# 美國癌症登記及老人醫療保險資料庫之發展與應用—論台灣癌症登記與健康保險聯結資料庫之可行性

潘憶文[1,*]　簡君儒[2]　施雅真[3]

根據美國國家癌症研究中心的統計，到2011年1月止，運用美國癌症登記(Surveillance, Epidemiology, and End Results Program，簡稱SEER)及老人醫療保險(簡稱Medicare)行政申報資料庫連結資料庫發表在同儕審查期刊的文章已超過650篇以上，在癌症醫療品質及成本相關的研究有卓著的成果。我國全民健康保險自1995年開始辦理，全民健康保險資料庫業已廣泛應用，另一方面我國癌症登記資料庫也漸趨成熟。若能將兩者結合，應可發揮相當程度的綜效。本文的目的是希望藉由他山之石，提供我國未來整合癌症相關資料庫供臨床、學術研究應用及政策評估之參考。本篇文章將針對以下內容作介紹：資料蒐集之行政層級及架構、資料庫的結構、附加應用軟體及程式介紹、資料庫驗證(Validation)與申請費用、資料庫的使用限制與病人個別資料保護，並以乳癌為例，說明該資料庫在臨床研究、醫療品質及癌症治療成本研究方面之應用，並提供建議。(台灣衛誌 2012；**31**(**4**)：299-310)

關鍵詞：癌症登記、醫療保險、行政申報資料庫

## 前　言

由於資訊科技的發達，次級資料的運用，尤其是健康保險申報資料庫，成為醫療服務研究的趨勢。美國癌症登記系統(Surveillance, Epidemiology, and End Results Program, 簡稱SEER)自1973年開始搜集癌症病例[1]，目前涵蓋全美老年人口的26%。該資料庫於1991年開始與美國老人醫療保險(簡稱Medicare)行政申報資料庫作連結，供學術研究使用，其應用範圍遍及於癌症預防、治療與臨終照護，並涵括醫療品質與成本的研究。

我國自1995年開辦全民健康保險，超過99%的納保率，使得全民健康保險資料庫(National Health Insurance Research Database, NHIRD)成為我國最大也最完整的母群體醫療保險資料庫；我國癌症登記資料自1979年開始蒐集，1996年更進一步由衛生署委託癌症登記小組開始進行較完整的癌症流行病學資料收集[2]。目前這兩個資料庫分屬兩個不同單位管理，並未作常態性與系統性的連結。藉由美國癌症登記與老人醫療保險資料庫(以下簡稱SEER-Medicare)發展的經驗，可供我國未來健康資料庫有效應用於學術研究與政策評估之參考。

[1] Information Services: Health Economic and Outcomes Research, Mckesson Specialty Health
[2] 中國醫藥大學醫學系放射腫瘤科
[3] 芝加哥大學醫學系醫院醫學科
[*] 通訊作者：潘憶文
聯絡地址：10101 Woodloch Forest Dr., The Woodlands, TX 77380, U.S.A.
E-mail: iwen.pan@mckesson.com
投稿日期：101年1月3日
接受日期：101年4月13日

## SEER-Medicare介紹

### 一、行政層級及架構(如圖一)

SEER-Medicare是由美國國家癌症研究所(National Cancer Institute, NCI)、SEER登記單位、及老人醫療與失能醫療保險服務中心(Centers for Medicare and Medicaid Services, CMS)等三個單位合作產出。

SEER-Medicare每三至四年連結更新一次，由SEER地方登記單位提供個人識別碼(簡稱ID)連結到Medicare的主檔ID後，將兩部分資料連結。第一次是1991年，最近一次是2009年，提供1973年到2007年的癌症登記資料，在保險就醫資料的部分，則蒐集到2009年為止。

由於臨床登錄系統時有變動，在資料庫更新時，基本上仍保留原變項名稱並加註變項適用的期間，如DAJCC為癌症登記分期版本六，僅適用於2004年以後的資料；AJCCSTG為癌症登記分期版本三，僅適用於1998年到2003年間的資料等，提供研究者在資料分析時完整的資訊，也避免資料被誤用。

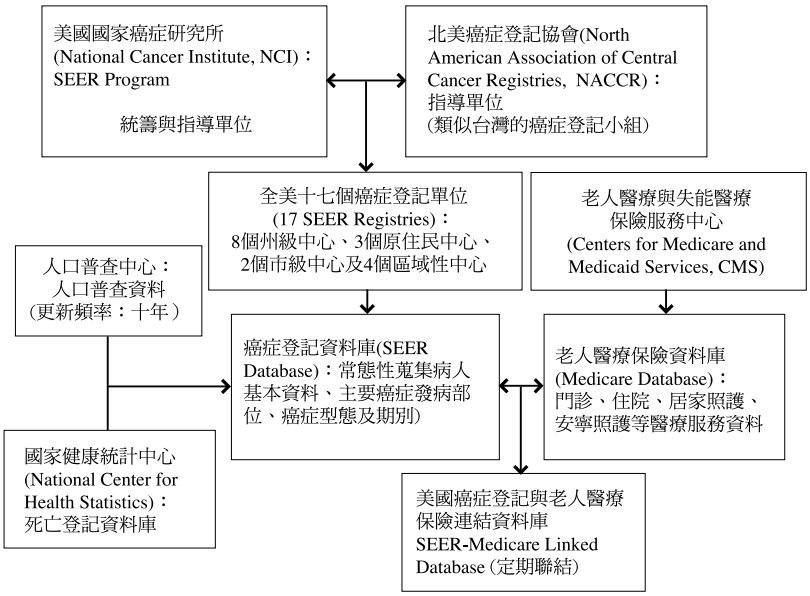
### 二、資料庫的結構

SEER-Medicare包括病人基本資料檔、Medicare申報檔、醫院基本資料檔、、及非癌症病人資料檔。各資料檔主要變項如表一，分述如下：

(一) 病人基本資料檔(Patient Entitlement and Diagnosis Summary File, PEDSF)：

這是核心檔案，一個病人僅有一筆紀錄並有個別經加密後的ID，所有的其他資料檔均需由此開展後，向外連結，並可作長期追蹤分析。資料檔的內容有幾個部分，一是SEER擷取來的；二是Medicare相關的，並有與全國死亡登記資料庫連結確認的死亡年月日及病人居住地人口普查及社經地位等資料。此外，最重要的部分，是SEER中鉅細靡遺的癌症臨床診斷資訊。

SEER地方登記單位在每個癌症病人進入系統後，便會持續蒐集他們的病史，並與Medicare資料完整連結。以2002年診斷為乳癌的67歲病人為例，病人雖然在2002年才進入癌症登記的系統，但在SEER-Medicare的資料檔中，會有病人自滿65歲後進入Medicare，也就是2000年後的所有就醫記錄(包括與癌症相關或非相關的)，




圖一　美國癌症登記與老人醫療保險聯結資料庫行政層級與架構

表一　SEER-Medicare資料檔主要變項一覽表

| 檔案名稱 | 主要變項 | 資料來源 |
|---|---|---|
| 病人基本資料檔(PEDSF) | 1. SEER 資料庫擷取來的包括病人的年齡、性別、婚姻狀態、死亡年月(不包含日)、出生地、種族、死因、居住地(郡別)、病人的郵遞區號(需要特別申請)及最多到十個的診斷別和每一個診斷當時的年月、病人年齡及居住地等相關資料<br>2. 被保險人出生年月日、保險資格、納保原因、有效保險期間，另有與全國死亡登記資料庫連結確認的死亡年月日，並有病人居住地區地理資訊、郡級(County level)的人口普查及社經地位等基本資料(非個別病人的資料)<br>3. 原發部位(Primary site)、後續發病的部位(至多到10個診斷)、組織類型(Histology)、側性(Laterality)、性態碼(Behavior Code)、腫瘤大小(Tumor Size)、癌症期別(staging)、區域淋巴結侵犯數目、AJCC癌症分期版本、腫瘤註記(Tumor Marker)、手術方式(包括乳房重建手術)、放射治療方式、和化學治療方式等 | 1. SEER<br>2. Medicare<br>3. SEER |
| 住院檔(MEDPAR) | 共同變項*、入出院日期、住院天數、10個ICD 9 CM診斷碼、6個ICD 9程序碼、DRG碼、洗腎病人註記及細項：如輸血費、部分負擔、其他保險給付及最後核付的金額等 | Medicare |
| 門診檔(OUTPAT) | 共同變項*、申報資料型態、服務起(迄)日、服務機構型態別、服務型態分類、申報費用、給付金額、主要保險(非老人保險)給付金額及代碼、病人出院狀態、醫療服務程序代碼(Health Care Financing Administration Common Procedure Coding System，HCPCS)、HCPCS細項代碼、收入中心(Revenue Center)單位代碼、資料年份、10個ICD-9診斷碼、6個ICD 9程序碼(ICD 9 CM Procedure code)、手術醫師代碼、主治醫師代碼 | Medicare |
| 醫師及醫事人員服務檔(NCH) | 共同變項*、服務日期、服務醫師(醫事人員)代碼、服務起(迄)日、服務機構型態別、服務型態分類、申報費用、給付金額、病人出院狀態、醫療服務程序代碼(HCPCS)、HCPCS細項代碼、單項服務ICD-9診斷碼、ICD-9主診斷碼Part B扣除額、部分負擔、申報金額、給付金額、醫事服務申報金額 | Medicare |
| 醫材檔(DME) | 共同變項*、申報資料型態、服務起(迄)日、服務機構型態別、服務型態分類、申報費用、給付金額、主要保險(非老人保險)給付金額及代碼、病人出院狀態、醫療服務程序代碼(HCPCS)、HCPCS細項代碼、NDC碼(藥品品項代碼，僅包括口服藥)收入中心(Revenue Center)單位代碼、資料年份、8個ICD-9診斷碼、單項服務ICD-9診斷碼 | Medicare |
| 居家服務檔(HHA) | 共同變項*、申報資料型態、服務起(迄)日、服務機構型態別、服務型態分類、申報費用、給付金額、主要保險(非老人保險)給付金額及代碼、病人出院狀態、醫療服務程序代碼(HCPCS)、HCPCS細項代碼、收入中心(Revenue Center)單位代碼、資料年份、10個ICD-9診斷碼 | Medicare |
| 安寧照護檔(Hospice) | 共同變項*、Hospice起始日期、ICD-9CM診斷碼(至多可以有10 個診斷碼)、醫療服務程序代碼(HCPCS)、收入中心(Revenue Center)單位代碼、病人出院狀態、主要保險(非老人保險)給付金額及代碼、老人保險給付金額等 | Medicare |
| 醫院檔(Hospital File) | 醫院識別碼、醫院評鑑別、權屬別、型態別、區域別、是否為醫學院附設醫院、是否可收住院醫師、床數、急性洗腎床、燒傷中心床、急診服務、居家照護、安寧照護、護士人數及其他設施等 | The Healthcare Cost Report(HCRIS) and the Provider of Service (POS) by CMS |

*共同變項：包括病人識別碼、保險身份別、癌症登記區域別、病人所在州碼、病人所在郡碼、病人的郵遞區號(需要特別申請)、醫院識別碼。

這提供醫療服務研究很重要的資訊，研究者可以病人發病前一年的就醫資料去計算共病(comorbidity)對疾病的嚴重度(illness of severity) 或資源耗用的影響。

(二) Medicare申報檔：

包括住院檔(MEDPAR)、門診檔(OUTPAT)、醫師及醫事人員服務檔(NCH)、醫材檔(Durable Medical Equipment, DME)、居家服務檔(Home Health Agency, HHA)和安寧照護檔(Hospice)，最新發行的2007版，增加處方藥附加保險(Part D event)檔，內容分別說明如下：

1. 住院檔：每一筆資料，代表病人一次住院的相關資料，住院給付採DRG包裹式給付，但不包含醫師費及其他特殊醫事服務費，該等資料會列在醫師及醫事人員服務檔。
2. 門診檔：泛指在醫院、診所、或其他場所門診申報資料，但一樣不含醫師費及其他特殊醫事服務費用。對於一筆申報資料而言，會有許多筆不同的資料細項，有點像台灣NHIRD的門診費用檔加上門診費用醫令檔，在資料使用上需十分小心，由於各地方Medicare委外處理申報的單位不同，資料的處理方式有些許的差異，有的費用是指整筆申報資料的總合，有的卻是單筆醫令項目的費用，如果研究主題與癌症成本有關，資料處理程序需十分的嚴謹。
3. 醫師及醫事人員服務檔：醫療服務發生的場所，可以是診所，也可以是醫院或是其他醫事機構。剛開始只納入醫師費的部分，後來隨著各項醫事人員服務費用陸續分列，已納入其他醫事人員服務費用資料。類似台灣的全民健康保險給付，將醫師診療費及藥事服務費分列。
4. 醫材檔：涵蓋了所有的醫用耗材、注射針劑、及輔具等的申報資料。在2000年後的癌症研究，逐漸重視這個檔案，原因是，某些注射用化學治療藥物，如果改以口服藥提供給病人，Medicare也納入給付，在此之前，Medicare不給付病人口服藥物。據分析在2005年時，已有10%的口服用化學藥物治療費用在該檔案當中呈現。
5. 居家服務檔：基本內容與醫師及醫事人員服務檔類似，但主要是居家照護提供的費用申報檔。
6. 安寧照護檔：基本內容與住院檔類似，但提供的服務為安寧照護。
7. 處方藥檔：2006年開始，Medicare開始提供處方藥附加保險(PART D)，大約有百分之六十的Medicare被保險人加入這項附加保險[3]，但符合條件的研究對象：癌症病人必須同時擁有 PART A(住院險)及PART B (門診險)且未參加醫療維護組織(Health Maintenance Organization, HMO)，只有百分之五十的人有PART D[4]，這項資料在2007年的資料檔中已納入。

(三) 醫院基本資料檔：

NCI曾經針對醫院特性變項與SEER-Medicare連結研究的可行性發表文章在Medical Care期刊上[5]。首先，是醫療服務量的評估，每一筆申報資料都有個別醫院加密的單一ID，如果研究的主題是針對老年人常見的疾病或癌症，雖然我們無法得到單一醫院的全部服務量，但是有可能推估到較真實的醫療服務量在醫院間的分佈，可以作為研究參考。但若研究的主題，如白血病或前列腺癌根治術，均為好發於年輕族群的疾病或適用於該族群的治療方式，那麼單以SEER-Medicare的資料，並無法推估到醫院正確的服務量。其次，醫院特性與研究主題相關的研究，透過特殊申請取得解密後的醫院ID，可以直接連結SEER-Medicare與CMS網站上公開的醫院檔。另外，也可以連結到美國醫院協會的醫院年度調查檔，但程序較為複雜，因醫院ID和前者不同，需另向美國醫院協會申請醫院ID對照檔，再加以連結。這兩項資料，均可以得到醫院的教學醫院屬性、病床數、權屬別、醫院所在地區別等相關的資訊。整體來說，由於醫院向CMS提供的資訊與他們申報給付金額的計算方式有關，因此CMS的資訊比較精確。2007年版新增醫院合併及結盟檔，所有連鎖醫院可藉由該檔案中的群體代碼整合。

(四) 非癌症病人資料檔

這個部分資料，主要是用於病例對照研究(case-control study)，針對癌症病人居住區域，從Medicare隨機抽樣另外百分之五的非癌症病人，這些資料，可以作為成本及比較性研究的對照組。

三、 輔助性分析及程式碼

SEER-Medicare的官網上，有提供研究者參考的基本資料分佈[6]、常用ICD-9處置碼的彙整(像是化學治療、放射線治療和癌症篩檢等)[7]、與診斷日期及治療日期的定義[8]等，希望研究者在各自的研究進行中有共同的語言。此外，也有許多應用程式供研究者使用。這些程式，是根據過去研究者的經驗所製作的。可以提供研究者對於數據及研究方法上可比較的基礎，讓研究者快速的得到基本數據(preliminary results)供申請研究計畫參考。

(一) SAS讀檔程式：

SEER-Medicare資料原始檔案是txt檔案，因資料庫本身變項很多，為避免沒有經驗的研究者資料讀取錯誤，在SEER-Medicare的網站上[9]，有提供完整的SAS程式碼，供讀取檔案用，這也讓研究者的研究結果不會因為個別資料的讀取方式錯誤而造成推論的謬誤。

(二) 計算共病(Comorbidities)[10]程式：

該程式是以SAS程式語言，運用Charlson's的演算方式[11]，並以SEER-Medicare的變項為基礎寫出來的巨集程式。其中有許多變項是可以依據資料庫及研究的內容而調整。包括像病人的保險資料來源，是門診、住院或醫師及醫事人員服務檔的資料、計算共病的期間、和ICD-9CM的診斷變項數目有多少，該程式可推廣應用在任何保險資料庫作癌症研究的共病計算。但因其納入計算的診斷不包括癌症，因此，如果推廣至應用到其他疾病研究時，需加回癌症的診斷碼。

(三) 其他軟體[12]：

其他軟體還包括Joinpoint[13]軟體，可用來繪製癌症長期趨勢圖，與簡易統計軟體、盛行率預測軟體等，詳情可以參考SEER的網站。(http://surveillance.cancer.gov/software/)

四、 資料庫的限制與使用說明[14]

(一) 研究對象及資料限制：

研究對象僅限於SEER癌症登記區域且為Medicare被保險人，此外，下列限制會影響到研究成果的應用及解釋，包括：一)病人參加HMO保險者會被排除在外，通常HMO的保險給付是採包裹式給付(lump sum)，並沒有細項的給付內容，無法透過申報資料得到所需的資訊。此外，在研究期間沒有完整Part A基本住院險和Part B附加門診險者，也需排除。二)榮民醫院的資料是獨立的，所有在其就診的資訊，都不在SEER-Medicare內。三)研究者無法從SEER-Medicare取得病人使用其他保險時所接受的服務資訊。四)無法從SEER-Medicare取得病人自費項目的資訊。

(二) 個人隱私權的保障：

研究者在申請SEER-Medicare時，需出示通過所屬機構人體試驗與倫理委員會(Institutional Review Board, IRB)審查的證明，並填具SEER-Medicare使用同意書。同意書主要的內容在於保障個人資料的安全，包括病人、醫師及個別醫院的隱私。資料庫中病人ID、病人居住地的郵遞區號、醫師及個別醫院的ID都是經過加密的，無法經由該ID連結到任何相關的資料庫。

(三) 解密資料的申請程序：

此程序十分繁瑣且嚴謹，通常是為了連結到外部資料以取得額外的資訊作研究，如病人居住地與就醫場所的距離對病人醫療可近性的影響研究。申請程序分為兩部分：首先，必須向NCI提出特殊需求申請，包括具體研究計畫及使用說明；其次分為三類，一是醫院，NCI可依計畫需求提供解密的醫院資料檔，二是醫師，研究者可在取得SEER-Medicare後，提供所需連結的加密醫師ID清單，提供給NCI委託的資料處理公司，由該

公司解密後連結到美國醫學會的醫師檔，再重新加密後，再將檔案交給研究者，所需的費用均由研究者負擔，三是病人資訊，僅限於病人居住地的郵遞區號，NCI視計畫需求，提供研究者SEER地方登記單位的聯絡方式，但需由研究者自行取得各單位同意後，才有可能取得資料，申請過程十分漫長。

(四) 資料庫保存及儲存：

資料庫取得後，僅得保存五年，五年後必須銷　或申請延期。通常SEER-Medicare的資料檔均是以光碟快遞到計畫負責人手上，而光碟本身是經過加密的，沒有密碼及相對的軟體無法開啟檔案。資料處理中心會在快遞光碟後，打電話給計畫負責人，口頭告知解密的密碼　，以避免資料庫在傳遞過程中遺失，造成資料外洩。NCI也要求研究者必須將資料儲存在有密碼及安全設定的電腦環境中使用，不得儲存於隨身可攜式的媒體中。

(五) 資料庫的使用授權：

資料庫僅授權予當初申請計畫者的機構使用，使用範圍必須涵蓋在原始申請的研究計畫中敘明，如有額外需要，應另提出計畫及說明。而資料庫的使用者也以最少(minimum)為原則，以與計畫相關的研究人員可參與使用。

(六) 研究成果的發表：

使用SEER-Medicare的研究成果，包括在各種雜誌上發表的文章，在投稿前，均需先送一份初稿經NCI審核同意後才得送出，其審查重點主要是確保個人資料的隱私被保障及發表內容是否與當初申請資料庫使用目的相符。通用的原則是，在任何表格的觀察值少於11者，均需用標記取代，不得列出實際的數字。

## 五、資料庫申請費用

申請費用的部分[15]，以美國的物價來說，SEER-Medicare的費用並不算高，但換算成台幣仍是一筆不小的研究支出。舉例來說，以單一癌症別來計價，假設研究者想研究1991年到2009年乳癌病人醫療費用的使用趨勢，大約需新台幣十萬元左右。

## 六、訓練課程

NCI每年舉辦一到兩次兩天一夜SEER-Medicare免費的訓練課程[16]，包括資料庫的結構，及Medicare相關申報資料的說明。藉由有許多實務操作經驗的講師介紹，提供研究者非常完整且實用的課程內容。

## 使用案例說明

乳癌研究為SEER-Medicare發表文獻的大宗，發展已十分成熟。如以乳癌(Intensity Modulated Radiation Therapy, IMRT)[17]研究為例，該治療盛行於2000年以後，可先設定研究的時間範圍，作為起始，使用2005年版的SEER-Medicare，分幾個階段以取得可分析的資料檔(analytic database) ，請參考表二。步驟一到九所需要的資訊，都可以從病人基本資料檔中直接刪選，步驟十至十三則需要連結基本資料檔及其他保險申報檔才能作確認。其中步驟一，為因應Medicare的設計而訂定的刪選條件，其他條件則需由臨床醫師確認刪選條件的合理性與重要性。另一方面，透過郡代碼(county codes)，SEER-Medicare也可與區域資源檔(Area Resource File, ARF)聯結[18]，取得醫事人力、醫院、貴重醫療儀器、區域內教育水準、與家庭收入低於貧窮線的比例等資料的分佈，該等資料可以作民眾就醫可近性(accessibility)及醫療市場競爭(market competition among hospitals)的評估。在IMRT保險給付成本方面，住院與門診需分別以前瞻式價格指標和Medicare經濟指標，將不同年度費用標準化成同一年的物價指數，並需依地區執業成本作調整。個別病人治療的總成本，則是依研究者的觀點(perspective)而定，可以放射線治療相關的保險給付作評估，也可以病人全部療程相關給付作評估。該研究結果發現，IMRT利用率與成本與區域給付政策有關。

表二　研究對象的刪選原則及個案數

| 步驟 | 刪選條件 | 個案數 | 使用的資料檔 |
|---|---|---|---|
| 1 | 選入個案為66歲以上之女性，診斷為乳癌且其診斷日期在2001年到2005年之間並診斷前後各一年內具有完整加保資料者 | 65,820 | 1.病人基本資料檔 |
| 2 | 如果個案在乳癌診斷後12個月內死亡者排除 | 65,243 | 1.病人基本資料檔 |
| 3 | 如果個案第二個癌症診斷在初診後12個月內者排除 | 62,209 | 1.病人基本資料檔 |
| 4 | 排除有癌症病史者 | 58,368 | 1.病人基本資料檔 |
| 5 | 組織類型與上皮起源不一致者(histology not consistent with epithelial origin)及肉瘤、淋巴瘤、及其他非上皮組織類型者 | 57,169 | 1.病人基本資料檔 |
| 6 | 排除原位乳葉癌 | 56,625 | 1.病人基本資料檔 |
| 7 | 排除癌細胞轉移者 | 54,868 | 1.病人基本資料檔 |
| 8 | 排除未知期別者 | 54,264 | 1.病人基本資料檔 |
| 9 | 排除未經病理確認者 | >54,253 | 1.病人基本資料檔 |
| 10 | 排除未經手術的病人 | 53,438 | 1. 病人基本資料檔<br>2. 住院檔<br>3. 門診檔<br>4. 醫師及醫事人員服務檔 |
| 11 | 排除病人沒有以CPT code記錄放射線治療處置者* | 27,600 | 1. 病人基本資料檔<br>2. 住院檔<br>3. 門診檔<br>4. 醫師及醫事人員服務檔 |
| 12 | 排除病人接受 brachytherapy治療者 | 26,747 | 1. 病人基本資料檔<br>2. 住院檔<br>3. 門診檔<br>4. 醫師及醫事人員服務檔 |
| 13 | 排除病人有多於40項處置記錄者 | 26,163 | 1. 病人基本資料檔<br>2. 住院檔<br>3. 門診檔<br>4. 醫師及醫事人員服務檔 |

資料來源：[17]

*CPT code: Current Procedural Terminology代碼，為美國醫學學會(American Medical Association)針對各項醫療處置因應各類保險給付所製訂的現行給付代碼。

## 討　論

根據NCI的統計[19]，到2011年11月止，運用SEER-Medicare發表在同儕審查期刊的文章共計有677篇，其中2006年至2011年間就有499篇，已進入豐收期，在癌症醫療品質及成本相關的研究有卓著的成果。由上述說明，我們可以了解NCI為了推動癌症次級資料研究的努力，有許多值得我國學習或借鏡的地方，分述如下：

### 一、資訊透明與公開

SEER-Medicare與台灣NHIRD一樣，都是使用公共資源建立的資料庫，應以社會大眾最佳利益為出發點與使用者付費的精神，善加利用。因此，基於資料共享，資訊透明的精神，2002年在SEER-Medicare發行的初期，NCI曾在Medical Care 40卷第八期的附

冊，發表一系列的完整介紹，包括資料庫總覽[1]、癌症成本的推估[20]、手術[21]、化療[22]與放射線治療[23]的應用、及醫院檔的使用[6]等等，提供資料庫清楚的輪廓供研究者參考[1,6]。相對來說，台灣癌記登記資料庫及NHIRD尚未整合，即以NHIRD而言，雖然已應用廣泛，提供的基礎研究資訊，近年來亦有大符的進步，但仍有改善的空間。

## 二、資料庫驗證(Validation)

對於研究的品質而言，資料庫本身的品質是研究成敗的關鍵。因此，資料庫的正確性、邏輯性及完整性十分重要。資料庫在連結後，需要經過許多繁瑣的驗證除錯過程，以確保資料沒有邏輯上的繆誤或是前後不一致的情形。

驗證的方式有很多，可使用電腦程式作基本邏輯的驗證、將保險資料庫與個別醫院的資料分佈統計作比對驗證、以及使用者的回饋。最後一項，往往扮演資料品質控管很重要的角色。筆者本身在研究期間，就曾發生NCI在資料公開後，又發出緊急通知要求更新某年度的醫師及醫事人員檔，即是NCI接到研究者回饋發現數據異常後的處理。

整體而言，SEER本就是為了流行病學研究而建立的資料庫，相較於Medicare的本質是健康保險的申報資料，前者的正確性及完整性自然會比後者好，但各有其優缺點。以死亡註記而言，Medicare通常在每個月底就會與官方死亡檔資料連結，以確保被保險人的資格，但沒有死因的資料；SEER有死因的資料，但死亡註記更新，可能與實際發生的時間有落差。

相對來說，關於NHIRD的資料驗證，目前僅有二篇文獻發表[24,25]，但是以中風和糖尿病為案例。資料庫中雖有變項可辨識住院死亡案例，但並不完整，研究者仍需向健康資料加值應用協作中心(The Collaboration Center of Health Information Application, CCHIA)另行申請相關資料。

## 三、資料庫的應用

自1991年至今，學者應用SEER-Medicare所作的研究，涵蓋了癌症防治的各層面，從癌症早期預防篩檢、到診斷治療、治療成效、治療成本、甚至到癌症臨終治療模式的分析等。以癌症部位來看，發表在期刊上的前三名依序是乳癌、大腸直腸癌、攝護腺癌[26]。以研究主題來看，發表量最多的前五名是治療方式、治療成果、醫療差別性、研究方法、癌症經濟、以及醫療體系，其他還包括癌症篩檢、危險因子、存活分析、臨終治療、醫療品質、癌症預防等。自開放供研究使用以來，已成為美國癌症防治的重要研究資源。台灣目前以健保資料庫發表在國際期刊癌症相關文獻自2002年至 2011年10月約有53篇，其中2011年就有22篇，仍有很大的進步空間。

## 四、個人資料保護與學術研究的平衡

個人資料保護與學術自由之間的平衡點，是需要被規範，更需要研究者共同遵循的，這是屬於學術倫理的一環。NCI針對個資保護，如前所述有因應不同對象、申請資料程序、資料保存及儲存方式、和研究資料發表等的具體作法。保護的對象包括病人和醫療服務提供者(個人及機構)等。就病人而言，除了基本的ID外，病人居住地的地理資訊，僅到郡級(相當於台灣的縣市別)，其他研究所需之重要社經地位變項，均僅提供病人居住所在地普查單位(可能是郵遞區號或是鄉鎮級的單位)十年一次的人口普查資料，且無法回推其地理資訊。醫療服務提供個人，包括醫師及其他醫療專業人員等，在資料庫中僅見其被轉碼後的ID。上述兩類，研究者如需進一步資訊，均需依前章所述提出特別申請，即使如此，研究主題及需求不清，申請仍十分容易被駁回。加上研究者在研究結果投稿發表前，均需經過NCI審核，為個人資料建立了相當完整的保護網。至於醫療機構的資訊取得雖相對容易，但亦僅限與研究有關的醫院病床數、服務科別、住院醫師收訓、教學醫院屬性等資訊。NCI

SEER-Medicare主要負責人，Dr.Warren[1]曾經在研習課程當中一再的強調，資料庫的整併，是許多單位合作與各界精英討論協調的結果。各單位願意授權開放資料庫使用，實屬不易。因此，在個人資料的保護上，必須作到百分之百的滴水不漏，才能繼續獲得各機構的支持與配合，讓相關的研究延續下去。這是一項承諾，20年來NCI也的確嚴格把關確保了這項承諾。

五、SEER-Medicare的研究侷限

SEER-Medicare的限制來自於兩個資料庫本身的侷限。SEER為NCI的一項大型研究計畫，僅涵蓋美國26%的人口，地理區域僅包括八個州、三個原住民保留區、二個市、和四個區域中心，雖號稱具全國代表性，但實際上仍有相當大的差距。由於Medicare為老人醫療保險，因此研究對象也僅限於老年人口，不包括年輕族群。此外，Medicare並未涵蓋所有的醫療費用，尤其在藥品方面，部分的癌症口服藥是不給付的；病人部分負擔及其私人保險負擔的費用也無法計算。相較這幾項限制，NHIRD涵蓋了全台百分之九十九以上的人口，不分年齡及地區，且醫療費用給付項目較為完整，有相對的優勢。

## 對台灣未來發展癌症研究資料庫之建議

過去一二十年來，癌症死亡率一直位居台灣十大死因之首，投注資源在癌症研究上，實屬必要。保險申報資料庫的研究雖有其限制，但與臨床試驗相比，成本較為低廉。較美國而言，台灣現有的癌症研究資料庫發展條件皆已具備，且保險資料並不限於老年人口，具有普遍性的特質，在學術研究發展上具有相當的優勢，只差整合。基於此，提供以下的建議：

一、定期連結並整合癌症相關資料庫，以利癌症相關研究

行政院衛生署自2011年設立CCHIA，開放研究者申請健康統計資料加值應用，基於個人隱私保護的前提下，多方資料庫連結，只能以“現場作業”，並僅能攜出統計分析結果。相較於常態連結的資料庫，若個別研究者的資料庫連結方式或分析程式有問題，不易稽核或重製。其次，研究者在不同的條件刪選下，產生不同的研究結果亦不容易重覆驗證。此外，資料庫可近性較差，也會增加研究成本。

建議參考SEER-Medicare的作法，每二至三年連結台灣癌症登記資料庫、NHIRD、及死亡登記檔作整合性的連結，建立台灣癌症研究資料庫。其優點是透過多重資料庫的連結，可以達到部分資料庫驗證的效果，建立資料庫的公信力。

二、強化資料庫的驗證，讓國際學術期刊認可

資料庫未經驗證程序，在發表投稿時，常會被編審質疑其研究成果或特別加註該資料庫未經驗證[27]。因此，除了將資料庫整合之外，參考SEER-Medicare的模式，在期刊雜誌上，以不同的主題介紹台灣癌症研究資料庫，也是使資料庫獲得國際學術界認同的好方法。癌症研究資料庫的領域很廣，包括現行最熱門的醫療經濟成本效益分析、藥物經濟學、高科技醫療儀器的使用評估、臨床治療模式的改進、醫療政策的評估與流行病學的追蹤調查等，目前也有研究者利用SEER-Medicare作臨床試驗後之驗證研究。

此外，基本統計數據的公開，包括每年各癌症部位之病人發生率及死亡率等，可作為研究者引用數據的參考。

三、開設訓練課程

目前國家衛生研究院每年定期舉辦NHIRD應用研習會，癌症研究資料庫或許可依資料連結的頻率，隔年舉行。主要的目的，是讓研究者能正確的使用資料庫，避免研究結果及建議的繆誤。訓練內容可包括各種研究的案例及各項資料檔案細部說明及使用計算方式。

四、加強資源共享

SEER-Medicare的網站有兩項很重要的資訊分享，一是分析工具：研究者可由SEER-Medicare官網外部連結至SEER的網站參閱癌症分級譯碼資訊及許多癌症研究相關有用的資訊，二是應用軟體程式，包括資料庫讀取程式、共病計算程式及簡易統計分析程式。以點列式讓研究者非常容易找到研究所需的資訊及特定主題說明。目前我國癌症登記資料庫除透過CCHIA申請現場連結分析外，並未廣泛開放外界申請。NHIRD網站雖已提供許多經驗分享資訊，但就最基本的資料讀取方式的統一：如提供不同軟體的資料讀取程式、相關參考統計數值、或是研究者可使用資源的連結等還有待加強。

研究資料庫屬於公共資源，應由政府推動訂定使用計畫與規範，並促進跨部會與機構的整合，相信對於台灣癌症研究發展上必能有所貢獻。

## 參考文獻


1. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care 2002;**40(8 Suppl)**:IV-3-18.
2. 賴美淑：台灣癌症登記發展與沿革。聲洋防癌之聲 2011；**(132)**：35-9。
   Lai MS. Development of the Taiwan cancer registry. Cancer Bulletin S.Y. Dao Memorial Fund 2011;**(132)**: 35-9. [In Chinese]
3. Health Services and Economics, National Cancer Institute. SEER-Medicare: medicare claims files. Available at: http://healthservices.cancer.gov/seermedicare/medicare/claims.html. Accessed August 25, 2011.
4. Health Services and Economics, National Cancer Institute. Number of part D enrollees. Available at: http://healthservices.cancer.gov/seermedicare/aboutdata/enrollees.html. Accessed August 25, 2011.
5. Schrag D, Bach PB, Dahlman C, Warren JL. Identifying and measuring hospital characteristics using the SEER-Medicare data and other claims-based sources. Med Care 2002;**40(8 Suppl)**: IV-96-103.
6. Health Services and Economics, National Cancer Institute. SEER-Medicare: number of cases table. Available at: http://healthservices.cancer.gov/seermedicare/aboutdata/numcases.html. Accessed November 18, 2011.
7. Health Services and Economics, National Cancer Institute. Procedure codes for SEER-Medicare analyses. Available at: http://healthservices.cancer.gov/seermedicare/considerations/procedure_codes.html. Accessed November 18, 2011.
8. Health Services and Economics, National Cancer Institute. SEER-Medicare: defining the date of diagnosis & treatment. Available at: http://healthservices.cancer.gov/seermedicare/considerations/date.html. Accessed November 18, 2011.
9. Health Services and Economics, National Cancer Institute. SEER-Medicare: SAS input statements. Available at: http://healthservices.cancer.gov/seermedicare/program/sas.html. Accessed November 18, 2011.
10. Health Services and Economics, National Cancer Institute. SEER-Medicare: calculation of comorbidity weights. http://healthservices.cancer.gov/seermedicare/program/comorbidity.html. Accessed November 18, 2011.
11. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;**40**:373-83.
12. National Cancer Institute, Surveillance Epidemiology and End Results, Accessing Datasets & Tools. Analytic software. Available at: http://seer.cancer.gov/resources/related.html. Accessed November 18, 2011.
13. National Cancer Institute, Surveillance Research, Cancer Control and Population Sciences. Joinpoint regression program. Version 3.5.2. Oct 2011. Available at: http://surveillance.cancer.gov/joinpoint/. Accessed November 18, 2011.
14. Health Services and Economics, National Cancer Institute. SEER-Medicare: privacy & confidentiality issues. Available at: http://healthservices.cancer.gov/seermedicare/privacy/. Accessed November 18, 2011.
15. Health Services and Economics, National Cancer Institute. SEER-Medicare: cost of acquiring data. Available at: http://healthservices.cancer.gov/seermedicare/obtain/cost.html. Accessed November 18, 2011.
16. Health Services and Economics, National Cancer Institute. SEER-Medicare training. Available at: http://healthservices.cancer.gov/seermedicare/

considerations/training.html. Accessed November 18, 2011.
17. Smith BD, Pan IW, Shih YC, et al. Adoption of intensity-modulated radiation therapy for breast cancer in the United States. J Natl Cancer Inst 2011;**103**:798-809.
18. Health Resources and Services Administration, U.S Department of Health and Human Services. Area Resource File (ARF): national county-level health resource information database. Available at: http://arf.hrsa.gov/purchase.htm. Accessed October 18, 2011.
19. Health Services and Economics, National Cancer Institute. SEER-Medicare publications by journal & year. Available at: http://healthservices.cancer.gov/seermedicare/overview/pubs_jour_year.php Accessed December 29, 2011.
20. Brown ML, Riley GF, Schussler N, Etzioni R. Estimating health care costs related to cancer treatment from SEER-Medicare data. Med Care 2002; **40(8 Suppl)**: IV-104-17.
21. Cooper GS, Virnig B, Klabunde CN, Schussler N, Freeman J. Use of SEER-Medicare data for measuring cancer surgery. Med Care 2002;**40(8 Suppl)**:IV-43-8.
22. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. Med Care 2002;**40(8 Suppl)**:IV-55-61.
23. Virning BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using SEER-Medicare-linked data. Med Care 2002;**(8 Suppl)**:IV-49-54.
24. Cheng CL, Kao YH, Lin SJ, Lee CH, Lai ML. Validation of the National Insurance Research Database with ischemic stroke cases in Taiwan. Pharmacoepidemiol Drug Saf 2011;**20**:326-42.
25. Lin CC, Lai MS, Syu CY, Chang SC, Tseng FY. Accuracy of diabetes diagnosis in health insurance claims data in Taiwan. J Formos Med Assoc 2005;**104**:157-63.
26. Danese MD. Overview of all SEER-Medicare publications through 2010. Available at: http://healthservices.cancer.gov/seermedicare/overview/seermed_pubs_through_2010.pdf. Accessed August 25, 2011.
27. Lai MN, Wang SM, Chen PC, Chen YY, Wang JD. Population-based case-control study of Chinese herbal products containing aristolochic acid and urinary tract cancer risk. J Natl Cancer Inst 2010;**102**:179-86.

# Development of the TCDB-NHIRD Linked Database: what can we learn from the SEER-Medicare Database in the United States?

I-Wen Pan[1,*], Chun-Ru Chien[2], Ya-Chen Tina Shih[3]

In the United States, the SEER-Medicare data link cancer patients in the SEER (Surveillance, Epidemiology, and End Results) Program with Medicare enrollment to identify cancer patients who are eligible for Medicare and to provide Medicare Claims for these patients. Statistics from the National Cancer Institute show that there are over 650 peer-reviewed publications using the SEER -Medicare database. This database has been the primary data source for health services research in oncology since its inception, and it has made tremendous contributions to policies related to cancer control, treatment, and surveillance in the United States. The National Health Insurance Research Database (NHIRD), established at the initiation of the Taiwan's National Health Insurance program in 1995, is a claims database. The Taiwan Cancer Database (TCDB), based on the Taiwan Cancer Registry, was established in 2002. The creation of a TCDB-NHIRD linked database could produce a SEER-Medicare-like database and has the potential to become an invaluable resource for cancer as well as policy researchers in Taiwan. This paper introduces the SEER-Medicare database, including data collection, software, applications of the database, data elements, structure, validation, the data request and application process, limitations of the database, and privacy and confidentiality issues. We then use a previously published breast cancer study to demonstrate the procedures involved in generating research from the SEER-Medicare database, followed by recommendations for the future development of a database that links TCDB to the NHIRD. (*Taiwan J Public Health. 2012;**31**(**4**):299-310*)

***Key words: Cancer Registry, Medicare, Claim database***

[1] Information Services: Health Economic and Outcomes Research, Mckesson Specialty Health, 10101 Woodloch Forest Dr., The Woodlands, TX 77380, U.S.A.
[2] Department of Radiation Oncology, School of Medicine, China Medical University, Taichung, Taiwan, R.O.C.
[3] Section of Hospital Medicine, Department of Medicine, The University of Chicago, Chicago, IL, U.S.A.
[*] Correspondence author.　E-mail: iwen.pan@mckesson.com
Received: Jan 3, 2012　Accepted: Apr 13, 2012

# 評論：台灣癌症登記與全民健康保險連結資料庫之可行性

本期綜論詳細介紹負責發行美國SEER-Medicare資料庫之相關機構、該資料庫之結構、以及相關研究資源與應用上之限制，並提出對台灣未來發展癌症研究資料庫之建議。台灣目前雖尚無類似SEER-Medicare資料庫之作法，由相關主管單位連結台灣癌症登記與全民健康保險資料庫後成為單一研究資料庫，但根據Medline與Web of Science等文獻索引資料庫之資訊，自2000年至2012年期間，以台灣癌症登記資料為研究材料所發表的研究論文，累計已經超過100篇以上。上述這些研究包含單獨使用癌登資料之研究[1-4]、串連癌登資料與健保資料之研究[5-9]、以及癌登資料串連其他資料(例如臨床檢驗檢查資料)之研究[10-12]，其數量雖不及應用SEER-Medicare資料庫之研究，但足見應用台灣癌症登記資料於研究上之潛力。台灣癌症登記資料庫經多年來的努力，已使得資料庫的品質與完整性達到國際標準，且達到族群為基礎的癌症登記系統(population-based cancer registry)，使用者宜定期查看資料庫的相關說明[13]。當使用次級資料進行研究時，需要先瞭解原資料收集的目的，例如健保資料庫主要是申報費用為主，癌症登記資料庫是癌症發生率的趨勢為主。必須考量原始資料間特性與目的的差異，進行必要的調整或妥協，才能在使用次級資料時增加其價值，並據以提出新的研究假說。

將多種研究資料庫串連成為整合性資料倉儲(integrated data warehouse)是未來醫療資訊系統在癌症研究的發展趨勢[14]，經由必要的資訊萃取(data extraction)、轉換(transformation)、整併(integration)、與除錯(cleansing)，提供符合諸如研究者、臨床醫事人員、行政管理者等各種利害關係人(stakeholders)各自不同需求的資料。提出SEER-Medicare資料庫的發展與維護為例，這樣的工作仰賴不同機構的協同合作，落實「資料治理」(Data governance)的精神，以確保不同來源資料在彙整後的正確性、完整性與一致性。

資料庫正確性的驗證為建立整合性資料倉儲時重要考量因素之一。台灣癌症登記資料庫採取多種措施以確保收錄資料的正確性，醫院根據病歷專人登錄的資料，先經過癌登系統的資料勘誤檢核，申報至癌症登記工作小組後，會再由研究人員再行勘誤檢核程序，需要時請醫院癌登同仁再修正申報。此外，衛生署國民健康局委託國家衛生研究院執行的「癌症診療品質認證」，也會進行癌登申報資料的隨機病歷再閱(medical records re-abstraction)，經由比對病歷資料與申報資料的一致性，確保癌症登記資料庫的申報品質[7,8]。然整合多重資料來源成為具有加值效果之大型研究資料庫，其資料正確性更要透過多元化的資料驗證措施[15-19]。

個人資料安全也是在探討串連不同資料庫時需要重視的議題。當病人期待要更先進，更好的藥物及醫療照護的同時，也要求他們的資料要被適當的保護。資料維護單位與病人就如同委託與受託的關係，需要以信任做核心，規劃對資料保密、資料安全、互相的盡責(accountability)、資料的擁有權等配套措施。以美國為例，政府針對HIPAA (Health Insurance Portability and Accountability Act, HIPAA)保護個人隱私法案的實施之時，IOM (Institute of Medicine of the National Academies, IOM)強烈建議國會應授權給衛生福利部，有別於一般的HIPPA的規定，針對所有健康有關的研究提出新的一套保護隱私的規定[20]。例如美國國家癌症研究院(National Cancer Institute, NCI)針對21個癌症中心建立多個資料庫的平台multiple data warehouse: the National

國立台灣大學公共衛生學院流行病學與預防醫學研究所
賴美淑
E-mail: mslai@ntu.edu.tw
聯絡地址：台北市中正區徐州路17號

Comprehensive Cancer Network outcomes database (NCCN)，也是一個整合性資料倉儲成功的例子[14]，它建立大多數的癌症病人長期追蹤資料。該資料庫與SEER最大的不同在於其收錄超過300個變項，包括轉移的位置、生物標記(Bio-makers)、治療的持續性、中斷的原因等。但是除了要將病人資料串到檢體收集的情形之外，這個資料庫運作被允許排除在HIPPA限制之外。

建立癌症研究整合性資料倉儲的長期目標，就是期望促進癌症照護的進步，包括高科技藥物、團隊照護、可近性、民眾付擔的起(affordable)的高品質的癌症照護體系等。但是這樣的工作需要凝聚主管機關、民眾、學界、癌症臨床專業人員與癌登人員的共識，投入相當的人力、物力始能達成，以目前人權團體對於健保資料用於研究用途上仍有異議，健保資料、癌登資料與死因登記分屬不同權責單位，且癌症登記工作經費逐年遭主管機關大幅刪減的現況下，除了排除上述困難，冀望台灣衛生主管單位更積極籌畫『建立癌症研究整合性資料倉儲』的長期目標，先以委員會勾勒出願景與行動藍圖，編列預算，結合現有的癌症登記資料庫、健保資料庫和死亡資料庫之外，還可進一步結合篩檢資料庫、疫苗注射資料、生物檢體組織庫等更多的國家資源，及醫院癌症中心資料庫等之外，同時提供資料倉儲之教育、訓練、推廣與結果分享，期使台灣癌症有創新的研究發展，有助提升國人的健康。

## 參考文獻


1. Chen PT, Kuan FC, Huang CE, et al. Incidence and patterns of second primary malignancies following oral cavity cancers in a prevalent area of betel-nut chewing: a population-based cohort of 26,166 patients in Taiwan. Jpn J Clin Oncol 2011;**41**:1336-43.
2. Chiang CJ, Chen YC, Chen CJ, et al. Cancer trends in Taiwan. Jpn J Clin Oncol 2010;**40**:897-904.
3. Lin CH, Chen YC, Chiang CJ, et al. The emerging epidemic of estrogen-related cancers in young women in a developing Asian country. Int J Cancer 2012;**130**:2629-37.
4. Wu SJ, Huang SY, Lin CT, Lin YJ, Chang CJ,Tien HF. The incidence of chronic lymphocytic leukemia in Taiwan, 1986-2005: a distinct increasing trend with birth-cohort effect. Blood 2010;**116**:4430-5.
5. Chang CH, Lin JW, Wu LC, Lai MS, Chuang LM, Chan KA. Association of thiazolidinediones with liver cancer and colorectal cancer in type 2 diabetes mellitus. Hepatology 2012;**55**:1462-72.
6. Chen WW, Shao YY, Shau WY, et al. The impact of diabetes mellitus on prognosis of early breast cancer in Asia. Oncologist 2012;**17**:485-91.
7. Kuo RN, Chung KP, Lai MS. Effect of the pay-for-performance program for breast cancer care in Taiwan. Am J Manag Care 2011;**17(5 Spec No)**:e203-11.
8. Kuo RN, Chung KP, Lai MS. Re-examining the significance of surgical volume to breast cancer survival and recurrence versus process quality of care in Taiwan. Health Serv Res 2012: doi: 10.1111/j.1475-6773.2012.01430.x. [Epub ahead of print]
9. Tang CH, Pwu RF, Tsai IC, et al. Costs of cervical cancer and precancerous lesions treatment in a publicly financed health care system. Arch Gynecol Obstet 2010;**281**:683-95.
10. Chen CJ, Yang HI, Su J, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 2006;**295**:65-73.
11. Huang YT, Jen CL, Yang HI, et al. Lifetime risk and sex difference of hepatocellular carcinoma among patients with chronic hepatitis B and C. J Clin Oncol 2011;**29**:3643-50.
12. Yang HI, Yeh SH, Chen PJ, et al. Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma. J Natl Cancer Inst 2008;**100**:1134-43.
13. Bureau Health Promotion, Department of Health, Executive Yuan, R.O.C. (Taiwan). Taiwan cancer registry annual. Available at: http://www.bhp.doh.gov.tw/BHPnet/Portal/StatisticsShow.aspx?No=200911300001. Accessed August 13, 2012.
14. Nass SJ, Wizemann T. Informatics Needs and Challenges in Cancer Research: Workshop Summary. Washington, DC: Institute of Medicine, 2012.
15. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care 2004;**42**:1066-72.

16. Earle CC, Nattinger AB, Potosky AL, et al. Identifying cancer relapse using SEER-Medicare Data. Med Care 2002;**40(8 Suppl)**:IV-75-81.
17. Penberthy L, McClish D, Pugh A, Smith W, Manning C, Retchin S. Using hospital discharge files to enhance cancer surveillance. Am J Epidemiol 2003;**158**:27-34.
18. Lamont EB, Herndon JE 2nd, Weeks JC, et al. Criterion validity of Medicare chemotherapy claims in Cancer and Leukemia Group B breast and lung cancer trial participants. J Natl Cancer Inst 2005;**97**:1080-3.
19. Leung KM, Hasan AG, Rees KS, Parker RG, Legorreta AP. Patients with newly diagnosed carcinoma of the breast: validation of a claim-based identification algorithm. J Clin Epidemiol 1999;**52**:57-64.
20. Nass SJ, Levit LA, Gostin LO. Beyond the HIPAA Privacy Rule : Enhancing Privacy, Improving Health Through Research. Washington, DC: National Academies Press, 2009.