平均在九成以上；(3)在就業準確樣本中利用投保薪資來推論個人薪資所得，在公保樣本平均低估約 10%，勞保民營受雇者的差異相當大：投保薪資最低的 1/3 投保人，和其調查薪資相比低估約 3 成，薪資最高 1/3 投保人其投保薪資高於實際薪資 6%；(4)兒童投保人為父或母時，與父或母同居比率為 92-94%，但同時和雙親同居則僅約 80-82% 左右；(5)與雙親同居的未成年兒童中，若家庭中有兩兒童以上時，將所有兒童投保於單一投保人比率為約九成以上。

本文各節大綱如下。第貳節先說明用來比對的家支調查與健保承保檔，以及合併後樣本基本特性；第參節則說明如何從承保資料推估投保者戶籍縣市、全職就業、薪資所得以及兒童家庭結構，並使用家支調查資料來確定這些推估的準確性；第肆節分析乳癌對職業婦女在存活、就業和薪資的影響，透過這一個研究實例，使讀者能將前述認定方法和實務操作結合。最後一節總結，並提供一些研究方向供讀者參考。

貳、資料說明

本節說明本文所使用兩大資料庫，分別是主計處的家庭收支調查（2005）和國家衛生研究院發行的健保承保資料檔（2005），並以 2005 年家支調查樣本串聯承保資料檔後樣本特性。

一、家庭收支調查

家支調查為行政院主計處每年針對居住台灣（非離島）本國籍國民所組成家戶，採隨機抽樣方式，針對家戶收入與消費支出所進行訪查。³ 該調查是台灣現有時間最長，⁴ 也是最權威的家戶所得和消費資料。訪查的項目包

3 家支調查採分層二段隨機抽樣方法，其中第一抽樣單位為村里，第二抽樣單位為該村里內家戶。抽樣過程中，依據各村里農、林、漁牧、與礦業之就業人口比率，先將村里區分為都市、城鎮、與鄉村，各層抽樣樣本戶數則以層內母體戶籍登記戶數，佔該區域母體戶籍登記戶數比例配置。

4 家支調查最早始於 1964 年，每兩年調查一次，但從民國 1970 年起，改為每年調查一次，

含四大類：家庭設備與住宅概況，家庭收支（經常性收入，非消費支出與消費支出），和家庭戶口組成（戶內成員的居住地區、性別、年紀、就業、教育等，以及各成員與戶長關係等）。

本文使用 2005 年家支調查樣本，取得這些受訪者確實的就業狀態、個人薪資、以及兒童家庭結構。由於家支調查清楚記載家戶內成員和戶長關係，是否共同生活，以及就業狀態等，我們可明確判斷該成員是否就業，以及兒童和父母依附關係；就居住地區而言，由於家支調查根據國人戶籍資料進行抽樣，問卷中清楚記錄了戶籍縣市。

比較複雜在於個人薪資部分。由於家支調查的一個主要目的在於了解家庭經常性收入來源，因此它記錄了相當詳盡的所得資料，包括受雇人員報酬、產業主所得、財產收入、租金收入、移轉收入和其他雜項收入；而在受雇人員報酬中，還另外包含本薪、兼業薪資以及加班費等收入，可說鉅細靡遺。基於本文主要是以家支調查和承保檔作對比，在投保薪資上，我們僅使用受雇人員報酬的本業薪資和承保檔中投保薪資比較。

二、健保承保資料檔

自全民健保實施後，除 1995 年和 1996 年部分外，⁵ 國衛院每年均將健保利用資料整理發行。由於母體資料龐大，國衛院將健保資料切割為三大部分：費用檔（住院和門診），醫令檔（住院和門診），以及基本資料檔。其中費用檔收錄每次健保就醫後，醫療機構所產生醫療費用概要；醫令檔則記錄就醫中醫師所使用檢查與處置；基本資料檔則包含病患的投保資料，以及就診院所和醫師基本特性。這些資料各自有不同功能，可彼此連結，串聯成有效的醫療利用資料。⁶

本文使用的健保資料為承保檔。該檔記錄三類資訊：第一類則為被保險

目前是由主計處中部辦公室、臺北市政府主計處及高雄市政府主計處負責辦理臺灣地區家庭收支調查計畫。

5 健保資料庫中紀錄 1996 年資料包括門診費用申請總表主檔，住院費用申請總表主檔，門診處方及治療明細檔，住院醫療費用清單明細檔，以及特約藥局處方及調劑明細檔。

6 有關健保資料庫檔案的處理原則，以及各子資料檔間如何串聯，請見連賢明（2008）第三節。

人（即病患）資訊，包括被保險人身份代碼、性別、年紀、投保身份（以被保險人納保或依附眷屬投保等）和稱謂代號。由於健保採強制納保，幾乎所有台灣國民均為被保險人；投保身分主要區分是否以被保險人身份納保或依附他人納保。假使被保險人透過他人納保，其稱謂代號會記錄被保險人和投保人間關係（如配偶、父母、子女、孫子女）。

第二類則為投保人資訊，包含投保人代號，投保金額，投保單位屬性（如公立機構員工，地區保人口等）及投保單位所在地。所謂投保人指的是保險費用的支付者，假使是被保險人為自己投保，投保人身份代碼和被保險人身份代碼兩者相同；相反的，若依附他人投保，兩個身份代碼會不同。舉例而言，未成年兒童通常透過父或母投保，因此，其投保身分為眷屬依附，稱謂代號會是子女，投保人代碼所記錄為父或母身份代碼。

除身分代碼外，投保人資訊還包括投保金額，投保類別，以及投保單位所在地。所謂保險類別指的是投保人投保機構的單位屬性，而單位所在地則是投保機構所在鄉鎮。根據目前健保設計，投保類別共有六類十四目，⁷不同投保類別在投保金額計算上不同，保費負擔比例也不相同。舉例來說，政大教師的投保單位所在地為台北市文山區，投保類別則為中央非事業機關公務人員（代碼 11A），至於投保金額在 2005 年是以全部薪資所得的 87.04% 計算；⁸若是以月薪十萬元的教授來說，每月健保保費約四千元，其中教師個人負擔 30%，其餘 70% 由雇主（即政府）負擔；相較起來，勞保則是以全部薪資所得作為計算基礎，而在保費分攤上，勞工佔三成，雇主佔六成，其餘一成由政府負擔。

最後一類資訊則是投保人加退保的原因和日期。依據健保承保規定，當

7 健保投保類別六大類中，第一類主要為公保和勞保（不含職業工會）投保人員；第二類為勞保和職業工會投保；第三類為農、漁、水利會會員；第四類為軍校生、替代役，無依軍眷和家屬；第五類為低收入戶成員；第六類為榮民、榮譽以及其他未涵蓋家戶投保人。有關詳細分類請參閱中央健康保險局網站（http://www.nhi.gov.tw/webdata/webdata.aspx?menu=18&menu_id=678&webdata_id=3436&WD_ID=722）（2009 年）。

8 根據全民健康保險法施行細則第 70 條之 1 第 2 項第 1 款，具有公教人員保險或軍人保險被保險人資格者，其投保金額 2005 年為 87.04% 乘以其俸（薪）給總額。2007 年 8 月則將該比例調整為 90.67%。

被保險人或投保人資訊變動時，會產生一次加、退保記錄。舉例來說，倘若投保人失業，導致被保險人轉換依附對象（如從父轉為母），則會產生加、退保情況；其他諸如投保人轉換工作、投保薪資變動，甚至工作地點改變，均會在投保資料檔產生加、退保情形。爲了確保資料串聯時能選擇正確承保資訊，建議將承保檔先行整理成以時間（年或月）爲基礎的資料格式，⁹ 此法不但能降低操作錯誤可能，也可方便日後將承保資料和其他資料串聯。

三、兩資料庫串聯後樣本特性

表一列出成功串聯當年度承保檔的 2005 家支調查家戶樣本基本特性。根據表 1，合併後共有 13,413 個家戶或 43,233 位受訪者，佔當年家支調查總家戶的 98.4%，顯示兩資料庫的串聯狀況良好。而在這些樣本中，男女樣本比例相當，受訪者年紀在 15 歲以下約佔兩成，16-40 歲及 41-65 歲者分別佔 33.28% 及 33.88%，而 65 歲以上有 13.08%。

表 1：2005 年家支調查與承保檔串連後樣本特性

平均數（百分比）			
性別		已婚	59.32%
男性	49.76%	教育程度	
年齡		國中以下	52.37%
15 歲以下	19.75%	高中	24.56%
16 歲—40 歲	33.28%	大學	21.12%
41 歲—65 歲	33.88%	研究所及以上	1.95%
65 歲以上	13.08%	居住地區	
就業	43.71%	北部	41.28%
年所得（萬元）		中部	19.98%
全部樣本	18.70	南部	31.21%
就業樣本	42.77	東部	6.23%
串聯戶數	13,413	串聯樣本數	43,233
全部戶數	13,637	全部樣本數	45,672

9 有關整理基本資料檔的基本原則，請參考連賢明（2008）。

在這些受訪樣本中，約六成爲已婚，教育程度則以國中以下佔一半以上，其次爲高中（24.56%），大學（21.12%），僅有 1.95% 爲研究所與研究所以上；43.71% 受訪者有工作，而其平均年度薪資所得爲 42.77 萬元。最後，這些樣本中約四成居住北部，三成住南部，兩成住中部，其餘來自於東部。

參、比較結果說明

本節使用家支調查和承保檔串聯後樣本，驗證採用承保檔推估戶籍縣市、全職就業、薪資所得、和兒童家庭結構等變數的可行性。我們先說明推估這些變數的原因與條件，再以串聯資料確認這些選樣條件的準確性。

一、戶籍縣市

在醫療經濟研究中，病患的居住地點是一個相當重要變數。居住地點不但會影響就醫選擇，也是計算就醫者的就醫距離（從居住地到就醫地）的關鍵變數，而就醫距離則是醫療經濟研究中不可或缺的工具變數（McClellan et al., 1994: 859-866）。可惜的是，健保資料中並沒有記錄病患居住地，惟一和居住地相關變數是「投保單位鄉鎮」。在現有投保規定中，地區人口投保人需透過戶籍鄉鎮區公所來投保；農、漁民、水利會員則需透過所屬或戶籍所在農、漁、水利會投保，這些人的投保單位鄉鎮理當和投保人戶籍鄉鎮應當一致，或可考慮以戶籍鄉鎮來作爲居住地的替代變數。

爲了瞭解投保鄉鎮和戶籍地的差距，我們選擇投保人年紀在 20 歲以上，且被保險人和投保人具有相同代碼。這些樣本透過自身投保，其居住地和投保機關所在地應高度相關。由於各個保險類別投保規定不一，表 2 依投保類別進一步區分，基於家支調查僅列出家戶樣本戶籍縣市，兩者比較時以縣市作單位。根據表 2 第一欄結果，公保或勞保投保人其投保機關縣市和個人戶籍縣市確實存在相當誤差。以公保人口而言，除地方公職人員外，僅約六成（58.61%）的投保者其投保縣市與戶籍縣市一致；勞保比例則更低，顯示這些樣本的戶籍縣市和投保機關縣市具有差距。根據公、勞保投保規定，這些投保人需透過工作公司、工廠投保，按理說投保單位縣市和戶籍縣市應相去

表 2：投保人投保縣市與全職就業精準性比較^a

家支調查承保資料	戶籍縣市 ^b	全職就業 ^c	非全職就業 ^c
	戶籍縣市 投保縣市	全職 全職	非全職 非全職
公保 (11A-E, 12F)	58.61% (786)	94.47% (1,266)	86.73% (565)
中央公教、公職人員 (11A)	41.61% (129)	98.99% (297)	88.05% (159)
非中央公教人員 (11B)	67.53% (470)	96.19% (682)	82.84% (268)
地方公職人員 (11C)	100.00% (105)	47.06% (68)	88.64% (44)
私立大專教職員 (11D)	24.49% (12)	100.00% (49)	82.35% (17)
私立中小教職員 (11E)	66.67% (22)	100.00% (33)	100.00% (10)
公營事業職員 (公保)(12F)	32.43% (48)	96.35% (137)	97.01% (67)
勞保 (12G,12H,13I,13J,13K,21)	55.87% (8,324)	77.50% (14,695)	79.73% (3,809)
公營事業工具 (勞保)(12G)	38.06% (94)	92.56% (242)	84.91% (53)
民營事業受雇者 (12H)	49.58% (4,509)	88.76% (8,922)	79.70% (2,310)
公務機關、學校工具 (13I,13J,13K)	76.69% (375)	94.45% (577)	86.73% (211)
職業工會會員 (21)	66.01% (3,346)	54.50% (4,954)	78.38% (1,235)
農會會員 (31Q)	95.17% (3,885)	50.05% (2,144)	80.52% (611)
水利，漁會會員 (31R,32)	83.44% (544)	53.17% (617)	69.62% (158)
地區人口 (62)	95.17% (3,432)	51.35% (2,405)	80.78% (1,025)
樣本數	16,971	21,125	6,168

a：所有樣本使用受訪者年紀介於 18-65 歲，且投保人身份代碼和被保險人身份代碼一致。括弧內為該投保類別樣本數。

b：承保資料中投保縣市為投保人投保單位縣市，家支調查戶籍縣市為受訪者戶籍地。

c：承保資料全職就業係指投保者自行投保，反之為無全職就業（含失業、兼職等）；家支調查全職就業為受訪者有就業且薪資所得為正，反之為無全職就業（含失業、兼職等）。

不遠。但實際上以工作所在縣市，推估戶籍縣市或居住縣市往往失真，主因是許多民營機構受雇者可能工作於某公司工廠，但投保於公司總部。舉例來說，許多公司總部設於台北市，員工工作地點卻散布全省，以台北市來推估所有員工戶籍縣市，自然會產生相當差距。

相較之下，農民、漁民、地區人口等保險類別，其戶籍縣市與投保縣市高度一致，特別是農民和地區人口，其準確性高於95%以上。這和前述推論一致：這些投保人需於戶籍所在鄉鎮投保，自然和戶籍縣市幾乎雷同。因此，研究者若能將樣本侷限於農民和地區人口等二類投保人口，應能取得可靠的戶籍地資料；但即使如此，仍有17%居民其戶籍地和居住地不一致（石曜堂等，2003）。研究者可參考Lien, Chou, and Liu（2008）一文，先利用病患投保鄉鎮和就診院所鄉鎮所計算就醫距離，再除去就醫距離超過一定範圍外樣本來降低可能誤差。

二、全職就業

在諸多社經變數中，一個重要變數是投保人的就業狀態。依據目前承保規定，當投保人有工作時，須自行透過雇主投保，不得由他人代保。因此，就業者的投保人身份代碼和被保人身份代碼理當一致；而投保類別，則依因雇主型態而有所不同。根據這項納保規定，可推論全職就業者的投保人和被保人身份代碼相同，從而於承保檔中認定就業；然而，此原則並非在每個保險類別中均成立。最明顯例子為地區人口，這些人口有相當比例為無業者，依規定卻需透過戶籍地戶政事務所加入健保，這類保險人和被保險人身份代碼因而相同；另外，農民，漁民，和無雇主工作者（如計程車司機），則分別透過所屬或戶籍地農會，漁會，和職業工會身份投保，^{10,11} 這些投保人工作性質特殊，其全職就業狀態不易清楚界定，以此原則來推論就業狀

10 根據健保投保規定，農會或水利會會員或年滿十五歲以上實際從事農業工作者（第三類第一目），無一定雇主或自營作業而參加漁會為甲類會員，或年滿十五歲以上實際從事漁業工作者（第三類第二目），此類投保人需透過所屬或戶籍農、漁、水利會投保。

11 職業工會投保人為無一定雇主或自營作業而參加職業工會者。

態也會產生誤差。¹² 也因此，現有研究多半以公保，勞保（不含職業工會）樣本來推論投保人的就業（魏郁純，2006）。

但使用公保，勞保（不含職業工會）投保人來推論就業是否正確呢？這仍是一個待驗證的問題。爲了瞭解承保檔所認定全職就業和實際就業差距，我們首先將投保人年齡限制於 18 至 65 歲，這和主計處所認定勞動人口相當；再則，我們從承保檔中選擇投保人和被保險人身份代碼相同者，透過其所申聯家支調查來判定是否就業且是否有薪資所得。¹³ 由於不同保險類別下所認定準確性存在差距，我們也依不同保險類別作區分，並將其比較結果列於表 2。

表 2 第二大欄列出承保檔所認定全職就業樣本，所對應家支調查的就業狀態準確性。¹⁴ 根據表二結果，就業正確率約可略區分三類：公保，勞保和其他。在這三類中，正確率最高是公保（保險類別代碼 11A-E、12F），平均就業資訊正確率爲 94.47%。事實上，若除去正確率最低的地方公職人員（47.06%），公保納保人口就業準確性在 96% 以上。其次爲勞保人口（保險類別代碼 12H，12G，13I，13J，13K，21），這些納保人口除職業工會外（21），其正確率介於 88%~96% 之間，而公營機關以勞保納保者（12G，13I，13J，13K），又比民營機構受雇者（12H）準確性來的高；最後一類爲其他人口，這類人口的所認定就業正確率最低，包括農民（50.05%），地區人口（51.35%），水利會或漁會會員（53.17%）。這些保險類別或因其投保規定自行納保和就業不必相關（地區人口），或是就業認定較複雜（農、漁民），顯示出保險類別變數在協助認定全職就業的重要性。

表 2 第三大欄則以承保檔中所認定非全職就業樣本，和其家支調查所認定非就業（失業、就業但無薪資所得或兼職）比較。雖然第三大欄中也依不

12 農、漁民因居住於農、漁村，在就業認定上通常標準較不齊一。舉例來說，農閒時間受訪者可能回答無就業，造成承保資料與家支調查在就業狀態認定上不一致。

13 在合併資料中，有 1,028 個成年人在承保資料顯示爲自己投保，但在家庭收支調查中回答有就業卻沒有記錄薪資所得。除雇主外、按理有業者理應有薪資所得，因此在家庭收支調查中認定全職就業採較嚴格標準：就業且薪資所得爲正。

14 家支調查直接詢問受訪者是否就業。然而，有些受訪者雖回答就業，但卻沒有薪資所得。因此，我們定義家支調查就業者必須同時爲就業且薪資所得爲正。反之則爲非全職就業者。

同保險類別區分，但因這些被保險人依附他人投保，投保類別僅代表所依附投保人資訊，並非被保險人個人資訊，投保類別的區分並不是這麼有意義。普遍來說，非全職就業的準確性低於全職就業準確率，這差距可能是因為勞保投保規定中，有些就業人口並不強制納保勞保（如規模在五人以下公司），¹⁵ 以致這些受雇者可能實際上有就業，卻需依附他人投保健保，造成非全職就業的準確性低於全職就業。

表 2 的結果還有兩點值得注意。第一，即使在較準確的全職就業狀態，承保檔所認定就業和家支調查所調查就業仍約 4% 差距。這差距可能是因為兩資料庫在聯結上的時間誤差。由於家支調查通常在當年度年底至隔年 2 月進行，和承保檔擷取當年度年底資訊，存在一到兩個月的時間差。受訪者有可能在此段時間工作發生變動，導致就業狀態改變，因而造成兩者差距。一般來說，工作穩定投保人發生這類可能性較低，這解釋了為何在勞保投保人中，公營事業機構工員（12G），公務機關、學校工員（13I，13J，13K）所判定就業狀態正確性較高，相對的，民營事業機關受雇者（12H）的準確度較低。

第二，由於勞保投保規定，承保資料所推估非全職就業，和家支調查所認定非全職就業間存在一定誤差，且這差距很難因特定樣本選擇而消除。若研究者想使用承保資料來推估投保人是否失業，可能會因所推估非就業狀態準確性較低而有所保留。為了降低這個問題的損害，一個可行方式是利用健保資料庫縱貫資料特性，將重點擺在投保人就業狀態改變（從就業轉至非就業），這樣應能有效降低承保檔中認定非全職就業的可能誤差。¹⁶

三、就業薪資

和就業相關的另一個重要社經變數為所得。由於所得資訊具有隱密性，

15 根據勞工保險條例第六條，年滿十五歲以上，六十歲以下之左列勞工，應以其雇主或所屬團體或所屬機構為投保單位，全部參加勞工保險為被保險人。一、受僱於僱用勞工「五人以上」之公、民營工廠、礦場、鹽場、農場、牧場、林場、茶場之產業勞工及交通、公用事業之員工。二、受僱於僱用「五人以上」公司、行號之員工。三、受僱於僱用「五人以上」之新聞、文化、公益及合作事業之員工適用之。

16 一般來說，除非投保人工作由公、勞保這類穩定工作，轉換成為較難判定就業投保類別工作（如開計程車），使用縱貫資料應該可精確推估就業狀態的改變。

即使透過嚴謹訪查，是否能取得可靠所得變數仍舊是一大疑問；再加上健保資料的原始目的是以保險支付為主，被保險人所得本來就不是蒐集重點，造成健保資料中對所得變數相當缺乏。然而，所得高低不但直接影響醫療使用，同時也常會影響消費者所面對的可能醫療選擇，可說是分析醫療需求的一個決定性因子。基於這個變數的重要性，我們試圖透過承保檔的投保薪資來估算薪資所得。雖說薪資所得僅是個人所得的一個項目，但是相當關鍵的一個部分，應可做為推估個人所得的重要參考。

表 3 比較承保檔中投保薪資和家支調查中調查薪資的差異。由於薪資所得僅存在於就業人口，樣本侷限在前述就業準確度較高投保類別，即公保（除地方公職人員外）和勞保（除職業工會外）人口。此外，為比較不同保險類別的差異，表 3 也依投保類別加以區分。從表 3 可知，公保人口的投保薪資平均為 53,628 元，而其家支調查薪資所得平均所得為 59,773 元，若以家支調查薪資為分母，兩者差距為 9.78%，但絕大多數都在一個標準差內；若再細分保險類別，公保各類保險類別的投保薪資約低估 10-15%，這低估主要是因為公保的投保薪資為全薪的八成七，自然存在一成以上差距。¹⁷

在勞保樣本上，投保薪資平均低於調查薪資約一成左右，但不同保險類別差距頗大。根據投保規定，勞保樣本是以全薪投保，雖說投保薪資有最高和最低限制，可能無法完全反應實際薪資，但因適用最高和最低級距比例相當低，¹⁸ 按理來說，健保投保薪資應該相當接近家支調查薪資。¹⁹ 然而，許多小型民營企業，長期被詬病利用低報薪資，來降低健保保費支出，因而導致投保薪資低於真實薪資。這個趨勢在表 3 中清楚顯示，對於公營企業以勞保投保者，投保薪資甚至高過調查薪資 4-8%；但在民營企業員工上，投保

17 資料比對時（2005 年）公保的投保薪資計算標準為全薪（含本俸、研究費、主管加給）的八成七，但自 2009 年 10 月起，公保投保薪資標準改為全薪（含本俸、研究費、主管加給、地域加給）的 93.52% 計算。

18 資料比對年度（2005 年）適用最低薪資（NT\$15,820）有 4.68%，最高薪資的僅有（NT\$131,700）0.72%。

19 根據健保局投保規定，健保投保薪資除本薪外還包括績效相關獎金。舉例來說，假使某公司員工領取四個月績效獎金，則會以原來十二個月本薪，加上四個月績效獎金為該年度薪資，再將年度薪資除以十四個月作為該員工的投保薪資。

表 3：承保檔所推估投保薪資準確性

	承保資料 投保薪資 ^a	家支調查 薪資 ^a	(家支—承保)／ 家支 (%)	差距 $\leq \pm 1$ s.d. ^b
公保 (11A,B,D,E,12F)	54,309 (15,895)	60,195 (19,819)	9.78%	90.12% (1164)
中央公教、公職人員 (11A)	57,947 (19,333)	65,387 (24,297)	11.38%	88.78% (294)
非中央公教人員 (11B)	50,824 (11,275)	55,976 (13,369)	9.20%	88.11% (656)
私立大專教職員 (11D)	63,394 (18,583)	75,299 (36,395)	15.81%	77.55% (49)
私立中小教職員 (11E)	45,921 (14,495)	53,197 (15,757)	13.68%	75.76% (33)
公營事業職員 (公營)(12F)	62,252 (19,771)	65,739 (21,190)	5.30%	89.39% (132)
勞保 (12G,12H,13I,13J,13K)	40,227 (26,830)	35,503 (20,145)	-13.31%	88.71% (8688)
公營事業工具 (勞保)(12G)	57,234 (18,148)	52,887 (18,664)	-8.22%	99.55% (224)
民營事業受雇者 (12H)	34,967 (20,214)	40,347 (27,562)	13.33%	87.75% (7919)
高收入 ^c	65,195 (22,124)	62,413 (37,428)	-6.16%	90.72% (1660)
中收入 ^c	34,162 (5,913)	40,265 (23,650)	15.16%	69.14% (3367)
低收入 ^c	18,673 (2,688)	28,511 (15,231)	34.51%	29.74% (2892)
公務機關、學校工具 (13I,13J,13K)	33,716 (13,893)	32,431 (12,677)	-3.96	96.51% (545)

a：括弧內為標準差。

b：計算差距是以健保投保薪資標準差計算。

c：高、中、與低收入為將民營事業機構受雇者之投保薪資分配切成三等分 (three quantiles)；低於 24,000 者為低收入，24,000 ~ 42,000 為中收入，高於 42,000 者為高收入。

薪資平均低於調查薪資 13.3%，兩者差距相當大。若將民營企業投保人投保薪資區分三等份，投保薪資在最低 1/3 投保人，其投保薪資低於實際薪資約三分之一；²⁰ 投保薪資在最高 1/3 投保人，其投保薪資反高於家支調查薪資約 6%，顯示勞保民營企業投保薪資所得越低者，低報的情況越嚴重。由此可知，使用健保投保薪資推估員工實際薪資，在勞保低薪資投保人的準確性偏低。

由於投保薪資在勞保低薪資部分有明顯低估，要使用健保投保薪資，一個可能方法是採「相對所得」進行分組。表 4 將表 3 中樣本，依投保薪資所得區分為高、中、低三組，並將調查薪資依所得同時區分三組來比較，以公保來說，超過六成以上樣本的投保薪資和調查薪資，均落在相同所得分組內；勞保樣本兩者相同比例較低，但投保薪資中低所得分組中，有高達六成也落在調查薪資中較低所得分組，顯示「相對所得」可能較「絕對所得」較能掌握勞保低所得的情況。這不失為一個可能的解決方式。

表 4：投保薪資與調查薪資在不同所得分組下比較

			家支調查薪資			樣本數
			低	中	高	
健保投保薪資	公保	低	74.94%	21.62%	3.44%	407
		中	17.99%	59.64%	22.37%	389
		高	3.53%	18.48%	77.99%	368
	勞保	低	64.02%	25.33%	10.65%	2,649
		中	33.71%	44.62%	21.67%	2,824
		高	6.09%	22.57%	71.34%	2,446

a：高中低收入的區分為以該所得變數的 1/3 等份進行切割。

b：公保不包含地方選擇人員（11C），勞保僅包含民營事業機構受雇者（12H）。

20 民營企業低報投保薪資的另一個解釋是勞工薪資所得通常會包含其他津貼（如交通費，餐點費，或住宿補貼），而這些津貼在計算投保薪資時多半會從薪資中扣除，導致投保薪資較調查薪資為低情況。但即使納入這個可能性，仍很難解釋投保薪資低於調查薪資 1/3 的情況。

四、兒童家庭結構

最後比較兒童家庭結構。在家支調查中，每個抽樣家戶均記錄家庭每位成員個人背景，以及各成員與戶長關係，可清楚推估每個家戶的家庭結構。但在健保資料庫中，所有資料均以個人為單位，無法推估每個被保險人的家庭組成。但家庭組成變數，在兒童醫療相關研究相當重要，不論是兒童是否與父母同住，兄弟姊妹數目，長幼排行順序，皆是決定兒童醫療需求的關鍵變數。

在推估兒童家庭結構時，或可嘗試使用承保檔中「投保人稱謂」來補足。該變數記錄了投保者與被保人關係，包含配偶、父母、監護人、或子女等；透過投保人稱謂和投保人代碼，可協助我們認定兒童父或母身分代碼。由於健保給付內容和保費多寡無關，在減少健保保費誘因下，絕大多數家長會選擇以薪資較低一方作為家中所有兒童的惟一投保人（韓幸紋、連賢明，2008: 589-623）。²¹ 在這個前提下，可將同一投保人名下所有子女找出，並據此認定兄弟姊妹數目和該兒童胎次，並判斷該兒童是否和父或母（投保人）同居。

為了瞭解這些推論可靠性，我們在承保檔中找出透過父或母投保的兒童家庭，與家支調查所記錄家庭結構比對。表五第一欄和第二欄中首先比較這些兒童是否與投保人（父或母）同住。由於兒童年紀可能會影響同居與否，我們將兒童依年紀劃分為 6 歲以下、6-12 歲、和 12-15 歲三組。以 6 歲以下兒童來看，當投保人為其父或母，與投保人同居比例高達 94.4%；6-12 歲和 12-15 歲為 93.3% 和 92.6%。亦即絕大多數的兒童均與投保人居住在一起。

表 5 第二欄列出透過父或母投保兒童同時和雙親同住比例。雖說使用健保資料無法判定是否與雙親同住，但我們可透過家庭收支調查，瞭解同時和雙親同住比例。和第一欄結果相較，和雙親同住比例普遍低於與投保人（父或母）同住比例一成以上：80.3%（0-6 歲），83.3%（6-12 歲），以及 82.8%

21 這個推論只適用夫妻投保薪資不同家戶。當夫妻投保薪資相同時（如投保類別為地區人口，農保，或職業工會），此時將子女投保在單一投保人並不會造成保費節省，但根據我們比較結果，即使是在這些投保類別中，家戶中子女仍大多投保在單一投保人名下。

表 5：家戶中未成年小孩依附狀況^a

	和父或母同居 ^b	和父且母同居 ^b	單一投保人 ^c (與父或母同居者)
兒童年紀			
0-6	94.4% (2,696)	80.3% (2,696)	
6-12	93.3% (3,251)	83.3% (3,251)	
12-15	92.6% (1,724)	82.8% (1,724)	
家中至少有 1 名 15 歲以下兒童			
1 名			90.8% (1,841)
2 名			96.1% (1,829)
3 名			95.9% (508)
>3 名			96.3% (54)

a：選擇兒童承保資料投保人為父或母

b：括弧內為兒童樣本數

c：括弧內為家戶樣本數

(12-15 歲)。這個差距可能有兩個：第一，父母因工作關係無法同住，導致僅有父或母與兒童同住；第二，父母婚姻離異，而變成單親家庭，自然無法與雙親共住。以 2005 年台灣 20-40 歲年齡層離婚率為 8-9% 來看，²² 後面這個可能性比較高。

最後，我們討論是否能透過承保檔找出兒童手足。和與父母同住相較，這個問題難度相對較高。第一，雖說現行保費規定，會給予父母動機將未成年子女集中在單一投保人，但承保規定並未明文禁止兄弟姊妹不能分別由父或母納保。第二，現有家支調查中僅調查家戶成員和戶長關係，當家中戶長

22 有關 20-40 歲之離婚率如下：20-24 歲為 8.09%，25-29 歲為 9.39%，30-34 歲為 9.83%，35-39 歲為 9.09%。資料來源請參見內政部統計處網站 (<http://sowf.moi.gov.tw/stat/year/list.htm>)(2005)。

爲年長長輩（如爺爺或奶奶），且家戶中爲大家庭共居時（有叔、伯、阿姨時），家支調查無法判定兒童家庭結構（亦即分辨哪位爲兒童父母）。第三，假使父母離異，各自擁有部分子女監護權，兄弟姊妹可能分別依附父母，造成同一投保人所依附兒童並非完整家庭結構；但需注意的是，此時家支調查中也可能無法認定完整家庭結構。

爲了適切比較，我們採取下列方式選取樣本。首先，我們以家支調查中所認定家庭結構爲基準。雖說此舉變相放棄認定父母離異兒童家庭。但受限於資料，我們僅能以家支調查爲比較基準；其次，我們排除家戶中具有叔伯或妯娌共居的現象，僅使用家支調查中非大家庭共居家戶。²³ 在這兩個原則下，表 5 從承保資料中挑出名下至少有一名 15 歲以下兒童的投保人，連結這些選擇後的家支調查家戶，再比較這些家戶中 15 歲以下兒童數目。

根據表 5 第三欄結果，2005 年的家支調查樣本所串聯承保資料中，共有 1,841 家戶有一名 15 歲以下之兒童。此時承保檔所推估家庭結構，和家支調查所得家庭中有一名子女且共居的比例爲 90.8%。隨著家中兒童的人數增加爲二名，該比例增爲高達 96.1%，而在三名和三名以上兒童時，比例約爲 96%。顯示出承保檔所推估家庭兒童數目，可能反映出子女數目增加後，投保人和兒童共居比例增加情況，和實際兒童數目應相當接近。

肆、實例說明

一、樣本選擇

爲了使讀者能清楚如何使用承保資料推估社經變數，本節擬透過一個實例，講解如何利用承保檔所推估社經變數進行分析。在這個實例中，我們分析罹患乳癌對受雇婦女存活、就業和薪資的影響。有關健康衝擊對就業、所得或消費影響分析在國外已經相當討論（Gertler and Gruber, 2002; Wagstaff, 2007），近來也有研究針對乳癌作分析（Bradley et al., 2005; Drolet et al., 2005;

23 比較時僅使用核心家庭或三代共居但妯娌不同住，排除辨識兒童父母困難家戶。

Bouknight et al., 2006)。²⁴ 相較之下，台灣由於受限於資料，無法有效結合病患醫療利用和社經變數，相關研究相對缺乏。

為分析乳癌對受雇婦女存活、就業和薪資影響，我們首先透過重大傷病證明明細檔 (HV)，找出 1997-2002 年間新罹患乳癌受雇女性作為實驗組樣本，將這些癌症患者身份代碼，與承保資料檔串聯取得各時期存活、就業和所得資料；其次以傾向分數配對法 (propensity score matching method, PSM)，找出與罹患乳癌婦女樣本特性「類似」的非癌症婦女，以此樣本做為控制組 (control group)，串連承保資料檔取得各時期資料。最後，比較兩組樣本在罹病 6、12 和 24 個月後的存活、就業和薪資差異。為使讀者清楚瞭解資料建構細節，下面一步步說明研究過程。

(一)、透過重大傷病證明檔，選擇罹患乳癌婦女

我們以下列條件選擇新罹患乳癌樣本：(1)國際疾病分類 (International Classification of Disease) 中診斷代碼為「174」，即女性乳房罹患惡性腫瘤 (不包括原位癌或其他良性及未明腫瘤)。(2)重大傷病卡申請日期於 1996 後的乳癌病患，這可協助刪除健保開辦前即罹患乳癌患者。(3)由於資料最晚至 2004 年，分析時選擇 2002 年以前罹患乳癌病患，以確保掌握罹病 2 年後的存活、就業和薪資變化。

(二)、串聯承保檔，取得存活、就業與所得資訊

確定罹癌樣本後，使用受雇乳癌病患身份代碼，串聯承保檔取得存活、就業和薪資資料。根據上節結果，我們選擇：(1)被保險人和投保人身分代碼

24 當婦女罹病時，可能因為生產力減低或者休閒效用提高而降低勞動供給 (McKenna, 1987; Shannon and Shaw, 2005)，但也有可能增加就業以確保就醫資源 (Sheu, 2002; Currie and Madrian, 1999)。近來文獻結果顯示，當婦女被診斷乳癌時，該健康衝擊對就業產生負面影響 (Bradley et al., 2005; Drolet et al., 2005; Bouknight et al., 2006)；隨著病情逐漸穩定，就業的負面衝擊下降甚或無衝擊 (Ganz et al., 1996; Satiriano and Gerald, 1996; Maunsell et al., 2004; Bouknight et al., 2006)；甚至持續就業的罹病婦女，其平均工時增加 (Bradley et al., 2002; Bouknight et al., 2006)。

相同的本國人，²⁵ (2)投保類別為公勞保，²⁶ (3)投保人年齡介於 35-60 歲者。由於乳癌病患從發病至死亡時間可能不長，我們將承保資料先轉換成以月為基礎的格式，再串聯取出存活、就業和所得資訊。有關存活的部分，我們以投保人於兩年內的退保紀錄，來推論這些病患的術後死亡。至於就業和所得資訊，前面已詳述，此處不再重複。

(三)、透過 PSM 配對取得控制組樣本

使用 PSM 配對有幾個要件：區分實驗組和控制組，選擇變數以及配對方法。在這個實例中，實驗組是重大傷病當中確診乳癌病患，控制組則是以健保局所發行 20 萬承保抽樣歸人檔，²⁷ 排除罹患癌症病患後所剩下樣本作為控制組。由於控制組樣本數遠大於實驗組，可以降低選樣配對時重複配對可能性。

在選擇配對變數上，則以確診乳癌「前一年」時病患年齡、投保金額、投保類別，作為配對變數。從承保檔中，可選擇這些變數進行配對，但配對時則以確診前一年，而非確診癌症（或罹癌後）的變數內容進行配對，以降低可能的配對偏誤。由於病患確診癌症前，可能已因身體不適自行調換工作，乃至辭職或退休，造成實驗組的配對變數可能包括癌症影響。若在這基礎上進行配對，會低估乳癌的健康衝擊；因此，在配對時特別選用確診前一年變數內容來降低配對偏誤。

最後有關配對方法。使用 PSM 方法配對時，按理應比較實驗組與控制組中變數特性，並將特性相近樣本進行配對。但隨著配對變數的個數增加，配對難度相對增加。Rosenbaum and Rubin (1983) 證明無混雜性假設 (unconfoundedness) 下，可將多維度的變數配對，簡化成單一維度的傾向分數 (propensity score) 來進行配對。²⁸ 但即使運用傾向分數，文獻上對如何以傾向

25 為確保被保險人為本國人，我們排除身份證字號小於 10 碼者。

26 公保除地方公職人員 (11C) 外皆納入，勞保則除職業工會外均納入。

27 國衛院以 2000 年承保者為母體。抽樣出 20 萬人 (共 4 組，一組 5 萬人) 樣本，並取得這些樣本由 1996 年至 2004 年的所有健保資料。

28 假設 Y_1 為接受實驗效果， Y_0 為沒有接受實驗效果， W 為樣本是否選擇加入實驗組 ($W=1$)

分數配對仍有不同看法。²⁹ 為簡單起見，配對時我們採用最近鄰域配對法 (nearest neighbor matching) 來計算傾向分數，並以一對一非重複對應 (one-to-one, non-replacement) 方法進行配對。

二、敘述統計與估計結果

表 6 第一欄列出公、勞保樣本實驗組婦女基本特性。由表可知，1997 年至 2002 年間罹患乳癌樣本共 6,355 人，其中 1997 為 827 人，其後每年逐漸增加，至 2002 年為 1,245 人。這些婦女罹病年齡以 51-55 歲者居多，約佔 3 成，其次為 46-50 歲和 56-60 歲，少部分患者年紀低於 40 歲。投保金額主要為 16,500 元以下 (26.7%)；隨投保薪資級距增加，罹病婦女比率逐漸下降。最後，就保險類別而言，88.8% 為勞保，其餘為公保。同表第二欄列出控制組樣本特性。由於控制組樣本是採傾向分數配對法配對，不論是婦女年齡，投保金額，投保類別等特性，和實驗組樣本相當類似，幾乎沒有明顯差別。

表 7 分別列出實驗組與控制組在罹病前後的存活和就業比率。由於配對是以罹病前一年以公勞保身份投保樣本為基準，因此罹病前一年，存活和就業率均為 100%；罹癌前 6 個月和罹癌確診時，控制組和實驗組不論在存活、就業或薪資上，差距均相當小，顯示確診前兩組特性相當，健康衝擊相當有限。但隨著罹癌時間增加，實驗組的存活情況穩定減少，到了兩年後，乳癌病患死亡約 5%，和癌症登記檔相較表 7 所得，乳癌病患 2 年後的存活比例稍低。³⁰ 這可能是因為樣本為就業婦女其健康狀況一般較非就業婦女為佳

或控制組 ($W=0$)。所謂無混雜性假設 (unconfoundedness) 指的是 $(Y_1, Y_0) \perp W | X$ ，亦即當控制所有可觀察到的共同因素 (X) 後，樣本是否加入實驗 (或政策及計畫) 和其政策結果獨立。

29 一般常見的 PSM 配對方法包括最近鄰域配對法 (nearest neighbor matching)，半徑配對法 (radius matching)，Kernel 配對法 (kernel matching method) 和階層配對法 (stratification method)。有關配對方法介紹，可參見 Abadie et al. (2004)。

30 根據癌症登記檔統計，88-92 年乳癌癌症患者的兩年存活率為 92.52%，較我們所估計的為低。有關癌症登記檔的存活數據，請見台灣癌症登記小組網站 <http://tcr.cph.ntu.edu.tw/main.php?Page=A5B3> (1999-2003)。

表 6：乳癌病患與對照組樣本比較

	實驗組	控制組
罹癌年度		
1997	827	827
1998	878	878
1999	1,075	1,075
2000	1,094	1,094
2001	1,236	1,236
2002	1,245	1,245
年紀		
35～40	447	449
41～45	951	954
46～50	1,545	1,543
51～55	1,962	1,954
56～60	1,450	1,455
投保金額		
16,500 以下	1,587	1,591
17,400～21,000	916	919
21,900～26,400	805	800
27,600～33,300	852	846
34,800～42,000	921	923
43,900～53,000	643	642
53,000 以上	631	634
投保類別		
公保 ^a		
中央公教、公職人員 (11A)	151	149
非中央公教人員 (11B)	625	624
私立大專教職員 (11D)	11	6
私立中小教職員 (11E)	10	4
公營事業職員 (公保)(12F)	78	71
勞保 ^b		
公營事業員工 (勞保)(12G)	66	70
民營事業受雇者 (12H)	5,185	5,201
公務機關、學校員工 (13I,13J,13K)	229	230
總計	6,355	6,355

a：不包括地方公職人員 (11G)

b：不包括職業工會會員 (21)

表 7：乳癌婦女存活、就業與薪資

實驗組	存活狀況(%)		就業狀況(%)		薪資狀況		
	控制組	實驗組	控制組	實驗組	控制組		
罹癌前 12 個月	100.0%	100.0%	100.0%	100.0%	30758 (15036)	30775 (15098)	
罹癌前 6 個月	100.0%	99.9%	95.9%	95.7%	30167 (15865)	30488 (16251)	
罹癌前	100.0%	99.8%	92.3%	92.8%	29757 (16317)	30076 (16767)	
罹癌後 6 個月	99.1%	99.6%	84.6%	89.9%	28540 (17249)	29717 (17182)	
罹癌後 12 個月	97.7%	99.4%	80.0%	88.0%	28112 (17542)	29648 (17749)	
罹癌後 24 個月	95.4%	99.4%	73.4%	84.1%	27267 (17987)	28879 (18128)	
罹癌人數						6,355	

所致。

相較之下，就業和薪資的變化比較大。罹癌半年後，實驗組就業較控制組約低 5%，一年後低約 8%，兩年就業比率差約一成，顯示罹患乳癌約使一成左右受雇者放棄全職工作；相同的，對於持續工作者，罹癌半年後實驗組薪資較控制組約低 1,200 元，一年約低 1,400 元，兩年後差距拉大為 1,600 元，顯示即使持續工作，乳癌患者有可能放棄升遷，甚至轉換較輕鬆工作，導致薪資持續較控制組下降。

伍、結論

近年來許多研究者使用健保資料庫從事相關研究。連賢明（2008）說明健保資料庫各個子資料庫的整理原則，並解釋如何連結醫院、醫師、病患各個子資料庫資訊，來建構一個完整醫療研究資料。但礙於健保資料的原始設

計，如何取得病患社經變數則是使用健保資料的另一個重大挑戰。這個問題不但加深健保資料使用難度，也阻礙了健保資料普及；特別在社會學科中，病患社經變數往往是研究重心，缺乏這些變數直接降低健保資料在社會學科的應用性。舉例來說，近來國外研究開始關注所謂「健康不均等」(health disparity) 議題，討論為何不同所得、教育水準民眾的醫療利用和健康狀態，會存在穩定甚至逐漸擴張的差距。³¹ 由於現有健保資料庫缺乏所得、教育變數，這類研究在國內發展相對受限，說明社經變數缺乏會限制了醫療研究者的議題選擇。

在主管機關協助下，本研究以 2005 年家支調查家戶為樣本，串聯健保資料承保檔，比對使用承保檔推估社經變數，包括戶籍縣市、全職就業、薪資所得，和兒童家庭結構的可靠性。根據比較結果，本文發現：(1)以投保單位縣市推論投保人戶籍所在縣市，在地區人口，農民，與地方公職人員準確性達九成五以上；(2)由於全職就業者須以被保險人身份投保，選擇投保人和被保險人身份代碼相同者，在公、勞保納保對象的就業狀態準確度相當高，平均達九成以上；(3)使用就業準確樣本，以投保薪資來推論薪資所得，在公保樣本平均低估約 10%，勞保中民營受雇者的差距最大，投保薪資最低的 1/3 平均低估三成，薪資較高的 1/3 則沒有差異；(4)未成年兒童其投保人為父或母時，與投保人同居比率達 94% 以上，但和雙親同居則僅約 85% 左右；(5)以承保資料投保於單一投保人名下兒童來推估兒童家庭結構，其準確性高達 90-96%。

根據比較結果，我們對未來研究提出幾點建議。第一，許多學者認為居住地的就醫選擇對醫療結果具關鍵性影響，但礙於健保資料限制，無法確知病患居住地點，本文提供一個方法來推估病患居住地，研究者可據此分析居住地對就醫選擇影響；第二，本文提供了一個方法來建構兒童家庭結構，這些變數應對從事兒童醫療相關研究應有相當助益；第三，本文說明如何利用

31 一個最明顯例子說明「健康不均等」(health disparity) 議題的重要性，美國 AHRQ (Agency for Healthcare Research and Quality) 自 2003 年起，每年針對國內健康不均等發佈年度報告 (National Healthcare Disparity Report)。

健保資料縱貫特性 (longitudinal)，來分析健康衝擊對病患就業和薪資的影響。雖說這個例子僅針對乳癌病患進行分析，但相同方式可應用到其他疾病，藉以瞭解不同疾病對病患個人生活的整體衝擊。這結果可提供主管機關決定政府在評估擴張健康保險或是其他社會保險（長期照護）時參考。

文末說明本文的研究限制。首先，本文僅使用 2005 年家支調查樣本進行比對。雖說家支調查樣本具有代表性，但我們提醒讀者整個比對是以單年樣本進行；其次，家支調查和承保檔在合併時，因彼此的調查時間不同，可能出現一至二月的差距，這對於一些於農曆年前變換工作的投保人其就業和薪資比對可能出現誤差；最後，家支調查訪查過去一年的就業、薪資，受訪者有可能存在記憶誤差 (recall bias)，而使得我們比較產生誤差。但即使如此，我們的比較結果顯示，以承保資料推估社經變數的準確性相當高，期望透過這些驗證，能吸引更多社會學科學者來使用健保資料。

參考資料

A. 中文部分

中央健康保險局

2009 〈投保單位選擇資料表〉。2009 年 11 月 17 日，取自 http://www.nhi.gov.tw/webdata/webdata.aspx?menu=18&menu_id=678&webdata_id=3436&WD_ID=722

內政部統計處

2005 〈婚姻狀況統計年報〉。內政部統計年報，2009 年 11 月 10 日，取自 <http://sowf.moi.gov.tw/stat/year/list.htm>

台灣癌症登記小組

1999-2003 〈癌症五年相對存活率〉。台灣癌症登記小組，2009 年 12 月 24 日，取自 <http://tcr.cph.ntu.edu.tw/main.php?Page=A5B3>

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〉，《臺灣公共衛生雜誌》22(6): 419-430。

主計處

2005 〈家庭收支調查〉。2009 年 9 月 4 日，資料取得自中研院網站 <http://srda.sinica.edu.tw/govdb/>

許績天、連賢明

2007 〈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》35(4): 415-450。

連賢明

2008 〈如何使用健保資料進行經濟研究〉，《經濟論文叢刊》36(1): 115-143。

國家衛生研究院

2005 〈健保資料庫承保檔〉，「全民健康保險研究資料庫」。2009年9月14日，取自 <http://w3.nhri.org.tw/nhird/index.php>

2008 〈健保資料研究成果清單〉，「全民健康保險研究資料庫」。2009年11月19日，取自 http://w3.nhri.org.tw/nhird/talk_07.htm

韓幸紋、連賢明

2008 〈降低部分負擔對幼兒醫療利用的影響：以北市兒童補助計畫為例〉，《經濟論文叢刊》36(4): 589-623。

魏郁純

2006 「健康狀態對就業影響—以癌症病人為例」，國立政治大學財政學系碩士論文。

B. 英文部分

Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens

2004 “Implementing Matching Estimators for Average Treatment Effects in Stata,” *The Stata Journal* 4(3): 290-311.

Bouknight, R. R., C. J. Bradley, and Z. Luo

2006 “Correlates of Return to Work for Breast Cancer Survivors,” *Journal of Clinical Oncology* 24(3): 345-353.

Bradley, C. J., H. L. Bednarek, and David Neumark

2002 “Breast Cancer Survival, Work, and Earnings,” *Journal of Health Economics* 21: 757-779.

Bradley, C. J., D. Neumark, and H. L. Bednarek

2005 “Short-Term Effects of Breast Cancer on Labor Market Attachment: Results from a Longitudinal Study,” *Journal of Health Economics* 24: 137-160.

Currie, J. and B. C. Madrian

1999 “Health, Health Insurance and the Labor Market,” *Handbook of Labor Economics* 3: 3309-3416.

Drolet, M., E. Maunsell, M. Mondor, C. Brisson, J. Brisson, B. Mâsse, and L. Deschênes

2005 “Work Absence after Breast Cancer Diagnosis: A Population-Based Study,” *Canadian Medical Association Journal* 173(7): 765-771.

Ganz, P. A., A. Coscarelli, C. Fred, B. Kahn, M. L. Polinsky, and L. Petersen

1996 “Breast Cancer Survivors: Psychosocial Concerns and Quality of Life,” *Breast Cancer Research and Treatment* 38: 183-199.

Gertler, P. and J. Gruber

2002 “Insuring Consumption Against Illness,” *American Economic Review* 92(1): 51-76.

Lien, Hsien-ming, Shin-yi Chou, and Jin-tan Liu

2008 “Hospital Ownership and Performance: Evidence from Stroke and Cardiac Treatment in Taiwan,” *Journal of Health Economics* 27: 1208-1223.

- 2009 “The Role of Hospital Competition on Treatment Expenditure and Outcome: Evidence from Stroke and Cardiac Treatment in Taiwan,” *Economic Inquiry* 48(3): 668-689.
- Maunsell, E., M. Drolet, J. Brisson, C. Brisson, B. Mâsse, and L. Deschênes
2004 “Work Situation after Breast Cancer: Results from a Population-Based Study,” *Journal of the National Cancer Institute* 96(24): 1813-1822.
- McClellan, Mark, B. J. McNeil, and J. P. Newhouse
1994 “Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?” *Journal of the American Medical Association* 272(11): 859-866.
- McKenna, R.
1987 “Overview of the Status of Patients with Cancer with Reference to Employment. Proceedings of the Workshop on Employment Insurance and the Patient with Cancer,” *American Cancer Society*: 2-4.
- Rosenbaum, P. and D. Rubin
1983 “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika* 70: 41-55.
- Satariano, W. A. and N. D. Gerald
1996 “The Likelihood of Returning to Work after Breast Cancer,” *Public Health Report* 111: 236-241.
- Shannon, C. S. and S. M. Shaw
2005 “If the Dishes Don’t Get Done Today, They’ll Get Done Tomorrow: A Breast Cancer Experience as a Catalyst for Changes to Women’s Leisure,” *Journal of Leisure Research* 37: 195-215.
- Sheu, S.
2002 “Labor Force Participation and Employer-Provided Health Insurance,” Ph.D. Dissertation, Texas A&M University (Unpublished).
- Wagstaff, A.
2007 “The Economic Consequences of Health Shocks: Evidence from Vietnam,” *Journal of Health Economics* 26: 82-100.

How to Construct Social-Economic Variables from National Health Insurance Data

Hsien-ming Lien

Professor

Department of Public Finance, National ChengChi University

ABSTRACT

Recently more and more researchers have conducted studies based on National Health Insurance Data (NHID). However, few studies are related to the social sciences due to the lack of social-economic variables in NHID. With the help from the Department of Health, which merged the sample of the 2005 Family Income and Expenditure Survey (FIES) with eligibility files of NHID, this study examines the validity of inferring four variables from eligibility files: residence county, full-time employment, working salary, and children's family structure. Our results show that: (1) The accuracy rate of inferring residence county based on one's county of enrolled units is above 95% for those obtaining coverage from a farmer's association, registration offices, or as public electorates; (2) The accuracy rate of inferring full-time employment status using self-insured status is above 90% for the sample of public employee insurance (PEI) or labor insurance (LI); (3) For PEI sample, the working salary inferred from one's insurance premium is approximately 90% of their reported working salary in FIES. For LI sample, the inferred salary is roughly 30% below the FIES working salary for low-paid workers (bottom 1/3), and 6% higher for high-paid ones (Top 1/3); (4) For children under 15 obtaining coverage through parents, more than 90% of them co-habit with at least one of their parents; (5) The accuracy rate of inferring sibling number and birth order based on eligibility files is at least 90% for children aged below 15. Finally, I use a research example to analyze how the incidence of breast cancer affects the survival, employment, and working salary of employed women.

Key Words: Taiwan National Health Insurance data, employment, working salary, family structure