

抽樣調查中戶籍資料適用性之探討

洪永泰*

《本文摘要》

本文以1990年國民健康調查和1994年國民醫療保健調查的資料，參考學者專家的研究成果，檢驗了一部分戶籍登記項目的正確性，證實婚姻狀況有6.9%紀錄與受訪者自述不符，教育程度則有25.6%不正確。在戶籍登記人口與現住人口的差異方面，若以「人」為計算單位時有不到九成的人「籍在人在」，超過一成「籍在人不在」，超過一成「人在籍不在」。若以「戶」為計算單位，則只有大約五成五的戶現住人口和登記資料完全相符，其餘「籍多於人」和「人多於籍」者各約兩成，這些現象也因地區而異，因戶內人數的多寡而異。此外本文並以1990年內政部戶口普查的百分之五抽樣資料建立臺灣地區戶內二十歲以上人口數的分布數據，證實如果等機率抽戶，再自戶中選擇一位受訪者，則理論上無法抽出一套有人口結構代表性的樣本。

關鍵詞：戶籍資料，現住人口，抽樣調查，戶中選樣

壹、前言

台灣在世界各國政府統計領域裡最引以為傲的就是擁有一套歷史悠久、資料完整的戶籍檔案。從民國三十五年開始每一個出生的人連同其家人的姓名、地址、性別、出生序、生日、本籍、教育程度、職業、婚姻狀況、配偶姓名、有一段時期甚至包括血型，

* 作者現為國立政治大學選舉研究中心副研究員。

都進入了政府的掌握。根據這個檔案而來的許多時間序列人口學數據不但在過去數十年間成爲政府各項施政建設規劃的重要參考指標，而且也提供了學術界許多領域諸如經濟、社會、政治、醫療衛生等極佳的研究題材，人口政策和家庭計畫研究就是一個很典型的例子。此外，在社會變遷的研究方面，它更是主要的數字依據。

事實上，不論是政府、民間、或是學術機構在引用這些資料時都會視需要而決定採用的項目，其中最大考慮因素是：（一）資料的正確性，（二）登記人口和現住人口的落差。前者最爲一般人以爲共識的是：登記人口的性別、年齡、和婚姻狀況大概不會錯，籍貫也許還好，但是教育程度和職業、行業的正確性則存疑。後者則要考慮用途，如果使用目的和選舉或公、私權益有關，則可以直接引用登記人口的數據，但是如果使用目的涉及經濟活動、交通建設、醫療衛生、工商發展等方面則登記人口的數據是不可冒然採用的。

關於戶籍登記資料的正確性，早期中研院經濟所曾經認真地求證過（Liu, 1967和Speare, et al., 1975），只是前者調查地區偏向局部性，代表性方面稍嫌不足，後者則年代久遠（1973年調查），似有重新驗證的必要。

其次是關於戶籍登記人口和實際上現住人口的差異問題，大家都知道一個地區的現住人口至少包括了戶籍登記在此且又常住於此、戶籍登記不在此但常住於此、以及因工作求學或各式各樣的原因而有部分時間停留在此的人，根據戶籍登記資料所得到的數據無疑的是和現實世界有極大差距的。熊瑞梅（1989）有系統地整理了有關戶籍登記之中「籍在人不在」和「人在籍不在」的調查估計並分析其相關因素。趙弘靜（1994）以官方調查爲依據分析現住人口與戶籍登記人口的差異。李隆安（1994）則以實際居住地爲抽樣基礎提供了較新的估計方法和數據。許博雄與洪永泰（1995）以追蹤「籍在人不在」人口的結果比較遷徙人口和「籍在人在」者的差異。姑且先不論上述各項研究的數據是否一致，我們可以確定的一件事是：如果我們以戶籍資料爲母體清冊進行抽樣調查，而且每戶只訪問一位合格的受訪者的話，則在最好的情況下也只能找到戶籍登記在此且又常住於此的人，他們很可能只佔實際上居住於該地區人口的一部分而已。從這個角度來看政府主計部門的許多重要調查，例如每月一次的人力資源調查，每年一次的家庭收支調查，每兩年一次的青少年狀況調查，老人狀況調查等，由於這些調查都以戶籍資料爲抽樣清冊（行政院主計處，1995）且其調查結果的推論對政策規劃和國計民生影響重大，參與其事者對於這方面的缺失都有相當的體會而亟思改進之道（賴美齡，1986；詹德松，1987；賴守仁，1987，1989）。

爲求更進一步瞭解上述諸多有關戶籍資料的運用缺失和困難，本研究有下列兩項目的：

- (一)驗證戶籍登記中各項資料的正確性，
- (二)探討以戶籍登記資料為抽樣清冊的適用性。

貳、戶籍資料的正確性

早期戶籍資料登記的項目很多，自1992年之後有較大的減化。根據民國八十一年六月二十九日公布施行的戶籍法第五條，戶籍登記的項目包括身分（出生、死亡、死亡宣告、認領、收養、終止收養、結婚、離婚、監護、指定繼承）、遷徙（遷入、遷出、住址變更、流動人口）、行業及職業（十五歲以上就業者）、教育程度（六歲以上者）。Speare等人（1975）在1973年的調查顯示戶籍登記的性別、年齡、婚姻狀況、和與戶長關係這幾個項目和受訪者所回答的資料誤差較小，教育和行、職業的差距就很大。胡幼慧（1989）於1987年亦曾針對臺北地區750位三十五歲以上民眾驗證其登記婚姻狀況與自述狀況的差異，結果發現有11.6%的人不相符。

表1和表2是利用江東亮、張明正、洪永泰（1995）的1994年國民醫療保健調查資料，分別從抽樣檔的戶籍資料和訪問結果檔之中驗證完訪樣本的婚姻狀況和教育程度的正確性，結果顯示婚姻狀況有6.9%不正確（表1斜角部分是正確資料，總共有 $20.6\% + 66.3\% + 1.7\% + 4.5\% = 93.1\%$ ），教育程度有25.6%不正確（表2斜角部分正確資料有 $37.1\% + 13.1\% + 16.2\% + 4.7\% + 3.1\% + 0.2\% = 74.4\%$ ）。性別和年齡幾乎找不到錯誤，而行業和職業方面由於譯碼的困難至今仍無法解決，只好從缺。

造成戶籍資料和訪問結果不一致有許多可能的原因，第一個可能是在調查過程中發生錯誤，包括對某一項目的某些等級定義不清楚、或無法明確歸類、或認知錯誤等，例如不同時代、不同年齡層的人對教育程度的認知也可能不同，常見的有師範、補校、士校、士官班、專修班、通訊電子學校、外語學校、公學校（日本制）、高等學校（日本制）等都很容易發生歸類的錯誤，其他如畢業、肄業、結業、或學分班等差異都可能產生來自受訪者或訪員的歸類錯誤；又如在婚姻狀況方面喪偶者在沒有被特別提示選項的情況下也有可能被受訪者或訪員歸類為已婚或未婚。第二個可能是在資料處理過程中發生錯誤，如譯碼、鍵入、或資料轉換等步驟。第三個可能的原因是受訪者認為問題敏感、或感受到社會壓力、或純粹是訪問情境、受訪者與訪員間的互動效果等，造成了教育程度高報、婚姻狀況不符等現象。有關這方面和相關因素的討論請見胡幼慧（1989）。

婚姻狀況的錯誤主要來自原登記為未婚，訪問時卻回答為已婚（佔未婚者的8.6%，全體的2.0%），另外是原登記為已婚，訪問時卻回答為喪偶（佔已婚者的2.4%，全體的1.7%），這兩種情形都可能屬於真正的身分變更，不算錯誤。其次是原登記為已婚，訪

問時卻回答為未婚（佔已婚者的1.9%，全體的1.3%），依常識判斷較有可能是上述三種原因之一（尤其是第三個原因），戶籍登記錯誤的機會較小。值得注意的是原登記為離婚者有13.7%回答「已婚」、7.4%回答「喪偶」，雖然這些人只佔總數的一小部分，但卻是學者研究的議題（胡幼慧，1989）。

戶籍登記教育程度的偏誤是很容易就可以預期的，表2的斜角右邊項目顯示面訪結果比戶籍登記的程度高：戶籍登記為小學及以下者面訪結果只有86.6%相符，13.4%實際教育程度更高；戶籍登記為國、初中者差異最大，面訪結果只有49.8%相符，有47.6%的人說他們的教育程度更高；戶籍登記為高中、職者有17.4%實際教育程度更高；專科的有8.8%、大學的有4.8%也是如此；總計所有項目面訪結果比戶籍登記的程度高的一共佔總數的22.6%。從另一邊看，表2的斜角左邊項目則顯示面訪結果比戶籍登記的程度低，其原因可能是來自受訪者和戶政事務人員認知上的差異，也可能是訪員的歸類失誤，不過這些人數比率不高，所有項目總加起來約佔總數的3%，影響不大。整個說來，假設受訪者回答的教育程度是對的，則官方數字（戶籍登記）小學程度或以下者多了4.7個百分點，國初中程度者多了約10個百分點，高中職程度者少了約8個百分點，專科程度少了4.8個百分點，大學及以上低估了1.7個百分點。

表1 1994年國民醫療保健調查戶籍登記婚姻狀況和面訪結果之差異：
臺灣省十五歲以上樣本

	面訪結果					合計
	未婚	已婚	離婚	分居	喪偶	
戶籍登記						
未婚	89.0 20.6	8.6 2.0	.8 .2		1.6 .4	100.0 23.2
已婚	1.9 1.3	94.9 66.3	.2 .2	.5 .3	2.4 1.7	100.0 69.8
離婚	2.1 .0	13.7 .3	75.8 1.7	1.1 .0	7.4 .2	100.0 2.2
喪偶	1.0 .0	3.4 .2	1.9 .1		93.7 4.5	100.0 4.8
合計	22.0	68.7	2.1	.4	6.7	100.0

註：第一排數字是依照戶籍登記每一項婚姻狀況計算的百分比，
第二排數字是佔總數的百分比，樣本數是4295。

表2 1994年國民醫療保健調查戶籍登記教育程度和面訪結果之差異：
臺灣省十五歲以上樣本

	面訪結果						合計
	小學以下	國初中	高中職	專科	大學	研究所	
戶籍登記							
小學以下	86.6 37.1	6.0 2.6	5.9 2.5	1.2 .5	.3 .1	.0 .0	100.0 42.9
國初中	2.6 .7	49.8 13.1	34.5 9.1	9.7 2.5	3.3 .9	.2 .0	100.0 26.3
高中職	1.1 .2	2.7 .6	78.8 16.2	13.2 2.7	3.4 .7	.8 .2	100.0 20.6
專科	2.1 .1	.7 .0	10.2 .6	78.2 4.7	8.4 .5	.4 .0	100.0 6.0
大學		2.1 .1	5.8 .2	9.0 .4	78.3 3.1	4.8 .2	100.0 4.0
研究所		7.7 .0			7.7 .0	84.6 .2	100.0 .3
合計	38.2	16.4	28.7	10.8	5.3	.7	100.0

註：第一排數字是依照戶籍登記每一項教育程度計算的百分比，
第二排數字是佔總數的百分比，樣本數是4770。

參、戶籍登記人口與現住人口的差異

戶籍登記人口與現住人口的差異來源可以分為兩部分：「籍在人不在」和「人在籍不在」。熊瑞梅（1989）引述1966年台灣省人口研究中心的估計前者是5%至6%，後者是2.5%，1973年劉克智的估計分別是13.6%和11.7%。行政院主計處（1994）對於「籍在人不在」的估計從1979年以來大都維持在8%至9%，1992年第一次跌至7.35%，而「人在籍不在相同縣市」的人佔了現住人口的7.09%。熊瑞梅（1989）的估計「籍在人不在」是6.8%，「人在籍不在」是10.2%。李隆安（1994）的新調查方法自其訪問成功者之中估計有17.9%「籍在人不在」，7.2%「人在籍不在」。本研究的估計，1990年的國民健康調查有10.25%的「籍在人不在」和13.99%的「人在籍不在」；1994年的國民醫療保健

調查則有13.07%的「籍在人不在」和11.45%的「人在籍不在」。

從以上這些數據看來，可見確切估計之不易。事實上，所有的估計並不能直接比較，各家的估計分母並不一致，可能使用現住人口數，也可能使用登記人口數，也有可能使用籍在人在、人在籍不在、以及籍在人不在的總和；此外，各家的抽樣設計不同，有些必須考慮加權處理的也許疏忽了，使得最後的數據可能會有些偏差。

從另一個角度看，「籍在人不在」和「人在籍不在」的百分比是否應該平衡呢？邏輯上，如果計算範圍相同而且是封閉的，亦即登記人口總數等於現住人口總數，則兩個百分比應該相等。以臺灣地區為定義範圍而言，這個說法是可以成立的，舉行政院主計處（1994）的統計為例，民國八十一年十月臺灣地區戶籍登記總人口數約是20714千人，現住人口數則約為20686千人，佔99.86%，其他的0.14%大多在國外（約24千人）及少數現住於福建省（約3千人），大致上說登記人口數和於現住人口數相差不多，計算上的小差別應可忽略。但是如果以臺灣省和北、高兩市，或以各地區、各縣市分別做為計算範圍，則問題就複雜多了。理論上，若計算範圍不同，則各該範圍內「籍在人不在」和「人在籍不在」的百分比也會因地而異，也因此才有研究分析的價值。但集合所有的分類範圍合併成為臺灣地區時，「籍在人不在」和「人在籍不在」的百分比仍應該大致相等才對。上述各項數據沒有得到這個結果大概是抽樣設計時並不以縣市或地區做為樣本數的分配依據，推論時卻分別以這些縣市或地區歸類，代表性和加權值都可能出差錯的緣故。另一個原因則是抽樣誤差，畢竟這些數據也只不過是一次抽樣調查的點估計而已。

表3和表4勉強將四項研究併排比較，引用這四項研究是考慮到他們都是全國性的抽樣架構以及在研究報告中都有地區的交叉數據可以直接採用，或「加工處理」後使用。各項數據都儘可能地重新編排計算使得比較可以具有意義。我們很容易就可發現主計處的各项數據最「保守」，這是因為調查定義「籍在人在」指的是「戶籍和人都在相同縣市」的緣故；李隆安（1994）的數據較「激烈」，兩次國民健康調查則居中，不過也可以看出從1990到1994「籍在人在」越來越少，「籍在人不在」越來越多，數據有向「激烈」走的趨勢。整體而言，如果只用「籍在人在」的受訪者做為抽樣調查的推論樣本的話，大概只能包羅大約九成的現住居民（臺北市更在這個數據之下），且先不要說還有訪問失敗的問題。

從地區的差異來看，臺北市自成一個等級：「籍在人在」比率最高，「籍在人不在」比率最低，顯示設籍在此的人較少「流失」，另外「人在籍不在」的比率超高，顯示其「吸人」的威力。如果只用「籍在人在」的受訪者做為抽樣調查的推論樣本的話，失真的情形最嚴重。高雄市是第二個等級，各項數據都類似台北市，只是數量稍次而已。高雄市和臺北市的「人在籍不在」比率都比「籍在人不在」高，表現出大都會吸引人口的

表3 各項研究戶籍地與現住地差異之比較（一）

地區	籍在人在/戶籍登記人數				籍在人不在/戶籍登記人數			
	1990 國健	1994 國健	1992# 主計處	1993** 吸菸行爲	1990 國健	1994 國健	1992# 主計處	1993** 吸菸行爲
戶籍地								
臺北市	95.58	91.66	97.00	88.63	4.42	8.34	3.00	11.37
高雄市	96.36	90.16	94.59	86.23	3.64	9.84	5.41	13.77
北部	91.67	92.27	96.10	86.17	8.33	7.73	3.90	13.83
中部	87.97	83.87	90.51	83.61	12.03	16.13	9.49	16.39
南部	84.99	84.46	89.09	75.39	15.01	15.54	10.91	24.61
東部	82.89	64.46	85.25	70.38	17.11	35.54	14.75	29.62
合計	89.75	86.93*	92.79	82.97*	10.25	13.07*	7.21	17.03*

註：北部不含臺北市，南部不含高雄市。

=：資料改編自行政院主計處（1994）台灣地區國內遷徙調查報告。

*：合計時採用加權計算。

**：依照李隆安（1994）調整分母重新計算。

表4 各項研究戶籍地與現住地差異之比較（二）

地區	人在籍不在/現住人數				籍在人在/現住人數			
	1990 國健	1994 國健	1992= 主計處	1993** 吸菸行爲	1990 國健	1994 國健	1992= 主計處	1993** 吸菸行爲
戶籍地								
臺北市	14.69	11.60	13.09	22.60	85.31	88.76	86.91	77.38
高雄市	21.13	9.29	8.66	13.77	78.87	89.67	91.34	86.23
北部	11.80	11.61	5.76	7.28	88.20	88.96	94.24	92.72
中部	11.40	11.17	6.78	7.56	88.60	88.83	93.22	92.44
南部	18.26	11.70	4.74	4.89	81.74	89.25	95.26	95.11
東部	10.00	14.29	6.81	6.15	90.00	87.38	93.19	93.85
合計	13.99	11.45*	7.09	9.26*	86.01	88.97*	92.91	90.72*

註：北部不含臺北市，南部不含高雄市。

=：資料改編自行政院主計處（1994）台灣地區國內遷徙調查報告。

*：由於北、高兩市擴大抽樣，合計時採用加權計算。

**：依照李隆安（1994）調整分母重新計算。

現象。臺灣省的北部和中部是第三個等級，數據和北高兩市有些落差，而且呈現「人口輸出」的特徵，「籍在人在」的比率稍低，「籍在人不在」比率開始攀高，而且都明顯高出「人在籍不在」比率甚多。南部和東部是第四個等級，雖然東部的數據比南部也有一段落差。「籍在人在」的比率最低，人口流失情形也最嚴重。

戶籍登記人口與現住人口在人口學上的基本特徵有無差異？如果有的話，是否會影響到其他認知態度與行為方面的推論？熊瑞梅（1989）以爲一個家庭之中「非核心家庭成員」及「大專教育程度」者最有可能居住地和戶籍地不一致。趙弘靜（1994）以整體性的資料推論「都市化程度、就業機會、和生活環境」是戶籍地和現住地差異的主要特徵。許博雄、洪永泰（1995）也證實了遷徙人口傾向生產年齡（35-44歲），大專教育程度，家庭人口少，受雇於私人機構，以及自營作業者，詳細情形本文不再贅述。

肆、以戶籍登記人口爲抽樣清冊的樣本代表性問題

在臺灣地區抽樣調查的眾多研究之中以戶籍資料爲抽樣清冊的情形可分爲兩大類型：一是以「戶」爲最終抽出單位，中選戶的戶中人口全查，通常是包括設籍者與現住者；另一種是以「人」爲最終抽出單位，方式可能是直接抽人，如根據選舉人名冊抽合格選民就是；也可能是先抽戶，再透過戶中選樣方式選擇一名合格的受訪者。第一種類型醫療衛生領域和政府主計部門較常使用，但資源耗費龐大，一般研究者較難得有機會採用。第二種類型則是常見的社會科學調查研究典型，也是本文要探討的主題之一。

以戶籍登記人口資料爲抽樣清冊在實務上必須要考慮的一個風險是：中選的樣本在家嗎？訪問成功的機會有多少？是不是應該事先膨脹一定倍數的抽出樣本數？這些問題的一個關鍵是：戶中人數，甚至是戶中合格人數，是否正確而適用？表5是1990和1994兩次國民健康調查有關戶中人數分布的驗證結果。兩次調查的數據並不穩定，使得推論基礎不是十分堅強。其中原因有可能是社會變遷的因素，也有可能是追蹤訪問努力的程度不同造成完訪率的不同（1990成功率是 $1031 \div 1203 = 85.7\%$ ，1994成功率是 $3119 \div 3814 = 81.8\%$ ）。無論如何，我們仍然可以說一人戶的量實際上並沒有戶籍資料上說得多，四人戶是臺灣地區戶口結構的最大宗，佔了將近四分之一。表面上看來，戶量的分布登記數據和現住情況雖有差異，但不是太離譜。可是如果我們更深入一點觀察，就可以發現問題其實大很多：表6和表7是每戶核對現住人口和戶籍登記人口的結果，兩次調查都證實戶籍登記人口和現住人口完全相符的只有五成六七左右！這裡要稍微留意的是表6和表7是以戶爲計算單位，而表3和表4是以人爲計算單位。登記人口和現住人口完全相符的情形在地區方面以北部較高，南部較低。在戶內人數方面以一人戶最高，三人戶及

表5 1990、1994國民健康調查戶籍登記人數與現住人數之分布

戶內 人數	臺灣地區 母體 百分比*	1990		1994 [≠]	
		戶籍 登記	實際 居住	戶籍 登記	實際 居住
1	13.4	13.0	10.4	15.1	13.1
2	12.7	15.0	14.8	14.8	16.8
3	14.6	15.6	14.1	15.4	15.4
4	21.5	23.7	24.6	20.8	23.2
5	18.6	17.5	18.0	16.5	16.6
>=6	19.2	15.2	18.2	17.4	14.9
樣本戶數		1031	1028	3119	3114

*：台灣地區母體資料根據內政部民國七十九年台閩地區戶口及住宅普查報告。

≠：由於北、高兩市擴大抽樣，合計時採用加權計算。

表6 1990、1994國民健康調查戶籍登記人口與現住人口相符情形：依地區分

戶籍地	1990					1994				
	完全 相符	入=出	入>出	出>入	樣本 戶數	完全 相符	入=出	入>出	出>入	樣本 戶數
臺北市	62.9	.8	28.2	8.1	124	57.8	3.8	21.9	16.5	552
高雄市	46.8	2.5	44.3	6.3	79	57.8	3.0	17.7	21.5	498
北部	61.6	1.7	20.9	15.8	292	64.4	1.7	19.0	14.9	717
中部	56.2	1.9	19.4	22.5	315	50.7	3.0	17.3	29.0	813
南部	52.4	3.9	23.8	19.9	206	52.7	2.6	16.8	27.9	463
東部	50.0	.0	11.1	38.9	18	17.1	3.9	13.2	65.8	76
全體	57.0	2.1	23.5	17.4	1034	55.6*	2.6*	18.2*	23.5*	3119

註：「入」表示戶內「人在籍不在」的人數，「出」表示戶內「籍在人不在」的人數。

* 由於北、高兩市擴大抽樣，合計時採用加權計算。

表7 1990、1994國民健康調查戶籍登記人口與現住人口相符情形：依登記戶內人數分

	1990					1994				
	完全 相符	入=出	入>出	出>入	樣本 戶數	完全 相符	入=出	入>出	出>入	樣本 戶數
戶內人數										
1	65.7	.0	34.3	.0	134	66.4	1.8	31.4	.4	471
2	58.3	.6	34.0	7.1	156	58.2	3.2	28.5	10.1	463
3	46.6	2.5	38.5	12.4	161	48.8	2.5	26.1	22.5	480
4	62.0	3.3	18.8	15.9	245	62.7	2.8	12.0	22.4	649
5	58.3	2.8	12.2	26.7	180	58.5	2.9	8.5	30.1	514
>=6	49.4	2.5	8.9	39.2	158	38.7	2.5	7.7	51.1	542
全體	57.0	2.1	23.5	17.4	1034	55.6*	2.6*	18.2*	23.5*	3119

註：「入」表示戶內「人在籍不在」的人數，「出」表示戶內「籍在人不在」的人數。

* 由於北、高兩市擴大抽樣，合計時採用加權計算。

六人以上者較低。其次，戶中「人在籍不在」和「籍在人不在」的戶數比率兩次調查有不同的估計，在1990年北、高兩市的「入>出」比率都比「出>入」超出甚多，1994年則有很大的改變，北市差距縮小，高市則流出多於吸入。北部地區則保持入多於出的狀況。中部和東部出得多、入得少。南部地區在1990年入大於出，1994年卻出大於入，且改變甚多。從戶內人數來看，登記戶內人數在三人以下者吸入大於流出，五人以上者則流出遠大於吸入，從1990年到1994年，各種家戶的流出率都顯著攀升，吸入比率則全面下降。綜上所述，如果我們事先以戶籍登記人口資料為抽樣清冊先抽出受訪者再前往訪問，不但高失敗率可以預期，和現住人口的脫節也是必定發生的結果。

先從戶籍資料抽出戶，再經戶中選樣過程抽出一位中選者的另外一個危機是：即使在完美的抽樣設計與執行以及百分之百的訪問成功率之下，我們很可能在理論上仍然不可能找到一群和母體結構相同的受訪者做為有代表性的樣本。這是戶口結構的因素使然。表5雖然列出臺灣地區戶內人數的母體分布（1990戶口普查）及一些抽樣調查估計，但社會科學界常用的戶中成人數（或二十歲以上人數）的分布數據卻不易取得。本研究自內政部1990年戶口普查資料檔中以各縣市為次母體隨機抽取百分之五的樣本，共244,957戶，623,945人（二十歲以上）進行分析。結果發現臺灣地區戶內二十歲以上人口數的分布中一人戶和兩人戶就佔了63.45%（表8），而一人戶之中30-39歲者佔32.34

%，二人戶之中30-39歲者佔45.33%，從或然率的計算，只要先抽戶，戶內選一人，則30-39歲的人一定會過度代表；同理，20-29歲的人大都集中在五人戶或更大的戶內，這些戶佔不到臺灣地區戶口數的12%，難怪從事抽樣調查實務工作的人老是在抱怨找不到年輕人。

表8 台灣地區二十歲以上人口之年齡分布：依戶內二十歲以上人口數分

	20-29歲	30-39歲	40-49歲	50-59歲	60歲及以上	佔總戶數%
戶內人數						
1人	14.53	32.34	17.10	9.51	26.52	19.97
2人	16.51	45.33	20.52	6.48	11.16	43.48
3人	27.62	19.73	18.78	14.86	19.02	15.35
4人	36.47	15.58	13.11	17.36	17.47	10.88
5人	43.57	14.88	9.45	18.61	13.49	5.96
6人	45.03	17.98	6.79	17.44	12.76	2.68
7人	45.15	20.96	5.55	15.40	12.94	1.06
8人	43.83	24.68	5.11	13.38	13.00	.41
9人	43.73	26.68	5.02	11.81	12.77	.12
10人及以上	43.39	25.98	7.07	10.80	12.76	.09
合計	28.13	28.22	15.77	12.57	15.31	100.00
中選機率*	22.84	32.64	17.48	10.70	16.33	100.00

資料來源：內政部1990年台閩地區戶口普查臺灣地區資料檔百分之五隨機抽樣。

* 先等機率抽戶，再每戶抽一人之中選機率。

伍、結論與建議

本文以1990年國民健康調查和1994年國民醫療保健調查的資料，參考學者專家的研究成果，檢驗了一部分戶籍登記項目的正確性，證實婚姻狀況有6.9%紀錄與受訪者自述不符，教育程度有25.6%不正確，如果受訪者回答的教育程度是對的，則官方數字小學程度或以下者多了4.7個百分點，國初中程度者多了約10個百分點，高中職程度者少了約8個百分點，專科程度少了4.8個百分點，大學及以上低估了1.7個百分點。

在戶籍登記人口與現住人口的差異方面，本研究估計以「人」為計算單位時有不到九成的人「籍在人在」，超過一成「籍在人不在」，超過一成「人在籍不在」，若以戶

籍資料進行抽樣調查，則頂多也只能包含九成不到的現住人口，而且因地區而異。若以「戶」為計算單位，則只有大約五成五的戶現住人口和登記資料完全相符，其餘「籍多於人」和「人多於籍」者各約兩成，這些現象也因地區而異，因戶內人數的多寡而異。

本文並以1990年內政部戶口普查的百分之五抽樣資料建立臺灣地區戶內二十歲以上人口數的分布數據，發現依照此一分布的年齡結構，如果等機率抽戶，再自戶中選擇一位受訪者，則理論上無法抽出一套有人口結構代表性的樣本。

其次，綜合以上的討論，本文亦針對不同的研究目的和研究對象提出以下的建議：

一、一般性建議：

(一)以臺灣地區為範圍的民眾婚姻狀況可依據戶籍登記數據，不過要考慮引用數據的時間落差，通常兩年之內是會有少數人有婚姻狀況的變更，總體的差距（亦即每一類婚姻狀況佔總數的百分比）應不會超過兩個百分點。

(二)根據戶籍登記資料而統計的教育程度數據最好不要引用，因為嚴重低估，寧可使用較穩定的抽樣調查數據。

二、以現住人口為對象之研究：

(一)採用李隆安（1994）方式抽樣，或使用戶籍資料抽樣亦可，但兩者均應儘可能地以「戶」為最終抽出單位。中選戶之所有現住人不論登記與否全查。如果堅持一戶訪問一人，則應根據科學的戶中選樣方式選出一人（不論登記與否只要合格全是候選人），然後根據該受訪者的抽取率和訪問成功率給予權值，再併入全檔分析。

(二)如果是電話抽樣，則在使用適當方法等機率抽出戶之後，依上述方式進行。

三、以戶籍登記人口為對象之研究（例如選舉）：

(一)以選舉人名冊為抽樣清冊，以「人」為最終抽出單位。

(二)如果無法使用選舉人名冊而只有戶籍資料可用時，也應跳過「戶」而直接以「人」為抽出單位。

(三)如果是電話抽樣，則在使用適當方法等機率抽出戶之後，依前述方式進行選人和加權，不過合格人數以設籍者為限。

(四)這三種做法都只能找到「籍在人」的人，即使透過加權處理仍然無法解決「籍在人不在」的缺失，在小範圍選區這個問題更是嚴重。在大範圍的選舉（例如總統選舉）則可以現住人口為對象，只是在問卷中詢問受訪者的戶籍即可。

四、不論是面訪或是電話訪問，授權給訪員根據中選戶內的合格人數（登記者和未登記者）來執行戶中選樣，並紀錄戶中合格者每人之性別年齡與戶籍資料及訪問結果應是不可避免的事，這也是從事抽樣調查實務工作者必須面臨和克服的一項任務。

參考文獻

詹德松

1987 「對家庭收支調查改進意見之商榷」，**中國統計學報**，25卷6期，頁11-24。

趙弘靜

1994 「臺灣地區現住人口與戶籍登記人口之差異」，**主計月報**，78卷2期，頁41-48。

江東亮、張明正、洪永泰

1995 一九九四年國民醫療保健調查，行政院衛生署委託研究期末報告。

行政院主計處

1994 中華民國八十一年台灣地區國內遷徙調查報告，台北。

1995 中華民國調查統計索引，台北。

熊瑞梅

1989 「影響遷移登記正確性的人口變項」，**人口學刊**，12期，頁29-51。

許博雄、洪永泰

1995 「民國七十九年臺灣地區國民健康調查遷移樣本之特性分析」，**中華公共衛生雜誌**，14卷3期，頁283-296。

胡幼慧

1989 「臺北地區登記與自述婚姻狀況差異之研究」，**人口學刊**，12期，頁53-66。

賴守仁

1987 「家庭收支調查改進之芻議」，**七十六年統計學術研討會實錄**，頁13-87。中國統計學社，七十六年四月十、十一日，台北。

1989 改進家庭收支調查之重要概念及方法，統計專題研究報告，行政院主計處，台北。

賴美齡

1986 「現行家庭收支調查制度之研究與改進」，**中國統計學報**，24卷10期，頁19-48。

李隆安

1994 「抽樣調查新方法的探討」，**調查研究**，試刊號，頁5-35。

林瑞雄、江東亮、洪永泰、張明正

1991 國民健康調查之規劃與試驗，行政院衛生署委託研究期末報告。

選舉研究

Liu, Paul K.C.

1967 The Use of Household Registration Records in Measuring the Fertility Level in Taiwan. The Institute of Economics, Academic Sinica.

Speare, A. Jr., P.K.C. Liu, K.S. Hwang, C.L. Tsay, and M.C. Speare

1975 "A measurement of the accuracy of data in the Taiwan household register," Academic Economic Papers 3(2), The Institute of Economics, Academic Sinica.

The Use of Household Registration Records in Sampling Surveys

Yung-tai Hung

Abstract

The study checks the accuracy of some items of the household registration records in Taiwan using data from the 1990 and the 1994 National Health Surveys and finds that the discrepancy rate between the marital status records and survey results is about 6.9%. There is also 25.6% inaccurate rate for the records of educational level. If counted by person, less than 90% of the people currently live in their registered households, over 10% of the people do not live in their registered households, and over 10% of the people live in places which they do not registered. If counted by household, only about 55% of the records are correct in both household registration and persons live in, other 45% are either greater or less than the registration record in terms of the number of people living in the household, or even the number is correct the persons' names do not match. This type of inaccuracy varies in different areas and also depends on the size of the household. The study also looks into the distribution of the household size using the 1990 census data and finds that it is not feasible to have a representative sample if households are selected first with equal probability and one adult respondent is then selected within the household due to the fact that a large portion of the younger people are clustered in bigger size households which account for only a much smaller portion of the total number of the households.

Keywords: household registration; population de facto; population de jure; sampling survey; selection within household.