

健康保險資料庫之學術應用：重大傷病證明 資料檔分析之經驗

陳建仁 林莉華

健康保險資料庫之學術應用，不僅有助於醫療保健工作之規劃、執行與考核，也有助於國民健康狀況、衛生需求及疾病致因之分析與探討。我國全民健康保險於一九九五年三月一日開始實施以來，即建立相當可觀之資料庫，可以提供學術研究所需之資訊。本文根據中央健保局委託分析「重大傷病證明領用保險對象檔案」之經驗，說明健康保險資料庫之學術應用範圍、應具備之特質、可能限制、以及未來改善提昇之方向。（中華衛誌 1997；16(6)：513-521）

關鍵詞：健保資料庫，應用範圍，特性，限制。

前 言

世界各國的健康保險資料庫經常提供重要資訊，作為衛生、保健、醫療之學術研究素材，對於加強醫療保健服務、促進全民健康有相當大的貢獻。以台灣地區公投保保者為研究對象所作之長期追蹤研究，發現B型肝炎病毒表面抗原帶原者，罹患肝細胞癌的危險性，遠高於非帶原者，其相對危險性達兩百倍[1]。因而促使衛生署積極推展全國性B型肝炎預防接種工作，獲致相當可觀的成效，近十年來的小兒肝細胞癌發生率顯著降低[2]。此一重要的世代研究，即是以公務人員保險之死亡給付檔和住院補助檔，協助研

究對象之發病與死亡狀況之追蹤。健康保險資料庫可應用於國民健康狀況與疾病型態之描述，衛生需求與衛生需要之評估，疾病危險因子與預防對策之探討，以及醫療保健工作之優先順序排定、規劃、執行與成效考核等。

我國全民健康保險於一九九五年三月一日開始實施以來，即著手建立各項與健康保險業務相關之資料庫，其中包括了承保系統、醫務管理系統、醫療給付系統、財務管理系統、決策輔助系統、其他系統等六大類，為數相當可觀。此類資料庫之建立，原係為因應全民健康保險各項業務之需要，但是其中之資訊，亦可提供衛生行政、醫院管理、流行病學等學術研究使用。舉例而言，醫療給付(P)系統之門診行政審核(PC)子系統的處方及治療明細檔(PCB_AMPS)以及住院行政審核(PD)子系統的住院醫療費用清單明細檔(PDB_PREAP)，即有可能提供門診及住院病例的主診斷和次診斷的《國際疾病傷害及死因分類第九版》(ICD9-CM)代碼。使用這些資料必須對資料檔的產生(主治醫師所作之診

國立台灣大學流行病學研究所

聯絡人：陳建仁

聯絡地址：台北市仁愛路一段一號

聯絡電話：(02)397-0800轉8359

傳 真：(02)351-1955

電子郵件：cjchen@ha.mc.ntu.edu.tw

投稿日期：86年1月10日

接受日期：86年9月9日

斷)、蒐集(醫院診所對診斷之登錄與申報)、建檔、核校、勘誤、品質管制、以至儲存備檔第一系列過程之完整性和正確性。

中華民國公共衛生學會之癌症登記工作小組，接受中央健康保險局委託，分析重大傷病證明領用保險對象檔案之資料，以期瞭解不同地區分局發放重大傷病證明的現況，以作為未來改進規劃之參考。本文根據分析之經驗，舉例說明健保資料庫之學術應用範圍、應具備之特質、可能限制及未來改善方向。

重大傷病證明領用保險對象資料檔之內容

依據《全民健康保險法》第三十六條之規定，保險對象有重大傷病者，免付自行負擔之費用，而重大傷病之範圍，由主管機關訂之。又依據《全民健康保險醫療辦法》第十條之規定，保險對象經確定診斷為重大傷病，得檢具保險特約醫院診所開立之診斷證明書，報經保險人核可後，發給重大傷病證明。中央健康保險局於一九九五年三月起，接受重大傷病證明卡之申請，本文分析之資料檔，係截至一九九六年一月之檔案。此資料檔共有十六欄位，分別是局別代碼、局別名稱、重大傷病類別、身份證字號、姓名、性別、醫事機構代碼、ICD9-CM代碼、診斷病名、申請日期、出生日期、地址、卡號、縣市別代碼、電話號碼、年齡。這十六欄位共有260格之資料，其中重大傷病類別共有25類、縣市別代碼有25碼，包括台灣有21縣市、台北市、高雄市、金門縣和連江縣。

重大傷病證明資料檔原係由六分局分別建檔，各分局所用之電腦軟體、欄位格式、譯碼規則、工作手冊都不相同，因此最初各分局之資料檔，有相當大的差異，完全無法合併分析。後來應委託單位之要求，逐筆資料重新核校、勘誤、補遺和建檔，但仍有相當比例之資料無法補正。

重大傷病證明領用保險對象資料檔之正確性

本資料檔原有216,505筆資料，經核對後發現有5,883筆(2.7%)屬於重複領證者，重複

領證係指身份證字號、姓名、重大傷病類別、和ICD9-CM代碼完全相同者；若身份證字號、姓名、重大傷病類別相同，但ICD9-CM不同者，則不列為重複領證。在六分局當中，以中區分局重複領證百分比最高(12.2%)，而以台北分局最低(0.2%)。在刪除重複資料，僅保留一筆資料的狀況下，尚留213,146名領證者進行資料分析。至於資料的正確性方面，身份證字號錯誤(不符合核校公式)者有2,603名(1.2%)，但身份證字號僅有九碼，卻有29,256名(13.7%)，其中以北區分局錯誤率最高(62.2%)，高屏分局也有36.3%的錯誤率，而以中區分局錯誤率最低(0.03%)。至於其他項目遺漏與錯誤之百分比，分別是性別0.1%，醫事機構代碼46.2%、ICD9-CM碼1.3%、診斷病名5.5%、申請日期19.1%、出生日期7.6%、年齡7.5%，此外尚有一名無重大傷病證明卡卡號，三名無姓名。

重大傷病證明資料檔錯誤和遺漏之發生，可能在工作手冊未臻完備、資料處理人員訓練不足，鍵入軟體缺乏偵錯設計、品質保證和品質管制不佳、準備時間不充份等。未來在建立健保相關資料檔時，必須考慮如何確保資訊之完整性、正確性與時效性。學界在使用健保資料時，也必須深入資料檔的原始目的、蒐集方法、可能限制和錯誤，以免得到不一致之結論。

重大傷病證明領用保險對象資料檔之分析結果

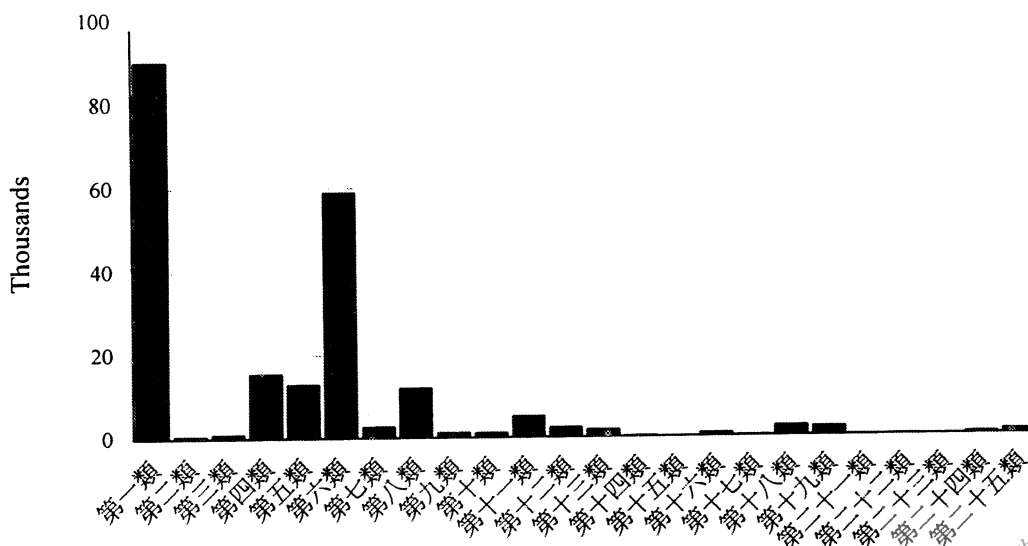
在二十五類重大傷病當中，以第一類(須積極或長期治療之癌症)，領用者最多，總共90,128名，佔所有領用證明者的百分比高達42.3%，其次為第六類(慢性精神疾病)領用者，共有58,962名，佔27.7%。由於不同類別之重大傷病之就醫率和領用證明率並不相同，而且也不會達到百分之百，因此僅按領用證明者人數，是不適宜用來估計各類重大傷病之盛行率。像第二十四類(麻瘋病)的領用證明人數，即有相當程度的低估。再者，同一被保險人同時罹患兩種以上之重大傷病，並不需要領用兩張不同疾病診斷的證明，因

此在較易罹患兩種以上重大傷病的高年齡層而言，利用重大傷病證明領用人數來估計其盛行率很可能會有顯著低估的困擾。各重大傷病類別的領證人數如圖一所示。

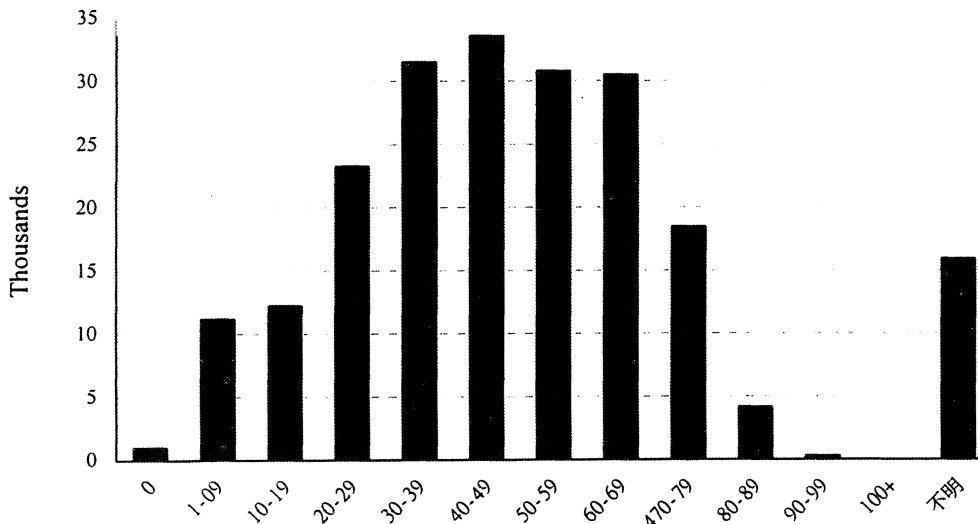
若以台閩地區全人口數來估算，各類重大傷病的每十萬人口領用率，分別是第一類(須積極或長期治療之癌症)的425.6，第六類(慢性精神疾病)的278.4，第四類(慢性腎衰竭必須接受定期血液透析治療者)的73.3，第五類(需終生治療之全身性自體免疫症候群)的61.1，以及第八類(心、肺、胃腸、腎、神經、骨骼系統之先天畸型及染色體異常)的57.1。這四大類的領用證明者之ICD9-CM碼，也有未明或錯誤的現象。其中第一類之未明或錯誤百分比約佔2.5%，其中1,333名之ICD9-CM碼未明，而883名ICD9-CM未詳細註明癌症部位；而有64名屬於非癌症之ICD9-CM碼，這部份有可能是譯碼或鍵入錯誤所造成。

若以年齡別領證人數來看，則以40~49歲最高，其次分別是30~39歲，50~59歲和60~69歲，如圖二所示。各重大傷病類別的男、女性領證人數，則如圖三所示，其中第一類(須積極或長期治療之癌症)、第三類(嚴重

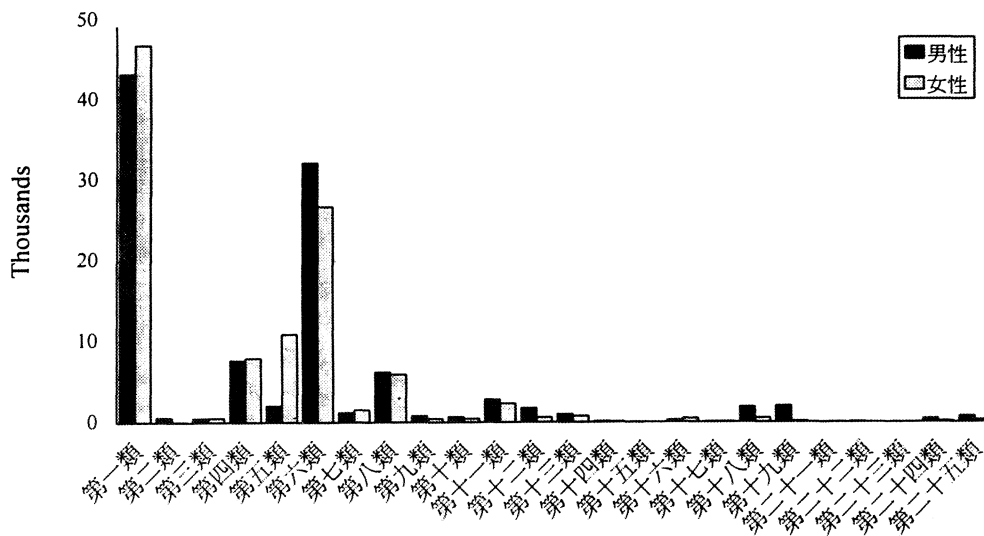
溶血性及再生不良性貧血)、第四類(慢性腎衰竭必須接受定期血液透析治療者)、第八類(心、肺、胃腸、腎、神經及骨骼系統之先天畸型及染色體異常)之男、女領證人數相近。至於女性領證人數明顯高於男性之重大傷病類別包括第五類(需終生治療之全身性自體免疫症候群)、第七類(G6PD代謝異常除外之先天性新陳代謝異常疾病)、第十六類(重肌無力症)和第二十一類(多發性硬發症)。其餘各類重大傷病類別之領證人數，均是男性明顯高於女性。男女領證人數之差異，究竟是導因於各重大傷病類別之發生率的不同，或是發病期間的不同，無法從領證資料中予以釐清。這必須配合癌症登記系統或其他疾病登記系統才能加以分析。即使根據新領證者來分析發生率，亦有可能低估疾病發生率。舉個例來說，如果重大傷病發生得很突然而且也相當嚴重，該名病患可能因猝死或來不及領證即告死亡，則其發病狀況將無法從領用資料檔得知；如果傷病發生之初即被診斷而獲得適切治療，則其嚴重度並未達重大傷病之標準，也不會出現在領用資料檔之中。因此，領證率可以說是重大傷病發生率，嚴重程度和存活期間的複雜函數，因此在應用重大傷



圖一 中央健康保險局重大傷病證明領用對象之各類重大傷病之人數分佈，民國八十四年三月至八十五年一月



圖二 中央健康保險局重大傷病證明領用對象之年齡分佈，民國八十四年三月至八十五年一月



圖三 中央健康保險局重大傷病證明領用對象之男女性各類傷病之人數分佈，民國八十四年三月至八十五年一月

病證明資料於學術研究時，必須十分謹慎和小心。像表一所示，在所有ICD9-CM的疾病種類當中，領證率最高的前十名疾病，分別是精神分裂症、慢性腎衰竭、情感性精神病、女性乳癌、子宮頸癌、直腸癌、肝癌、類風濕性關節炎、肺癌和紅斑狼瘡，其順位和領證率與台灣地區居民常見中老年病的排名順位、發生率或盛行率均有明顯出入。

重大傷病證明領用保險對象資料檔與癌症登記資料檔之比對

根據重大傷病證明領用資料檔以及癌症登記資料檔[3]分析所得之五大癌症領證率和報告發生率，分別如表二所示。其中可見領證率均一致地高於發生率，這是因為發生率僅指某一年內每十萬人口的新發生癌症人

表一 領證率高居前十名之疾病名及領證率

| 疾病名稱(ICD9-CM) | 每十萬人口領證率 |
|---------------|----------|
| 精神分裂症(295) | 183.2 |
| 慢性腎衰竭(585) | 71.4 |
| 情感性精神病(296) | 55.5 |
| 女性乳癌(174) | 53.5 |
| 子宮頸癌(180) | 43.0 |
| 直腸癌(154) | 36.6 |
| 肝癌(155) | 30.2 |
| 類風濕性關節炎(714) | 29.5 |
| 肺癌(162) | 28.2 |
| 紅斑狼瘡(710) | 27.0 |

表二 台灣地區領用重大傷病證明之前五名癌症的領用證明率與癌症申報登記率之比較

| 癌症部位 | 重大傷病證明領用率* | 癌症申報登記率 | 相對比值 |
|------|------------|---------|------|
| 女性乳癌 | 53.5 | 20.5 | 2.6 |
| 子宮頸癌 | 43.0 | 18.7 | 2.3 |
| 直腸癌 | 36.6 | 8.7 | 4.2 |
| 肝癌 | 30.2 | 19.2 | 1.6 |
| 肺癌 | 28.2 | 18.8 | 1.5 |

* 每十萬人口比率，以民國八十三年台閩地區人口數21,177,874為分母

數，而領證率係指每十萬人口的現有接受積極或長期治療之癌症人數。若比較領證率與報告發生率之相對比值，則以肝癌、肺癌最低，子宮頸癌、乳癌次之，而以直腸癌最高，這可能是前二者的疾病嚴重度較高，而存活期較短的緣故。這也說明了利用重大傷病證明領用資料檔來估計疾病發生率、盛行率或存活期間都必須謹慎小心。

進一步就縣市別之膀胱癌領證率和發病率來比較，如表三所示，無論是重大傷病證明領用率或是癌症申報登記率，都是以台南縣和嘉義縣居前三位，兩者均能反映烏腳病盛行地區膀胱癌較盛行的狀況。這可是這三縣市之膀胱癌發生率已經遠高出台灣其他地區，所以雖有各種資料蒐集不足的限制，重大傷病證明資料檔仍可看出端倪。

表四是比較重大傷病證明領用資料檔與癌症申報登記檔之年齡別子宮頸癌患者分佈

狀況。從表中可以得知癌症申報登記檔中的2956名新發生子宮頸病例的年齡分佈，在40～49歲達到高峰；而重大傷病證明領用檔中的9066名所有領證之子宮頸癌病例的年齡分佈，則在50～59歲達到高峰。兩者的差異，可能是子宮頸癌患者被診斷後，仍有相當比例會存活相當一段期間，以致領證者之年齡較老於發病者。

重大傷病證明領用保險對象資料檔 之流行病學應用

現有之癌症中申報登記資料檔和全國死因資料檔，常被國內之長期世代研究利用來作為追蹤研究對象是否罹患癌症或死於特定疾病的連鎖資料來源。換言之，這兩大資料檔對於國內流行病學研究具有相當重要的參考價值。最近在美國新格蘭醫學雜誌上發表

表三 台灣各縣市之膀胱癌病患領用重大傷病證明率及癌症申報登記率

| 縣市別 | 重大傷病領證率* | 癌症申報登記率 |
|-----|----------|---------|
| 台北縣 | 11.7 | 4.0 |
| 宜蘭縣 | 11.6 | 4.7 |
| 桃園縣 | 10.1 | 3.1 |
| 新竹縣 | 7.2 | 3.0 |
| 苗栗縣 | 9.3 | 3.0 |
| 台中縣 | 9.8 | 3.8 |
| 彰化縣 | 10.4 | 2.7 |
| 南投縣 | 11.5 | 2.4 |
| 雲林縣 | 11.1 | 3.8 |
| 嘉義縣 | 22.9 | 9.4 |
| 台南縣 | 23.7 | 8.4 |
| 高雄縣 | 13.9 | 5.5 |
| 屏東縣 | 13.5 | 6.8 |
| 台東縣 | 2.7 | 2.0 |
| 花蓮縣 | 3.9 | 1.1 |
| 澎湖縣 | 18.3 | 6.5 |
| 基隆市 | 15.1 | 5.5 |
| 台中市 | 12.9 | 3.6 |
| 台南市 | 20.4 | 7.4 |
| 高雄市 | 18.2 | 6.8 |
| 台北市 | 17.4 | 4.2 |
| 新竹市 | 12.1 | 5.0 |
| 嘉義市 | 8.4 | 5.6 |

* 每十萬人口比率，以民國八十三年台閩人口數21,177,874為分母

表四 重大傷病證明領用資料檔與癌症申報登記資料檔之子宮頸癌患者年齡分佈之比較

| 年齡 | 重大傷病證明領用檔之百分比 | 癌症申報登記檔之百分比 |
|-------|---------------|-------------|
| <20 | 0.3 | 0.03 |
| 20~29 | 4.4 | 2.5 |
| 30~39 | 14.1 | 21.6 |
| 40~49 | 23.6 | 24.4 |
| 50~59 | 25.4 | 22.1 |
| 60~69 | 21.0 | 18.5 |
| 70~79 | 9.4 | 8.9 |
| 80+ | 1.8 | 2.0 |
| 合計 | 100.0 | 100.0 |

的B型肝炎預防接種導致兒童肝細胞癌發生率之下降[4]，即使用了這兩大資料檔進行連鎖比對，以確定肝細胞癌的發生與死亡現況。

我們比較重大傷病證明領用檔、癌症申

報登記檔和全國死因檔之中的子宮頸癌患者數，結果如表五所示。從表中得知三項資料檔的數據有相當大的出入，其中重大傷病證明領用檔有資料，而癌症申報登記檔沒有資

料的有5,412名；而重大傷病證明領用檔沒有資料，而癌症申報登記檔卻有資料的有6,135名。這有可能是重大傷病證明領用者當中，有相當大的比例並不是在有50床以上之醫院接受診治，所以未被列入癌症申報登記檔；也有可能是新發生的子宮頸癌病例當中，有相當大的比例痊癒而不需長期治療，所以未能領用重大傷病證明。國內學者在利用現存資料檔來追蹤世代研究對象之疾病發生狀況時，仍可參考重大傷病證明領用資料檔，進行資料連鎖比對，以瞭解其是否已罹病而領有重大傷病證明。

全民健康保險資料庫之學術應用 未來應努力之方向

良好的可供學術研究使用之資料庫應具備：(1)完整性(completeness)、(2)周延性(comprehensiveness)、(3)保密性(confidentiality)、(4)正確性(accuracy)、(5)易取性(accessibility)、和(6)可用性(availability)等特質。由於全民健康保險局各項資料檔之原始目的均在於行政作業管理，所以在學術應用上，必須考慮各資料檔之上述特性。有些資料檔有相當高的完整性，但周延性和正確性卻不足，例如在PCB_AMPS(處方及治療明細)資料檔中，詳細列明各項檢驗的內容及給付金額，但卻未包含檢驗數據或判讀結果，所以對疾病診斷之正確性即無從查核。研究者仍必須前往各醫療院所查閱相關病歷，才能確定診斷。有些資料檔有相當高的診斷正確性，但完整性和

周延性卻不足，例如重大傷病證明領用資料檔中，鍵入及譯碼錯誤不考慮時，各項疾病之診斷都相當明確。但是輕微或致死的傷病即未列入。

未來要促進全民健康保險資料庫之學術應用，可從下列三方面努力：

- (一) 加強資料庫之完整性與正確性：無論在電腦硬體設備或軟體程式方面，都必須予以擴充強化。各項資料檔之內容、欄位設定、譯碼規則、鍵入偵錯都應該建立詳盡可行的工作手冊，再配合資訊管理人員的專業訓練，始可確保全民健康資料庫之完整性與正確性。進一步則定期連鎖比對其他全國性之醫療保健資料檔，例如癌症申報登記檔，全國死因檔等，以檢視不同檔案間之出入。必要時宜委託研究機構分析和評估資料庫之系統與作業之品質保證。
- (二) 加強學術應用並開放資料連鎖：全民健康保險資料的學術應用可以從學術研究公用檔的建立和個人資料連鎖的開放著手。公用資料檔並不包括任何個人或醫療院所的身份資料，所以並無保密性之困擾。但是個人資料的連鎖比對，則必須有個別身份資料。任何健康保險資料都應該確保其隱密性，不可對外公佈任何個人資料。為了加強學術應用，特別是資料連鎖比對之開放，又不得不具有個人身份辨證的方法。國外經常採用的是身份證字號之轉碼，以保存個人身份辨證號碼，但又不曾公開個人隱私。

表五 重大傷病證明領用檔、癌症申報登記檔和全國死因檔當中子宮頸癌之患者數

| | 癌記登記檔 | 重大傷病檔 | 全國死因檔 | 人數* | 百分比 |
|----|--------|-------|-------|--------|-------|
| | 有 | 有 | 有 | 241 | 1.5 |
| | 有 | 有 | 無 | 3,676 | 23.4 |
| | 有 | 無 | 有 | 138 | 0.9 |
| | 無 | 有 | 有 | 236 | 1.5 |
| | 有 | 無 | 無 | 5,997 | 38.1 |
| | 無 | 有 | 無 | 5,176 | 32.9 |
| | 無 | 無 | 有 | 273 | 1.7 |
| 總數 | 10,052 | 9,329 | 888 | 15,737 | 100.0 |

* 聯結三檔並去除重複後之人數

未來在開放資料連鎖時，轉碼工作必須統一由特定之學術應用資料中心來進行。

- (三) 整合醫療保健資料並設立衛生資訊中心：國內與醫療保健相關之資料檔都相當龐雜，且由不同負責單位保存管理，像全民健康保險局、衛生署、環境保護署、勞工安全委員會等，都建立有與國民健康息息相關之資料。若能加以整合和集中管理，勢必可以提供學術應用上的重要參考。基於專業和成本之考量，有必要設立全國衛生資訊中心，以負責推動相關資料之學術應用。

參考文獻

1. Beasley RP, Hwang LY, Lin CC, Chien CS. Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22707 men in Taiwan. *Lancet* 1981; **2**:1129-33.
2. Chen DS, Hsu HM, Sung JL, et al. A mass vaccination program in Taiwan against hepatitis B virus infection in infants of hepatitis B surface antigen-carrier mothers. *JAMA* 1987; **257**:2599-603.
3. 衛生署，中華民國八十三年癌症登記報告。台北市：行政院衛生署，1997。
4. Chang MH, Chen CJ, Lai MS, et al. Nationwide hepatitis B vaccination and the incidence of hepatocellular carcinoma in children in Taiwan. *New Engl J Med* 1997; **336**:1855-9.

USE OF NATIONAL HEALTH INSURANCE DATABASE IN ACADEMIC RESEARCH: EXPERIENCES FROM ANALYSIS OF MAJOR DISEASE CERTIFICATION PROFILE

CHIEN-JEN CHEN, LIH-HWA LIN

The database of health insurance is important for the planning, monitoring and evaluation of health care programs. It is also useful for the assessment of health need and exploration of risk factors for diseases. After the implementation of national health insurance program in Taiwan, a large-scale database been collected for administrative and managerial purposes. It may also provide essen-

tial information for academic researches. Based on our experiences in the analysis of major disease certification profile of national health insurance program, we describe the applicability, essential characteristics, limitations and future perspectives of national health insurance database. (*Chin J Public Health. (Taipei): 1997; 16(6): 513-521*)

Key words: *health insurance database, applicability, essential characteristics, limitation.*

Graduate Institute of Epidemiology, National Taiwan University.