

公元 1980 至 1997 年本國衛生署死因資料庫準確性之確認與補正

謝功毅¹ 陳保中^{1,2,3} 王榮德^{1,2,3,*}

GONG-YIH HSIEH¹, PAU-CHUNG CHEN^{1,2,3}, JUNG-DER WANG^{1,2,3,*}

¹ 國立台灣大學公共衛生學院職業醫學與工業衛生研究所，台北市仁愛路一段一號

Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, No. 1, Jen-Ai Rd., Sec. 1, Taipei, Taiwan, R.O.C.

² 國立台灣大學公共衛生學院職業病防治示範中心

Center for Research of Environmental and Occupational Diseases (CREOD), College of Public Health, National Taiwan University.

³ 國立台灣大學醫學院附設醫院內政部

Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

* 通訊作者Correspondence author. E-mail: jdwang@ha.mc.ntu.edu.tw

目標：目前國內將死亡診斷書建立成死因資料庫有衛生署及內政部兩個單位。本研究之目的係使用內政部死因資料庫中正確及完整部份來補正衛生署自公元1985至1997年死亡之身分證字號及戶政資料，同時填補衛生署自公元1980至1997年教育及村里兩變項及衛生署自公元1980至1984年之身分證字號，以作為未來相關研究之參考與應用。**方法：**首先將兩死因資料庫內身分證字號及戶政資料皆相同之個案，先填補衛生署死因資料庫之教育及村里之資料，然後把不可連結之部分以身分證字號或戶政資料各做連結碼，來補正衛生署死因資料庫內缺漏或錯誤之戶政資料或身分證字號，及填補教育及村里之資料。同時建立兩資料庫間互相無法連結之個案資料庫。**結果：**第一部分，自公元1980至1997年以身分證字號與戶政資料為連結碼，經由兩資料庫互連，可連結出1,259,600個個案。連結不上之個案利用身分證字號與戶政資料分別比對後，可補正衛生署之戶政資料有3,168個個案，身分證字號有33,036個個案。另外，兩資料庫可連結之部分，教育及村里變項補正百分比都達98%以上。第二部分，公元1980至1984年，主要填補衛生署身分證字號。以戶政資料包括出生年月日、死亡年月日、縣市代碼、鄉鎮代碼全部變項為連結碼可以填補身分證字號達88.83%。公元1980至1997年，兩資料庫未連結上之個案未作補正者，共有衛生署 46,305個個案及內政部63,157個個案，可能是失去追蹤及延遲申報個案。**結論：**本研究結果可使世代研究及存活分析在使用衛生署的死因資料庫範圍往前推至公元1980年；研究族群之區域分層只能切割到最小單位鄉鎮，但補正後可使之細分到村里。另外衛生署及內政部未連結上之個案可建立檔案作為延遲申報個案及失蹤建檔個案之參考，以便作為將來研究追蹤死亡個案之依據。(台灣衛誌 2002；21(5)：329-338)

關鍵詞：死亡登記、資料庫、衛生署、內政部。

Verification and correction of error for death registration data of the Department of Health R.O.C. between 1980 and 1997

Objective: Both the Department of Health (DOH) and Ministry of Interior (MOI) of Taiwan have established a computerized death registry based on the death certificates of the dead in Taiwan during the past two decades. The purpose of this study is to verify the accuracy of the dataset of death registration for the DOH with the dataset from the MOI. Because the data of the DOH during 1980-84 lacked the identification number (ID no.), our second objective is to establish guidelines for replacing the missing ID no.. **Method:** Firstly, decedents in the two datasets linked by same ID no. and demography including dates of birth, death, and places of living between 1985 and 1997 were identified and codes of education and village taken from the MOI were filed in those of the DOH. Decedents that could only be linked by ID no. or demography between 1985 and 1997, were refiled the demography or ID no. after making sure of the correction in the MOI. Secondly, due to the lack of information of ID no. in the DOH for decedents between 1980 and 1984, we linked the two decedents with social demography including dates of birth and death, area codes of detailed address, and then imputed ID no. into dataset of the DOH based on dataset of the MOI. **Result:** There were 1,259,600 decedents identified from 1985-97 with completely the same ID no. and demographic data. After verification of the data, we refiled the missing demographic data for 3,168 subjects and the missing ID no. for 33,036 subjects into the dataset of the DOH. According to the study results, there were 0.0512% of total subjects with complete overlapping demographic data. Thus, after linking the two datasets with the same complete demographic data, we were able to refile 88.83% of the missing IDs for all the death certificate data of the DOH during 1980-84. There were 46,305 cases in the dataset of the DOH and 63,157 cases in that of the MOI during 1980-97, that were unlinkable. This was probably because they were lost track of or there was a delay in registration. **Conclusion:** This study corrected many errors, filled in ID no. for 1980-84, and imputed information of education and village into the death registration dataset of the DOH. This can be useful for future studies. Beside, the unlinkable cases in the datasets of the DOH and MOI were identified and a separate dataset files for future usage was built up. (Taiwan J Public Health. 2002;21(5):329-338)

Key words: death registration, dataset, Department of Health, Ministry of Interior.

前言

資料庫對於流行病學之研究佔有相當重要的地位，尤其資料庫完整性及正確性對於研究之結果有重大的影響性。目前國內建立許多資料庫如死因資料檔、新生兒出生檔、勞保資料檔、公保資料檔等等。所有的世代研究[1,2]及存活分析[3,4]多必須依賴正確的死亡診斷書。

目前國內將死亡診斷書建立成死因資料庫的有衛生署及內政部戶政司兩個單位，而其兩者之內容有所差異；此點可由其形成之歷史來解釋：衛生署死因資料檔自公元1952年起由台灣省政府衛生處負責收集，資料收集方式由鄉鎮市區公所於每月初至當地戶政事務所影印前一個月之死亡證明書，經註記後送各衛生局初核，核定原因及填註各欄代號[5,6]，自公元1971年起開始利用電腦集中處理，但是沒有輸入死亡者姓名、戶籍地址、職業別及住址，且公元1971至1984年間亦未輸入身分證字號。在美國，財稅資料中對一個人之存活與死亡資料保存極為完整，本國亦然。內政部戶政司公元1978至1997年死因資料電子檔是委託財稅資料中心建立，但是公元1979年該年電子檔遺失，自公元1998之後才由內政部戶政司自力建立；除了輸入身分證字號外，尚有教育程度及村里代碼，但未記載死因代碼及性別。（下文『內政部戶政司』以『內政部』簡稱之）

回顧過去有關使用台灣死亡資料檔案之研究，有理論性的研究[7,8]、死亡原因的驗證[9]及探討社會[10,11]、經濟[12]、疾病[13]等因子研究，但肇因於資料庫完整性的缺乏，每每使資料在應用上有所限制。鑒於衛生署從公元1985年後才建立身分證字號，因此在職業世代暴露追蹤研究中，往往無法探討1985年之前的死亡情形，而導致較早建廠之公司的職業死因研究無法有一個完整之時間趨勢解釋。在環境污染評估方面，暴露分層如果只能細分到鄉鎮為單位，則其精確度實在大有問題。另外，在應用死因資料檔案之

研究，其人數之準確相當重要，而在衛生署延遲申報之部分卻只從公元1996年開始，因此如能加以追溯俾可以減少死亡率之低估，使研究更加準確。

基於上述之各種理由，死因資料庫之補正實在是目前研究相當重要之議題，本研究收集內政部及衛生署的資料來做比對，用較正確的檔案來補正不正確的檔案。因此，本研究的目的如下：1.比較兩個死因資料檔案之不同與其完整性，並加以補正。2.衛生署延遲申報檔案只從公元1996年開始建立，而內政部每年死因資料檔是有建立延遲申報檔，所以用內政部死因資料檔來填補衛生署延遲申報檔案。3.衛生署死因資料中自公元1980至1984年並無建立死亡個案之身分證字號，所以用內政部死因資料檔來填補衛生署之身分證字號。4.用內政部死因資料檔中教育及村里變項來填補衛生署此二種變項。

材料與方法

資料庫所含之變項

本研究範圍為公元1980至1997年，資料來源取自衛生署死因資料檔案(不包含延遲申報個案)計有1,815,134個個案，及內政部死因資料檔案計有1,829,924個個案。衛生署資料含有變項：死亡者身分證字號、性別、縣市代碼、鄉鎮代碼、出生年民國前後、死亡者出生日期、死亡者死亡日期、死亡地點、死亡場所代碼、死亡種類代碼、職業代碼、婚姻狀況代碼、國際詳細死因分類號碼、「800」以後損傷及中毒之國際詳細死因分類、診斷者代碼、國際基本死因分類號碼、國際簡略死因分類號碼。內政部資料變項與衛生署資料不同之處，在於具有衛生署沒有建立之死亡者教育程度、死亡申請登記日期、村里代號；但沒有建立國際詳細死因分類號碼、「800」以後損傷及中毒之國際詳細死因分類、診斷者代碼。

身分證字號之規則(內政部戶政司公佈的規則)

身分證字號之錯誤是指根據內政部戶政司公佈的身分證檢查碼檢查辦法來檢查不符

投稿日期：91年6月20日

接受日期：91年12月25日

合者。身分證字號之檢查號碼檢查方法是依據中華民國五十九年十二月十八日臺內戶字第三五五六四號令所公佈；其檢查方法如下：首先照附表所列數字將英文字母代號換為數字，再按照十位數字，第一位乘一，第二位乘九，第三位乘八，第四位乘七，，最後一位乘一，將各位相對數字所乘之積相加，然後前項所得之和除以十，求得其餘數，最後以十減去前項所得之餘數即為檢查號碼。但由於此數碼仍為0到9中之任一數，故在隨機之假設下，此數錯誤機率為0.1。身分證字號其正確範圍第1碼26個英文字母A Z；第2碼為性別碼男女以1 2為代表號；第3 9碼為阿拉伯數字0 9共10個代表號；第10碼為身分證字號檢查碼共10個代表號0 9。

戶政資料之範圍及準確性

戶政資料是指出生年月日、死亡年月日、縣市代碼及鄉鎮代碼之綜合資訊；其中偶而有一項以上缺漏無填寫者，視為戶政資料之缺漏。戶政資料錯誤是指出生年月日、死亡年月日、縣市代碼及鄉鎮代碼任何一個變項不符合其正確範圍者。其正確範圍如下，出生年：民國1年至86年(公元1912至1997年)、出生月：1至12月、出生日：1至31日；死亡年：民國69年至86年(公元1980至1997年)、死亡月：1至12月、死亡日：1至31日；縣市代碼：A至Z、鄉鎮代碼：01 33。假定錯誤完全是隨機發生時，任何一個變項有錯誤之機率，各為其正確範圍之倒數；例如身分證字號第1碼26個為英文字母；第2碼為性別2個代表號；第3 9碼為阿拉伯數字10個代表號；第10碼為身分證字號檢查碼共10個代表號。身分證字號可能排列出之最多組數為 $26 \times 2 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 = 5,200,000,000$ ，因此其隨機錯誤之機率為 $1.92E-10$ ，同理戶政資料錯誤機率為 $5.76E-12$ 。但是登錄與輸入牽涉到人為疏失，因此這些計算數字為此資料錯誤機率之下限。

補正方法

本研究擬以衛生署死因資料庫為主體，

而以內政部死因資料庫之身分證字號及戶政資料來補正。連結比對之方法綜述如下：首先篩檢在個別資料庫內身分證字號及戶政資料皆相同的個案；以便找出同一個案重複輸入之機率，其次比對身分證字號與戶政資料之有無缺漏、錯誤及重複。之後，倘兩資料庫間以身分證字號及戶政資料皆為連結碼互相連結，完全相同之個案則表示兩邊為同一個案，這些個案可利用內政部資料來補正衛生署之村里及教育變項。未連上之部分，先以戶政資料為連結碼，在兩資料庫間互相連結，可連結上之部分補正衛生署之村里及教育變項及身分證字號錯誤部分。其次，再以身分證字號為連結碼，在兩資料庫間互相連結，可連結上之部分補正衛生署戶政資料錯誤部分；而有關村里及教育變項補正部分，除了戶政資料可補正外，另外還有戶政資料個案正確時，則比對內政部出生及死亡年月日如有4個以上變項相同則為可補正之個案。內政部用身分證字號或戶政資料與衛生署無法連結上之部分，可作為補充衛生署死因資料庫內因延遲申報或不明原因未建檔之個案，另外開列加入其中，以備有存活追蹤研究之參考。

補正原則

本研究補正有缺漏，或數字超出正常範圍，或檢查碼有錯誤者，其原則如下：1.衛生署死因資料庫之被補正變項(身分證字號、戶政資料)是缺漏或錯誤，意即戶政資料有任何一個變項缺漏或不在正確範圍，且身分證字號缺漏或檢查碼錯誤才納入可補正之個案。2.內政部死因資料庫之補正變項(身分證字號，戶政資料)必須是完全正確，才納入可補正之個案。

連結原則

當連結碼(身分證字號、戶政資料)連結成功，但如有重複連結或其他因素無法判定是否連結成功，則以其他資訊(戶政資料、身分證字號)來做比對。其戶政資料比對優先順序如下：死亡年月日、出生年月日、縣市代碼及鄉鎮代碼。此時，另一戶政資料不同之

個案，即被判定為連結失敗。當使用戶政資料為連結碼，比較衛生署與內政部間之兩個身分證字號，必須確認兩者間之身分證字號內十碼必須超過8個相同，才可被判定為連結成功之個案。補正個案與被補正個案兩者間之身分證字號比較優先順序依第1碼、第2碼、第10碼、第3-9碼。

連結碼選取原則

當以戶政資料為連結碼時，必須先檢視及比對哪些戶政資料特异性最高。以衛生署及內政部公元1985至1997年死因資料檔為基準，建立戶政資料(出生年月日、出生年月、死亡年月日、死亡年月、縣市代碼、鄉鎮代碼)各種組合為連結碼，比對出其重複連結比率，依此原則來決定各種組合為連結碼之優先順序，本研究以重複連結個案比率最少者為連結碼之標準，作填補之依據。

結 果

公元1980至1997年兩資料庫基本特徵比較

首先針對公元1985至1997年間兩資料庫均有身分證字號及戶政資料者相比，當身分證字號且戶政資料一樣視為相同個案時，衛生署之死因資料庫重複輸入率較內政部高出約七倍(見表一)；在身分證字號缺漏及身分證字號檢查碼錯誤方面衛生署都有較高的百分率，可以大致推論內政部死因資料在個案數目及身分證字號較衛生署為準。因此本研究使用內政部死因資料中之身分證字號及戶政資料，來補正衛生署的缺漏與錯誤可能是較合理且減少錯誤之作法。

補正公元1980至1997年衛生署死因資料身分證字號、戶政資料、教育及村里變項

兩資料庫因為身分證字號且戶政資料完全相同被剔除之個案數，衛生署有2,410個及內政部361個(見圖一)。然後，以身分證字號且戶政資料為連結碼兩資料庫互連可連結出1,259,600個個案，而連結不上之個案衛生署計有118,437個個案及內政部計有132,134個個

案。再單獨用身分證字號或戶政資料分別比對兩資料庫，身分證字號相同者有65,810個個案，其中有296個個案身分證檢查碼不正確，補正衛生署戶政資料有3,168個個案，即可補正教育及村里變項，而戶政資料缺漏且錯誤有1個個案無法補正，其他有62,641個個案的戶政資料是在正確範圍內，比對內政部出生及死亡年月日4個以上變項相同者達60,260個個案，可補正教育及村里變項。另一方面戶政資料相同者有43,124個個案，補正衛生署身分證字號有33,036個個案，但有4個個案身分證字號缺漏與352個個案身分證字號檢查碼錯誤無法補正。另外，教育變項補正1,368,313個個案及村里補正1,343,999個個案；兩資料庫可連結之部分，兩個變項補正百分比都達8%以上(見表二)。兩資料庫以身分證字號或戶政資料為連結碼未連結上之個案來補正，共有衛生署9,503個個案及內政部23,213個個案，可能是失去追蹤或延遲申報個案。

戶政資料特异性

公元1980至1984年衛生署資料庫中並無建立身分證字號，因此為了使兩資料庫連結有規則遵循，其建立規則方法如下：從公元1985至1997年，共取得衛生署1,380,447個個案及內政部1,392,095個個案為比對其重複比率之對象。兩資料庫內先篩選出身分證字號正確且唯一的個案，再以身分證字號且死亡年份為連結碼使兩資料庫互相連結，可以連結上有1,320,787個個案；以這些個案為基準。建立戶政資料(出生年月日、出生年月、死亡年月日、死亡年月、縣市代碼、鄉鎮代碼)各種組合為連結碼使兩資料庫互相連結後之重複比率；結果發現以戶政資料全部變項為連結碼時，其重複個案連結率為0.0512%，是各種組合中為重複連結比率最少者(見表三)，定其為連結碼之標準。

補正公元1980至1984年衛生署死因資料身分證字號、教育及村里變項

接著進行衛生署身分證字號補正，從公元1980至1984年，主要填補衛生署身分證字

表一 公元1985年至1997年衛生署與內政部死因資料庫身分證字號及戶政資料之重複或缺漏率

特徵	內政部 N (%)	衛生署 N (%)
身分證字號且戶政資料*重複 ¹	720 (0.052)	4,812 (0.35)
身分證字號缺漏一碼以上 ²	0 (0)	350 (0.03)
身分證字號重複(戶政資料不同人) ²	12 (0.001)	537 (0.04)
身分證字號檢查碼錯誤 ²	3,769 (0.27)	36,461 (2.64)
戶政資料*有一項以上缺漏 ²	376 (0.03)	3,226 (0.23)
戶政資料*重複(身分證字號不同人) ²	26 (0.002)	68 (0.05)
總計	1,392,095	1,380,447

* 戶政資料是指是指出生年月日、死亡年月日、縣市代碼及鄉鎮代碼之綜合資訊。

¹ 衛生署死因資料檔案計有1,380,447個個案，及內政部死因資料檔案計有1,392,095個個案。

² 身分證字號且戶政資料為連結碼兩資料庫互連，而連結不上之個案衛生署計有18,437個個案及內政部計有132,134個個案。

號。以戶政資料全部變項為連結碼可連結上內政部91.53%，只挑內政部資料中身分證字號無誤者，則可以填補身分證字號達88.83% (見表四)。且檢查其縣市、鄉鎮代碼都合平邏輯，進行村里及教育補正，結果教育變項補正388,141個個案及村里補正364,498個個案；兩個變項補正百分比都達83%以上。兩資料庫未連結上之個案，共有衛生署36,802個個案及內政部39,944個個案，可能是失去追蹤及延遲申報個案。至於連結碼放寬標準之得失，在本文之討論部分多加了一個議題『不同戶政資料組合之特異性』中有詳加討論。

討 論

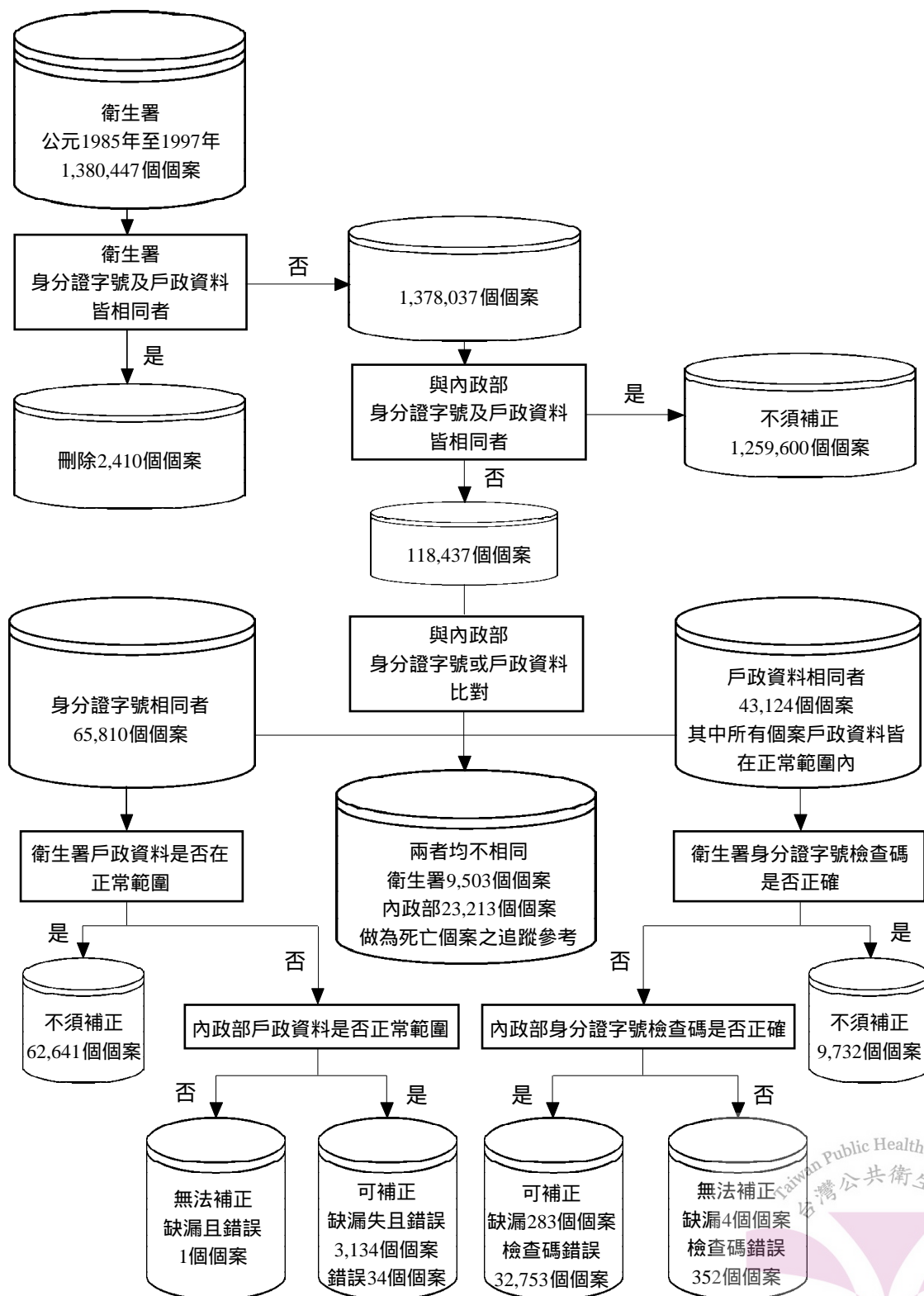
本研究以衛生署死因資料檔為主體，使用內政部死因資料檔正確資訊來補正，重新整理出一個死因資料檔案，可提供未來研究之參考。以下針對戶政資料及身分證字號之準確性、連結策略之檢討、不同戶政資料組合之特異性、補正後資料庫之運用與限制等進行討論，並對未來應用死因資料建檔與校正提出建議。

戶政資料及身分證字號之準確性

各行政機關建立檔案，是以其本身之目

的及用途為導向，因此其所關注的焦點也不同。以死因資料檔而言，衛生署之目的是提供本國生命統計資料，因此對死亡相關資料如死亡種類代碼等可能較為重視，但個案之基本資料與少數失蹤的個案可能就未建檔。然而，內政部與其立場可能較不相同，其建檔的目的是為了追蹤報稅與各種行政用途，剛好可能與衛生署有互補性。本研究即利用此特性，來進行補正之工作。兩資料庫共有變項部分之完整性，綜合在表一上。就個案數而言，衛生署死因資料建檔是從公元1971年開始，但其延遲申報案檔只從公元1996年才建立；而內政部死因資料檔和延遲申報案檔都是從公元1978年建立，但公元1979年當年資料卻已經毀損。在個人資料方面，身分證字號衛生署在公元1984年之前是未建檔的，而內政部從建檔開始即有建立；比較公元1985年至1997年兩資料庫之身分證字號，發現衛生署有缺漏而內政部均完整。衛生署死因資料之身分證字號重複率及檢查碼錯誤較內政部為高；另外戶政資料方面，衛生署有一項缺漏者個案及重複率都較內政部偏高。但是出生日期在民國前後之變項分析結果，因內政部有3%缺漏個案而衛生署無缺漏個案，為兩資料庫比對方便起見，此項不列入戶政資料庫之比對項目。在性別碼方面只有衛生署才有建立，且比較其性別碼與身分

圖一 公元1985年至1997年衛生署死因資料檔案補正身分證字號及戶政資料之流程



表二 公元1985年至1997年衛生署死因資料檔案補充教育及村里資料代碼

		與內政部分身分證字號且 戶政資料相同者	與內政部 戶政資料相同者	與內政部 身分證字號相同者
校對結果				
可連結個案		1,259,600	43,124	65,810
教育資料*	正確	1,259,406	43,115	65,792
	錯誤	141	4	14
	缺漏	53	5	4
村里資料**	正確	1,238,267	42,554	63,178
	錯誤	15,146	391	711
	缺漏	6,187	179	1,921

* 教育資料：依據教育程度查記作業要點附件一(中華民國86年11月10日內政部台(86)戶字號及教育部(86)統86127441號會銜函訂定)，如符合其教育程度代碼為正確；不符合者為錯誤；未填寫者為缺漏。

** 村里資料：依據中華民國台灣地區各市鎮鄉村里代碼行政院主計處公告，如符合其村里代碼為正確；不符合者為錯誤；未填寫者為缺漏。

表三 公元1985年至1997年內政部與衛生署死因資料庫，以身分證字號且死亡年份為連結碼對連，內政部與衛生署死因資料庫所連結上之個案有1,320,787人，在以出生年月日、死亡年月日、縣市代碼、鄉鎮代碼作各種組合之連結碼所產生之連結率。

連結碼	連結產生重複個案數 N(%)	假定完全隨機 情況重複機率*
出生年月日 + 死亡年月日 + 縣市代碼 + 鄉鎮代碼	676 (0.0512)	7.53E-12
出生年月日 + 死亡年月日 + 縣市代碼	3,290 (0.2491)	2.49E-10
出生年月 + 死亡年月日 + 縣市代碼 + 鄉鎮代碼	6,939 (0.5254)	2.34E-10
出生年月日 + 死亡年月 + 縣市代碼 + 鄉鎮代碼	8,133 (0.6158)	2.34E-10
出生年月日 + 死亡年月日	47,739 (3.6144)	6.46E-09
出生年月 + 死亡年月日 + 縣市代碼	76,205 (5.7697)	7.71E-09
出生年月日 + 死亡年月 + 縣市代碼	83,819 (6.3461)	7.71E-09
出生年月 + 死亡年月 + 縣市代碼 + 鄉鎮代碼	191,318 (14.4852)	7.24E-09
出生年月 + 死亡年月日	1,183,146 (89.5789)	2.00E-07
出生年月日 + 死亡年月	1,270,964 (96.2278)	2.00E-07

* 假設各變項均為獨立，理論上可能產生重複之機率。

表四 公元1980年至1984年衛生署死因資料庫以出生年月日、死亡年月日、縣市代碼及鄉鎮代碼做各種組合之連結碼，連結內政部死因資料庫可補正身分證字號、教育及村里百分率。

變項名稱	連結上之 個案數	連結上之 百分比	可補正 身分證字號 之個案數	可補正 身分證字號 之百分比	可補正 教育之 個案數	可補正 教育之 百分比	可補正 村里之 個案數	可補正 村里之 百分比
出生年月日 + 死亡年月日 + 縣市代碼 + 鄉鎮代碼	397,885	91.53	386,134	88.83	388,141	89.29	364,498	83.85
出生年月日 + 死亡年月日 + 縣市代碼(再增連上之 個案)	2,848	0.66	2,779	0.64	2,785	0.64	2,429	0.56
出生年月 + 死亡年月日 + 縣市代碼 + 鄉鎮代碼(再 增連上之個案)	8,683	2.00	8,417	1.94	8,440	1.94	8,022	1.85
出生年月日 + 死亡年月 + 縣市代碼 + 鄉鎮代碼(再 增連上之個案)	7,155	1.65	6,947	1.60	6,943	1.60	6,650	1.53
連結上之個案	416,571	95.83	404,277	93.00	406,309	93.47	381,599	87.79
未連結上之個案	衛生署 內政部	18,116 21,258	4.17 4.86					
未補正之個案			30,410	7.00	28,378	6.53	53,088	12.21
總和	衛生署 內政部	434,687 437,829	100.00 100.00	434,687	100.00	434,687	100.00	100.00

證字號第二碼亦有差異性0.6%存在，因此本研究不把出生日期民國前後及性別碼納入連結碼。內政部有建立個案之村里變項代碼是補正的一個重要資訊，對未來研究環境流行病學可能有相當大的幫助，故在本研究中特別把它填補到衛生署資料中。基於上述內政部資料除了死因資料未能鍵入之外，大致上準確性比衛生署好，因此本研究用內政部死因資料檔來補正衛生署之死因資料，以便未來在世代研究及存活分析使用。

連結策略之檢討：未來使用身分證字號或戶政資料比對流行病學研究個案是否死亡之準確性推估

本研究發現公元1985年至1997年身分證字號相同之個案戶政資料仍連結不上者超過6

萬筆以上；而戶政資料相同之個案身分證字號仍連結不上者超過4萬筆以上(見圖一)。是否連結策略上有誤造成此種現象？仔細探討這些個案，發現衛生署62,641個個案中，比較兩資料庫出生及死亡年月日六個變項中，五個以上變項相同者共有56,018個個案。同樣的方法在兩資料庫戶政資料均相同的3,124個個案中，有9,732個個案身分證字號檢查碼錯誤，比較兩資料庫身分證字號十個號碼中，八個以上號碼相同者共有8,288個個案。根據表3之各碼相同重複機率很低上來想，檢測是否兩資料庫無法相連結之這些個案仍是同一人，可能在人為輸入時誤差造成無法相連結。可見我們以上連結策略並未造成誤差使未來連上之個案假性增加。同時可以大致推論，將來使用此校正後資料庫，以身分證字

號來比對作流行病學追蹤是否死亡準確性可達98.58% ($= (1,259,600 + 65,810 + 32,753 + 283) / 1,378,037$)；反之，如果以戶政資料為連結碼作追蹤是否死亡，約可達4.76% ($= (1,259,600 + 43,124 + 3,134 + 34) / 1,378,037$)。

不同戶政資料組合之特異性

公元1980至1984年間兩資料庫間檢視及比對哪些戶政資料特異性最高，在本研究中為了補正嚴謹度之考量，因此以重複連結個案比率最少者為連結碼之標準，作填補身分證字號之依據。我們發現當出生及死亡年月日、縣市與鄉鎮代碼均相同時，重複比率為0.0512%以下。如果欲使用本資料庫者，有其他研究目標欲放寬連結碼標準，定重複連結比率到小於1%之組合時(見表三)，可放寬至缺鄉鎮代碼、或缺出生日、或缺死亡日3種組合為連結碼，可以再增加填補身分證字號4.18%，教育變項4.18%案及村里3.94% (見表四)。由上可知，把連結碼條件放寬雖然可以增加填補百分率，但是其重複連結比率相對的增加，如何取得最佳條件可依照研究本身之特性來決定。不過本研究已先將所有戶政資料均拿來比對且相同者，補上了身分證字號供大家方便使用。(目前已做好光碟，有意者請向衛生署統計室申請死因資料庫通過，轉給作者且附上工本費即可提供。)

補正後資料庫之運用與限制

此補正過之檔案，在年代方面使用身分證為連結碼時，以往研究範圍只能從公元1985年之後的，本研究結果可使研究範圍往前推至公元1980年；先前研究族群之區域分層只能切割到最小單位鄉鎮，但補正後可使之細分到村里。另外連結後內政部未連結上衛生署之個案，可建立衛生署公元1986年之前未建檔之延遲申報個案及失蹤建檔個案，作為將來研究追蹤死亡個案之依據。

補正後重新整理之死因資料檔案，仍然有部限制存在。在衛生署死因資料檔中應補正而不能補正之個案有兩類，1.未連結上之個案，2.連結上，但內政部無正確資訊來補

正之個案。經過補正後的死因資料庫，不管是以身分證字號或戶政資料為連結碼都仍然還有機會產生重複連結之個案。此時必須再佐以其他相關資訊加以比對來判定正確連結之個案。對於本研究無法補正個案，在未來應該更進一步找出原始書面資料加以校對訂正，以力求資料之完整及正確。

對未來死因資料建檔與校正提出建議

比對兩資料庫公元1985年至1997年資料補正之結果，衛生署身分證字號每年補正率約有2-3%；而戶政資料約在0.01%以下，其補正率未因年代改變而有所減少，因此建議衛生署對死因資料在未來應定期進行補正工作。除了建檔時考慮與內政部連線將村里、教育變項納入之外，對延遲申報個案須補建檔，且對失去追蹤個案可每年定期(例如1-2次)配合戶口普查或校正、失蹤人口、刑事案件、出入境資料來加以比對更新。另外如果可以將現有未建立電子檔案之死亡證明書書面資料建為電子檔案，將可能更方便查詢與校正以增進其準確性。

誌謝

本研究係由國科會計畫(NSC 90-2320-B-002-127-M56)補助部份研究經費。感謝衛生署及內政部提供死因資料。

參考文獻

1. Chen CC, Kuo CJ, Tsai SY. Causes of death of patients with substance dependence: a record-linkage study in a psychiatric hospital in Taiwan. *Addiction* 2001; **96**:729-36.
2. Manjer J, Andersson I, Berglund G, Bondesson L, Garne JP, Janzon L et al. Survival of women with breast cancer in relation to smoking. *Eur J Surg* 2000; **166**:852-8.
3. Martikainen P, Valkonen T, Martelin T. Change in male and female life expectancy by social class: decomposition by age and

- cause of death in Finland 1971-95. *J Epidemiol Community Health* 2001;**55**:494-9.
- 4.Svartbo B, Olov BL, Gunnarsson T, Steen L, Ribe M. Survival during and after hospitalization: a medical record linkage. *Int J Health Care Qual Assur Inc Leadersh Health Serv* 1999;**12**:13-7.
5. 台灣省政府衛生處：國際疾病、傷害及死因分類表。台灣省政府衛生處，1995。
6. 行政院衛生署：國際疾病傷害及死因分類表(1975 Revision)。台北：行政院衛生署，1981。
- 7.Lee WC. Characterising life-span variability in a population using the life-table-based Lorenz-curve analysis. *J Epidemiol Biostat* 2000;**5**:315-20.
- 8.Lee WC. The meaning and use of the cumulative rate of potential life lost. *Int J Epidemiol* 1998;**27**:1053-6.
- 9.Lu TH, Lee MC, Chou MC. Accuracy of cause-of-death coding in Taiwan: types of miscoding and effects on mortality statistics. *Int J Epidemiol* 2000;**29**:336-43.
- 10.Liang J, Bennett JM, Krause NM, Chang MC, Lin HS, Chuang YL et al. Stress, social relations, and old age mortality in Taiwan. *J Clin Epidemiol* 1999;**52**:983-95.
- 11.Liu X, Hermalin AI, Chuang YL. The effect of education on mortality among older Taiwanese and its pathways. *J Gerontol B Psychol Soc Sci* 1998;**53**:S71-82.
- 12.Chiang T. Economic transition and changing relation between income inequality and mortality in Taiwan: regression analysis. *Br Med J* 1999;**319**:1162-5.
- 13.Chen CJ, You SL. Epidemiology of cervical cancer in Taiwan. *Gynecol Oncol* 1997;**67**:115-6.

附表

縣市別	代號	數字換算	縣市別	代號	數字換算	縣市別	代號	數字換算
台北市	A	10	苗栗縣	K	19	花蓮縣	U	28
台中市	B	11	台中縣	L	20	台東縣	V	29
基隆市	C	12	南投縣	M	21	澎湖縣	X	30
台南市	D	13	彰化縣	N	22	陽明山	Y	31
高雄市	E	14	雲林縣	P	23	金門縣	W	32
台北縣	F	15	嘉義縣	Q	24	連江縣	Z	33
宜蘭縣	G	16	台南縣	R	25	嘉義市	I	34
桃園縣	H	17	高雄縣	S	26	新竹市	O	35
新竹縣	J	18	屏東縣	T	27			

