

研究紀要

比較入選機率分組與其他加權方法 對電話調查樣本的調整：上網率的推估

杜素豪

杜素豪 中央研究院人文社會科學研究中心副研究員。通訊地址：台北市南港區研究院路二段128號中央研究院人文社會科學研究中心，E-mail: suhao@gate.sinica.edu.tw。感謝中央研究院資助作者執行97-98年前瞻計畫的〈依據次樣本加權的成效評估與無反應項目、隱藏因素之插補分析〉子計畫；行政院科技部資助社會學研究所張荳雲教授主持的「臺灣地區社會變遷基本調查五期四次全球化組」；以及（前）行政院研究發展考核委員會於民國97年7-9月執行的「數位落差調查」。本文能順利完成承蒙洪永泰老師的悉心指導、羅婉云小姐的協助分析、以及兩位匿名審查人寶貴的意見，僅此一併致上萬分的謝意。

收稿日期：2014/8/22，接受刊登：2015/2/9。

中文摘要

傳統的事後分層與多重反覆加權法可以降低抽樣誤差，但在改進推估偏差的效果仍然有限。本研究採用「入選機率調整法（propensity score adjustment, PSA）」，對2008年的電話訪問樣本進行18歲以上臺灣民衆之上網率的推估，進行上述兩種加權方法以及其納入PSA前後之估計效果的比較。在幾種入選機率值（propensity scores, PS）的調整方法中，本研究採用次樣本分組法，因此也同時比較四到十個分組之間估計效果的差異。

本研究以同一時段蒐集的「臺灣地區社會變遷基本調查五期四次全球化組」為PSA的參考樣本；（前）行政院研究考核委員會執行的「數位落差調查」所推估的全臺灣民衆的上網率為擬母體黃金標準，亦即PSA與兩種傳統加權之估計效果的比較基準。結果證實納入原調查權數（來自事後分層與多重反覆加權）後，在五個次樣本分組時，上網率的估計值最接近數位落差調查的估計值。比起納入事後分層權數，納入多重反覆加權權數的PSA在估計誤差上相對較小。

關鍵詞：事後分層、多變項反覆加權、入選機率調整法、次樣本分組

A Comparison of Propensity Score Sub-classification and other Calibration Methods based on a Telephone Sample to Estimate Internet Usage

Su-Hao TU

Center for Humanities and Social Sciences

Academia Sinica

Abstract

While post-stratification and raking calibration methods can reduce sampling errors, they have estimation limitations. The author adopts propensity score adjustment (PSA) to estimate Internet usage based on data collected from a 2008 telephone sample. Comparisons were made among post-stratification, raking, post-stratification PSA, and raking PSA. Stratification was used to produce PS weights and to compare estimated Internet usage for seven sub-classifications. The Taiwanese Social Change Survey conducted during the same period was used as the reference sample for PSA, while official statistics based on a Digital Divide Survey as the benchmark for bias reduction comparisons. Results indicate that (a) Internet usage estimates based on PSA adjusted according to base weight (i.e., survey weight from post-stratification or raking) using five sub-classes were more accurate than other estimates, and (b) bias reduction based on PSA adjusted by raking exceeded that of PSA adjusted by post-stratification.

Keywords: Post-Stratification, Raking, Propensity Score Adjustment, Sub-Classification

一、前言

一般的電話抽樣調查結果是否能代表抽樣母體主要受到兩大因素的影響，分別是抽樣底冊的母體涵蓋率與樣本無回應（Lee and Valliant 2008）。在抽樣底冊的母體涵蓋率方面，利用住宅電話號碼為代表母體的抽樣底冊時，會有兩個降低涵蓋率的來源。第一個來源是住宅電話的普及率低或使用率降低（Duncan and Stasny 2001）。普及率低主要是發生在低度現代化的社會中，對於住宅電話普及率高達約百分之九十六¹的臺灣來說並非關鍵的影響（洪永泰、洪百薰、林宇璇、呂孟穎、許勝懋、吳淑惠、卓仲彥、徐書儀 2014）。相反的，比較關鍵的問題主要是來自於手機越來越普遍，越來越多民衆不再裝設住宅電話，造成住宅電話的使用率逐漸下降，那麼同樣也會產生住宅電話涵蓋率降低的問題（Lee and Valliant 2008; Sirmath, Martin, Hoaglin, and Battaglia 2009）。第二個來源是不願意公開家戶電話號碼的比例。由於擔心個人基本資料外洩，在個資保護意識日益抬頭之際，不願意讓電信公司公開住宅電話號碼的家戶越來越多，涵蓋率勢必越來越低。至今已有修正涵蓋率不足的方法，亦即對電話號碼底冊中所抽樣出來的電話號碼進行末尾二到四碼的隨機撥號。但是，隨機撥號之後空號或非住宅電話號碼的比例上升，也因此衍生另一個涵蓋率降低的問題（洪永泰、黃永政 2000）。

影響電話抽樣調查結果是否具有樣本代表性的第二個重要因素是樣本無回應。可能產生回應率低的原因通常包括無人接聽、受訪者不在家、拒訪或其他無法順利進行電話訪談的中止原因。電話接通率降低可能是答錄或來電顯示功能的普遍使用，更方便一般民衆過濾電話，也可

¹ 根據行政院主計總處的官方調查資料，2011年臺灣地區家戶住宅電話的普及率是96.06%（洪永泰等 2014）。

能是因為家戶在同一電話號碼上用傳真機或上網專線的比率提高（Lee and Valliant 2008）。另外，電話接通後最主要的困擾是拒訪率。在臺灣各式各樣的民意調查、市場調查或選舉調查愈來愈多，加上詐騙案件越來越多，民眾自然更加強心理的防衛，拒訪率因此越來越高，也越來越不容易挽回拒訪的樣本個案（侯佩君 2008）。

要預防或解決以上影響樣本代表性的兩個重要的因素實屬不易，一般的作法多是在調查資料蒐集之後，以加權方法對調查結果進行事後的估計調整。各種加權方法包含事後分層（post-stratification）、多變項反覆加權（raking）、多重加權（multiple weighting）、線性加權（linear weighting）與GREG（generalized regression estimation）等（杜素豪、瞿海源、張荳雲 2012: 202-206）。這些加權方法多著重在利用某些社會人口特徵對樣本個案進行權重的調整，以達到樣本在這些社會人口特徵的分布能近似於母體。調整的結果是產生一個可以代表放大或縮小這些社會人口特徵之樣本個案的權數，權值小於一表示縮小樣本個案可以代表母體中相同（似）社會人口特徵者的權重。相反的，大於一則是放大權重。換言之，權值大小可代表更多或更少在相同（似）社會人口特徵之中未被涵蓋的調查樣本或未回應問卷題目的研究對象（Kalton and Flores-Cervantes 2003; Lee and Valliant 2008）。

不過，只著重在社會人口特徵的加權方法仍有改進的空間（Kalton and Flores-Cervantes 2003）。加權後，樣本中這些社會人口特徵的分布的確會接近於母體，但並不表示其他非社會人口特徵（如個人行為與態度）的分布也接近於母體（洪永泰 2007）。其主要的理由是推估到母體時，那些未被納入調查或無回應的樣本在某些行為或態度題目的回答可能不見得近似於相同社會人口特徵的受訪者（成功樣本）。此問題在投票率與政黨得票率的估計中已得到證實（劉從葦、陳光輝 2005）。在

針對多變項反覆加權法（*raking*）的研究也發現加權的效果有限（李隆安、韓成業 2008）。因此，若加權處理時能再考慮其他非社會人口特徵變項（例如與研究旨趣相關的認知、態度或行為特徵等），調查結果的估計應會更準確（洪永泰 2006）。

至今的確已有考慮了社會人口與非社會人口特徵變項的加權方法，以GREG（*generalized regression estimation*，有時稱為*linear weighting*）為例，² 其原理來自事後分層法，主要是藉由迴歸模型的建立計算權數，可避免某一分層樣本數為零而無法計算該分層權值的問題。其步驟首先是選定用來推估調查結果（亦即目標變項，*target variable*）的輔助變項（*auxiliary variables*，即加權所需的人口或非人口特徵變項），依據迴歸模型推估的結果進行權數的調整，權值大小的調整，其原則與事後分層一樣，調整到樣本特徵的分布與母體一致為止。迴歸模式推估的準確度受到輔助變項與目標變項之相關程度的影響，相關程度越高，估計標準誤越低（Kalton and Flores-Cervantes 2003; Bethlehem and Biffignandi 2012a: Chapter 10）。

但是，利用迴歸模式的加權方法仍有一些缺點。首先，很有可能所產生的權數是負值。其次，無法納入太多的輔助變項，造成操作複雜度提高。再者，加權所需來自於母體的輔助變項中，社會人口變項比較容易取得，但認知、態度或行為等相關的資料則得來不易（Särndal 2007）。若有一種調整方法可以克服以上傳統加權方法的限制（例如：輔助變項的數目，負的權數等），又可達到加權的效果，讓抽樣調查結果可以適當的調整到具有樣本代表性則會更好。

有鑑於此，根據文獻的建議，應用入選機率調整法（*propensity*

² 事實上事後分層是GREG的特例，其差異是輔助變項在前者只能是類別形式，而後者則不限（Kalton and Flores-Cervantes 2003; Bethlehem and Biffignandi 2012a）。

score adjustment, PSA) 於調查結果的推估會是可解決的辦法，尤其是來自非機率抽樣之自願性樣本的調查（例如網路調查）。PSA是利用另一個適當選擇的參考樣本（reference sample）與共變項（covariates），對觀察樣本進行條件機率的預測，再利用預測的機率值（入選機率值，propensity scores, PS）對樣本個案做適當的權重調整（Lee and Valliant 2008）。³ PSA應用在樣本推估的相關研究中，國外已有比較不同的主題、不同共變項、不同參考樣本、不同預測PS的迴歸模式、或不同PS調整方法的相關探討（Judkins, Morganstein, Zador, Piesse, Barrett, and Mukhopadhyay 2007; Lee 2006; Srinath, Martin, Hoaglin, and Battaglia 2009）；在臺灣則僅有針對投票結果的估計範例（杜素豪、羅婉云、洪永泰 2009）。其中，國外較少研究進行原機率抽樣調查的權數（原調查權數，base weight）納入PSA與否，以及納入不同原調查權數PSA之估計效果的比較（Duncan and Stasny 2001）。然而，對機率抽樣的調查樣本進行PSA時，這會是基本且重要的議題，本研究因此選擇臺灣尚無研究的主題——「臺灣地區民衆的上網率」，探討於PSA中納入原調查權數的必要性，並比較納入多變項反覆加權與事後分層後的估計效果。

具體而言，本研究利用一個全國性機率抽樣的電話訪問調查，比較多變項反覆加權、事後分層、未納入原調查權數的PSA、納入多變項反覆加權的PSA、與納入事後分層的PSA，五種上網率的估計結果。在評估利用PS修正樣本推估的成效時需要有好的且來自大樣本的母體估計值，以及許多調查同時都有問到的題目。使用「上網率」符合這個要求，可以比較各種估計方法以及不同資料蒐集方法的優劣。理論上，若本研究確定了最佳估計效果的PSA之後，此估計方式即可以適用於所有欲進行推估的結果變項，並非僅止於上網率的估計（Rosenbaum and

³ 關於PSA的原理、背景與應用，在下一節有更詳細的說明。

Rubin 1983)。由於我們並沒有臺灣民衆上網率的母體黃金標準，本研究以具有公信力、由政府統計調查加權後推估的臺灣民衆上網率為接近於母體的標準（擬母體黃金標準），當作估計效果的比較基準。此政府統計調查（『數位落差調查』）有多個特色：來自機率抽樣的樣本；樣本數大；由在訪問與資料蒐集品質有控管嚴密的民意調查機構執行；對成功樣本採兩階段嚴密的權數調整；在上網率的估計上不僅具備全國代表性，推估每一縣市的上網率時，也具備縣市的代表性。⁴

除此之外，本研究選擇具有不少優點且普遍使用的次樣本分組法進行PS的調整（Bethlehem and Biffignandi 2012a），比較不同估計方法的效果，也同時進行不同次樣本分組數的比較。雖然文獻建議用四到五分組即可以降低90%的估計誤差（Cochran 1968; Rosenbaum and Rubin 1984），這早期的研究結論主要是應用在樣本數較少的生物實驗。不過，理論上，估計誤差的降低也會受到研究樣本規模及PS相同之頻率的影响（Harder, Stuart, and Anthony 2010）。前人研究顯示，五分組時估計誤差的降低比例不見得可達到90%，反而是十分組後比較有可能（Imai and Dyk 2004）。多少分組數比較適合仍值得再探索，本研究因此進行四到十個次樣本分組的比較。

⁴ 因進行小區域樣本推估（small area estimation）所需，為了具備縣市的代表性，在某些縣市抽取的樣本數比較高，在未經加權的情況下，幾個人口特徵在縣市的樣本分布與母體的差異較大。此調查有兩套權數，其一適用於推估各縣市，另一則適用於推估全國（詳見該調查的執行報告）。本研究所引用的是後者，即加權後估計出來有全國代表性的民衆上網率，並非加權前的上網率，因此並不會影響PSA的比較基準。

二、文獻

入選機率調整法亦稱為傾向分數分析法（propensity score analysis），主要是有利於對無實驗設計中個案隨機分派機制的觀察性研究（observational studies）進行因果的推論或效果的分析。在缺乏控制組可以比對時，此法可以利用不同的配對或分組的技術，對非實驗法所蒐集的觀察性資料進行因果效應的估計（Rosenbaum and Rubin 1983, 1984; Rosenbaum 2002, 2005）。根據PSA的基本原理，若個案在某一原因變項（研究處理，treatment）上會不會受到影響（有沒有處理，treated or untreated），這兩種狀態之間的變異可以被一組與原因變項相關且可觀察的特徵變項（observed covariates，亦即共變項）所預測時，我們即可用這一組共變項，先以邏輯迴歸分析產生個案會受到影響（有處理，treated）的預測機率值（入選機率值，⁵亦即傾向分數，propensity scores, PS）。再利用入選機率值進行個案在兩種狀態之間的配對，最後計算觀察研究結果在兩種狀態的平均差異，估得因果（處理）效應（treatment effect）（Rosenbaum 2002, 2005）。理論上，配對的結果，各共變項在兩狀態之間即可平衡分布（balancing），亦即配對後的個案兩種狀態之間會具有相似的共變特徵，如此，我們可以獲得精確的估計（Zanutto 2006）。

異於一般因果推論的研究，PSA依據的是反事實架構（counterfactual framework）的理論觀點，不需要對共變項與觀察研究結果之間的關係做任何的理論性假設（例如線性或非線性），也不需要考慮共變項之間的共線性（collinearity）。只要選擇好理論上有可能與原因變項（研究

⁵ 本文命名為入選機率值而非傾向分數，主要目的是呈現此數值為經過邏輯迴歸模式所預測的機率值。

處理，treatment）相關的共變項，預測入選機率值時不需參考邏輯迴歸係數。另外，此法估計的不是個別個案的影響效果，而是平均的效應。對於不確定研究假設該如何設定的觀察法研究，這是一個比較能捕捉觀察事實，降低估計誤差的分析方法（Harding 2003）。在生物醫學相關領域中普遍運用已久（Rosenbaum 2002, 2005），也已應用在社會學相關議題的研究中（Czajka, Hirabayashi, Little, and Rubin 1992; Harding 2003; Zanutto 2006）。

將PSA的原理應用在樣本推估或調查結果，估計是利用可信賴的參考樣本（reference sample），透過有理論基礎所選擇的共變項，預測研究樣本（本文稱為觀察樣本）相對於參考樣本の入選機率值，再以配對、次樣本分組或其他方法調整入選機率值。從涵蓋率或無回應的觀點來看，觀察樣本假設為被母體涵蓋或有回應的樣本（treated group），⁶ 參考樣本則是未被母體涵蓋或無回應的樣本（untreated group）（Bethlehem and Biffignandi 2012b; Austin 2011）。具體的步驟是，第一、選擇一個經過嚴密之抽樣與加權的參考樣本。此參考樣本的母體涵蓋率與樣本回收率至少要比觀察樣本高，且是透過較嚴密且高品質控管之調查方法所蒐集的。第二、選擇與待推估之調查結果息息相關且參考樣本與觀察樣本中共同具有的共變項（社會人口與 / 或非社會人口特徵變項）。第三、利用邏輯迴歸分析預測入選機率值。第四、選擇一種入選機率值的調整方法，依據入選機率值對兩樣本個案做平衡性地配對。第五、對配對後的個案做必要的權重調整。第六、總計調整後的結果即得最後的估計值（Lee and Valliant 2008, 2009）。

承上的原理與步驟，PSA的優點是待估計的調查結果與入選機率值無關；入選到觀察樣本的結果不會影響共變項；樣本是否被涵蓋或接受

⁶ 亦即擬進行結果估計的調查樣本。

訪問不受任何因素的影響。如前所述，依據反事實架構的理論觀點，PSA不需對共變項與調查結果的關係做理論性的假設，類似資料採礦，更能貼近估計的實況。不過，估計的準確度會因共變項、參考樣本與入選機率值之調整方法的選擇而異；前兩者常是研究者無法改變的，需依賴現成的資料，入選機率值的調整方法則相反。

根據文獻，入選機率值的調整方法主要包含四種。第一種也是最早使用且應用最久的方法，就是依據相同或近似的入選機率值對觀察樣本與參考樣本進行個案的配對（*matching*）。最常見的是一對一的配對。但因配對的結果可否被替代，一對一配對的效率如何改進，以及如何精準配對又延伸其他多種的一對一或一對多的配對方法，例如：貪婪配對（*greedy matching*）、最鄰近配對（*nearest neighbor matching*）、最適配對（*optimal matching*）、以及完全配對（*full matching*）等等，共有超過20種的配對方法（Rosenbaum 2002; Austin 2011, 2014）。⁷不過，配對方法是需要花較多的時間處理，無法配對到個案的比率會比較高，尤其是選擇一對一或不替代時。

第二種是入選機率倒數（權數）法（*inverse propensity scores*）。⁸主要是利用入選機率值（PS）的倒數當作權數，取代原觀察樣本的權數。此法是在已知觀察樣本之抽樣機率時比較可行，其缺點是所產生的PS權數若有明顯的極端值時（例如入選機率值非常低，或者處理的觀察樣本個案數少），估計誤差會變大（Schonlau, Soest, Kapteyn, and Couper 2009; Stuart 2010; Austin 2011; Bethlehem and Biffignandi 2012b）。第三

⁷ 各配對方法均各有優缺，讀者可參見本文的參考書目或更多其他的論文，更深入瞭解其實際步驟與推薦的方法，內文不再詳述。

⁸ 文獻中也有*inverse probability weights*，*response propensity weighting*等不同的名稱。

種是多變量調整法 (covariate adjustment)。主要是將入選機率值當作自變項，納入迴歸模式分析欲估計的結果變項後所轉換而得的權數。這是唯一需要假設入選機率值與結果變項之間具關係線性與否的方法。也因此產生另一層如何恰當地設定適當迴歸模式的不確定性，為較不普遍的方法 (Austin 2006, 2011)。

第四種是次樣本分組 (subclassification或Stratification)，是依據入選機率值的排序對觀察樣本與參考樣本的個案進行次樣本分組 (Luellen, Shadish and Clark 2005; Austin 2011)。事實上次樣本分組是多對一的配對方法，較容易操作、有效率、且為相關研究普遍使用。主要是此法可單純地將樣本個案分組；可利用一般統計軟體設計程式語法；且方便於PSA的步驟中納入來自複雜抽樣設計的權數 (即原調查權數，base weight或survey weight)，獲得最適的估計值 (Yanovitzky, Zanutto and Hornik 2005; Zanutto, Lu, and Hornik 2005)。若不考慮原調查權數，此法簡單來說就是依據每一次樣本分組中觀察樣本與參考樣本在案數上的相對比例，以及兩樣本總數的相對比例，產生每一次樣本分組的修正係數 (即PS權數)。然後利用修正係數對每一次樣本分組中待估計變項的案數做同步的調整。總合各次樣本分組中調整後的新案數即可獲得觀察樣本的PSA估計值 (Lee and Valliant 2008)。⁹ 其優點主要是分配入選機率值時，可以讓每一分組中包含來自參考樣本與觀察樣本的總案 (人) 數相等或幾乎相等；每一分組個別產生估計值 (效果)，總估計值可平均掉所有分組在組內兩樣本之間的差異。綜合以上，若入選機率值與估計變項間存在非線性關係時，會產生比起其他方法更小的估計偏差 (Luellen, Shadish and Clark 2005; Yanovitzky, Zanutto and Hornik 2005;

⁹ 詳細的步驟與公式可參考Lee (2006) 與杜素豪、羅婉云與洪永泰 (2009)。不過這兩篇論文採用的是模擬分析，與本文相異，請讀者注意。

Bethlehem and Biffignandi 2012b)。

PSA在國外已熱烈討論，其主題多圍繞在檢驗PSA是否可以超越傳統加權方法只利用社會人口變項產生權數，容易造成樣本代表性失真的問題。前人研究之一是透過模擬分析，且多半將PSA應用在網路調查的自願性樣本（Lee 2006; Schonlau, van Soest, and Kapteyn 2007; Lee and Valliant 2009; Schonlau et al. 2009）；其次是電話調查樣本（Duncan and Stasny 2001; Srinath et al. 2009）。也有比較網路與電訪調查樣本之PSA效果（Schonlau et al. 2004）。其他的主题則包括PSA的理論與技術性介紹（Lee 2006; Lee and Valliant 2009; Bethlehem and Biffignandi 2012a）以及 PSA與傳統加權方法的比較（Judkins et al. 2007; Loosveldt and Sonck 2008; Srinath et al. 2009）。

綜合前人的研究，在估計政治態度、投票、種族態度、家庭財產、以及健康狀況、行為與態度等各種主題上已確定PSA可降低估計誤差。若研究目的是在估計實驗或影響效果，PSA步驟中不納入原調查權數（base weight）是適當的，而推估的調查樣本沒有原調查權數可納入，也是沒辦法的。可是，若目的是調查樣本的推估而且有已知的原調查權數時，爲了提高估計的準確度，則有必要在進行入選機率值（PS）調整並產生PS權數時，利用已知的原調查權數計算出最後修正的PS權數（adjusted weight）（Zanutto, Lu, and Hornik 2005; Lee 2006; Lee and Valliant 2009）。如此修正的結果，根據前人針對電訪的研究結果顯示，納入原調查權數後，PSA推估誤差的降低幅度比未納入原調查權數或其他傳統加權者還要高（Duncan and Stasny 2001）；在推估網路調查的模擬分析則發現推估誤差的降低幅度最高可達80%或95%（Lee 2006; Lee and Valliant 2009）。

在臺灣的相關研究顯示，利用多變項反覆加權法於推估面訪調查

的效果有限，不見得準確（劉從葦、陳光輝 2005；李隆安、韓成業 2008）。至今已有發展能考慮到多層抽樣設計，且比事後分層和 / 或多變數反覆加權效果更佳的「最小差異加權法」（黃紀、張佑宗 2003）與「多層次貝氏定理估計的事後分層加權」（俞振華、蔡佳泓 2006）。但是，在比較各種電訪加權方法的研究中也顯示推估的結果差異不大（洪永泰 2007）。有鑑於臺灣電話調查之涵蓋率與樣本回應率日益降低，杜素豪、羅婉云與洪永泰（2009）引進PSA，也參考Lee（2006）的模擬分析步驟，進行各總統候選人得票率在選後調查的估計，評估結果顯示PSA雖然增加推估的變異（標準誤），仍可有效降低估計誤差。不過，此研究也僅止於模擬分析，相對於國外諸多相關研究，仍有很多方法相關的議題尚待研討。

三、資料與分析

（一）觀察樣本、參考樣本與擬母體黃金標準

不同資料的選擇的確會影響估計的可靠性。根據PSA的原理與本研究的目的，關鍵主要在於參考樣本。換言之，選擇不同的觀察樣本與擬母體黃金標準，PSA的估計成效會有程度上的差別。觀察樣本的選樣誤差越大，PSA估計成效會更好。本文只引用數位落差調查經加權後估計的全國上網率為各式PSA估計之成效的比較基準，若有真正母體的上網率，無誤地呈現PSA的成效會更理想。不過，觀察樣本與擬母體黃金標準的選擇對本研究評估不同估計方法的優劣，其結論應該不會有所改變。但是，選用不同的參考樣本時，對本研究在推估方法之比較的結果會有影響。因此，選擇理想或適當的參考樣本很重要。不過，不同參考

樣本的比較有待未來的驗證，非本研究探討的範圍。可以確定的是本研究已經根據PSA的基本原理，選擇了符合條件的參考樣本。至於觀察樣本，本研究選用的抽樣調查資料主要是作者所執行，且經過機率抽樣與嚴密執行之調查方法所蒐集的。而擬母體黃金標準則選擇在臺灣唯一可取得、政府所公告、數據可信度較高、來自大規模抽樣調查結果所估計出來的民衆上網率。

本研究的觀察樣本、參考樣本與擬母體黃金標準來自三個抽樣調查資料。觀察樣本是民國97年7-8月執行的「前瞻計畫電話訪問調查（以下簡稱「前瞻電訪」）」，¹⁰ 使用電腦輔助電話訪問系統（CATI）進行調查訪問，以臺灣地區年滿18歲及以上且家中有電話之民衆為調查對象，以臺灣地區住宅用戶電話號碼中華電信93-94年住宅電話資料庫為抽樣底冊，進行分層兩階段系統抽樣（stratified systematic sampling）。首先以臺灣23個縣市作23個分層，根據行政院內政部統計處民國96年12月底公布之人口統計資料中的縣市人口數，以等比例原則分配各分層需抽取的人數。其次，利用全國電話資料庫以系統抽樣原則隨機抽出各縣市所需的住宅電話號碼，並進行末尾兩碼的隨機撥號。最後利用洪氏戶中抽樣法（洪永泰 1996），針對戶中合格受訪者之男女兩性所占人數，抽出受訪對象，最後完訪1,050案。

參考樣本來自同時段執行的「臺灣地區社會變遷基本調查五期四次全球化組（以下簡稱「變遷面訪」）」，以臺灣地區年滿18歲及以上之中華民國國民（78年12月31日以前出生者）為母體，臺灣地區戶籍資檔為抽樣底冊（sampling frame），分層三階段PPS抽樣法（probabilities proportional to size, PPS）抽出受訪對象，完訪數為2,067（張荳雲、廖培

¹⁰ 資料來源：〈依據次樣本加權的成效評估與無反應項目、隱藏因素之插補分析〉研究計畫，97-98年中央研究院前瞻主題計畫子計畫一，計畫總主持人李隆安。

珊 2008)。

擬母體黃金標準¹¹來自(前)行政院研究發展考核委員會於民國97年7-9月執行的「數位落差調查」,經加權後所估計出來具有全國代表性的民衆上網率。此調查採電腦輔助電話訪問系統(CATI)進行訪問,以聯合報系民意調查中心97年3月更新之臺灣地區25縣市住宅電話用戶名冊作為抽樣底冊,再以25縣市作為副母體進行隨機抽樣,並對抽出的電話號碼進行末尾兩碼的隨機撥號。調查對象為居住在臺閩地區之普通住戶內年滿12歲之本國籍民衆,各縣市預定樣本數是以內政部統計處民國97年7月公布之各縣市12歲以上人口數為計算標準,在推估各縣市95%的信心水準與抽樣誤差小於±4%前提下配置樣本數為15,000份。另外,針對臺北市、臺北縣、桃園縣及高雄市等人口數眾多縣市補足樣本至800(抽樣誤差±3.5%)。最後完成的有效樣本數為16,131。由於觀察樣本與參考樣本均為臺灣地區年滿18歲的國民,本研究引用的是刪除12-17歲與金馬地區樣本後(樣本數為13,238),經加權估計的上網率(行政院研究發展考核委員會 2008)。

(二) 加權與插補

本研究以性別、年齡、教育與設籍的地理區域,對觀察樣本進行多變項反覆加權(raking)與事後分層,產生兩種原調查權數(base weight)。Raking是依序以性別、年齡(18-29歲、30-39歲、40-49歲、50-59歲與60歲及以上共五個年齡層)、教育(國小及以下、國初中、高中職、大專及以上共四個教育等級)、以及地理區(北北基、桃竹苗、中彰投、雲嘉南、高屏澎與宜花東共六個地理區)進行調整,直到

¹¹ 此擬母體異於前人研究中以模擬分析產生的擬母體,詳見Lee(2006)。

每一變項的樣本代表性檢定結果與母體無統計顯著性的差異為止。

事後分層是以性別、年齡（五分類，18-29歲、30-39歲、40-49歲、50-59歲、60歲及以上）、教育程度（三分類，國中及以下、高中職、大專及以上）與地區（四分類，北、中、南、東）四個變項進行交叉分層，產生各分層的權值。但因觀察樣本數僅1050，四維交叉分析後，產生某些分層的案數小於五甚或等於零的問題，本研究進一步合併某些年齡、教育或地區分層。¹²

參考樣本採用與觀察樣本相同的四個變項進行多變項反覆加權。¹³ 而當作擬母體黃金標準的數位落差調查是以性別、年齡進行raking，除了進行縣市內樣本的加權外，另依據各縣市12歲以上人口占全國12歲以上人口比例進行第二階段加權還原。除了可推估到在隨機抽樣中無法取得足夠推論之樣本數的縣市之外，也可推論到全國。

在預測入選機率值（PS）的迴歸分析中，無論是來自觀察及參考兩樣本的所有共變項均需無任何遺漏值。¹⁴ 一般而言，處理遺漏值的方式有二：刪除遺漏值與插補。PSA是藉由控制觀察與參考樣本的多個共變項所產生的入選機率值，做兩樣本分組的配對，使兩樣本在這些共變項

¹² 詳細的分類合併方式不在文內說明，歡迎對此有興趣的讀者向作者洽詢。另外，雖然職業是重要的分層變項，但由於需要變數交叉分布後的母體數據，臺灣並沒有地區、性別、年齡、教育程度經四維交叉後各細小分層中的職業分布數據。即便有，在樣本數不足的情況下，無法符合多個變數交叉後每一分層細格需含五案以上的要求。

¹³ 與社會變遷調查的加權原則相同，唯本研究採的地區是地理區域而非抽樣分層，目的是獲得較正確合理的地理區代表性。

¹⁴ 觀察樣本的年齡有3案，教育程度2案，地區17案，省籍16案，目前無工作1案，目前工作行業16案，目前工作職業19案。參考樣本的省籍有3案，目前工作行業1案，目前工作職位1案。

上建立平衡的分布，由此在統計上控制了變異後即可達到縮小選擇性的偏誤（Rosenbaum and Rubin 1983）。刪除遺漏值可能會影響預測入選機率值的精準度，因此本研究選擇插補以取得完整的共變項數據。

在各種插補法中，尤以多重插補法最為完整，可改善單一插補法的缺點，透過對同一遺漏資料進行多次估計，將多次估計結果產生整體之平均數來插補遺漏值，可使遺漏值誤差降至最小。本研究採多重插補法，以迴歸方程式於每一個遺漏值分別產生出五個插補值，獲得五份完整的資料集，將所有資料集結果合併產生整體之平均數來插補各共變項中的遺漏值。其優點是獲得一個更大且更具代表性的樣本進行PSA，避免遺漏值所造成的推估偏誤（Rubin 1987; Sinharay, Stern, and Russell 2001; Little and Rubin 2002 ;Yanovitzky, Zanutto, and Hornik 2005）。

（三）變項與描述統計

依據PSA的原理，我們主要是利用參考樣本與觀察樣本共同的共變項所預測而得的入選機率值為個案配對的依據。雖然「變遷面訪」中也有上網率的資料，而「數位落差調查」也有相似的共變項，但均非PSA所需（Zanutto 2006）。因此，以下僅說明擬黃金母體與觀察樣本在上網率的題目設計，以及觀察樣本與參考樣本共同之共變項的測量方式。關於上網率，「數位落差調查」的問法是「請問您目前平均一天花多少時間上網？」，回答是「上網的分鐘數」；而「前瞻電訪」的題目是「請問您最近一個月內，平均每天會上網多久？」，受訪者的回答是「上網的小時與分鐘數」。雖然回答方式與參考時間點不太一樣，但經過資料轉換最後可以得出目前或最近一個月受訪者上網的百分比。

本研究選用的共變項包括性別、年齡、教育、省籍、地區、有無工

作、行業、職位，以及是否會使用電腦。每一變項的分類與對照組均在表1中說明。從表1可知，參考樣本的男女比例相近，觀察樣本的女性比例高於男性（13.4%）。在年齡方面，參考樣本中18-29歲、60歲及以上的比例比較高，其他各年齡層所佔比例則是在觀察樣本中比較高。兩樣本均是高中職及大學以上，臺灣客家與臺灣原住民以外的族群，以及設籍於北北基的比例最高；有工作者的比例也都有六成。在行業方面，兩樣本均是無工作的比例最高，其次依序是服務業、商業運輸及通信業、製造業。兩樣本在各職位的分布除了無工作之外，稍有差異，觀察樣本次高的是助理（半）專業人員，參考樣本則是管理人員與服務工作人員及售貨員。觀察樣本中會使用電腦的比例高於參考樣本達20%。

表 1 觀察樣本與參考樣本變項的描述

變項	觀察樣本 前瞻電訪			參考樣本 變遷面訪		
	n	%	mi2(%)	n	%	mi2(%)
樣本數	1,050			2,067		
性別						
男性	455	43.3	43.3	1030	49.8	49.8
(女性) ¹	595	56.7	56.7	1037	50.2	50.2
年齡						
18-29歲	166	15.9	15.8	461	22.3	22.3
30-39歲	213	20.3	20.3	393	19.0	19.0
40-49歲	273	26.1	26.0	400	19.4	19.4
50-59歲	210	20.1	20.3	390	18.9	18.9
(60歲以上) ¹	185	17.7	17.6	423	20.5	20.5
教育程度						
國小及以下	165	15.7	15.7	446	21.6	21.6
國(初)中	117	11.2	11.2	242	11.7	11.7
高中職/士官學校	299	28.5	28.6	536	25.9	25.9
專科	189	18.0	18.0	280	13.5	13.5
(大學及以上) ¹	278	26.5	26.5	563	27.2	27.2

表 1 觀察樣本與參考樣本變項的描述 (續)

變項	觀察樣本 前瞻電訪			參考樣本 變遷面訪		
	1,050			2,067		
樣本數	n	%	mi2(%)	n	%	mi2(%)
省籍						
臺灣客家	119	11.5	11.4	257	12.5	12.4
臺灣原住民	10	1.0	1.6	33	1.6	1.6
(其他) ¹	905	87.5	87.0	1774	85.9	85.9
地區						
北北基	318	30.8	30.3	519	25.1	25.1
桃竹苗	145	14.0	14.1	330	16.0	16.0
中彰投	192	18.6	19.0	409	19.8	19.8
雲嘉南	173	16.7	17.0	343	16.6	16.6
高屏澎	157	15.2	15.0	367	17.8	17.8
(宜花東) ¹	48	4.6	4.6	99	4.8	4.8
有工作	660	62.9	63.0	1341	64.9	64.9
行業						
農林漁牧 / 礦業及土石採取業	59	5.7	5.6	75	3.6	3.6
製造業	149	14.4	14.5	367	17.8	17.8
水電燃氣業 / 營造業	48	4.6	4.9	100	4.8	4.8
商業 / 運輸倉儲及通信業	181	17.5	18.0	376	18.2	18.2
服務業	208	20.1	20.0	422	20.4	20.4
(無正式工作者) ¹	389	37.6	37.0	726	35.1	35.1
職位						
管理人員	84	8.1	8.0	244	11.8	11.8
專業人員	94	9.1	9.1	124	6.0	6.0
助理(半)專業人員	129	12.5	12.4	192	9.3	9.3
事務工作人員	77	7.5	7.5	165	8.0	8.0
服務工作人員及售貨員	70	6.8	7.2	216	10.5	10.4
農林漁牧	60	5.8	6.4	51	2.5	2.5
技術工及有關工作人員	41	4.0	4.0	126	6.1	6.1
機械設備操作工及組裝工	46	4.5	4.4	119	5.8	5.8
非技術工	41	4.0	3.9	94	4.5	4.5
軍人	0	0	0	9	0.4	0.4
(無正式工作者) ¹	389	37.7	37.0	726	35.1	35.1
(會使用電腦) ¹	911	86.8	86.8	1326	64.2	64.2

註：1. 括號內的變項組別為入選機率邏輯迴歸模型中的對照組，分別是女性、60歲及以上、大學及以上、其他省籍、宜花東地區、目前有工作、行職業為無正式工作者、會使用電腦。

2. mi是經由多重插補後的結果。

(四) PSA步驟

準備好了參考樣本 n^R 與觀察樣本 n^T 之後，首先是合併兩樣本，再利用邏輯迴歸分析，產生對照於參考樣本之觀察樣本的預測機率值（亦即入選機率值PS）。其次是產生PS權值，本研究採次樣本分組法，依據排序（rank）後的入選機率值均分兩樣本為四到十個次樣本分組，每一分組所包含面訪參考樣本與觀察樣本的總數原則上均相同。¹⁵再者，本研究依據Lee（2006），以面訪參考樣本raking後的各組比例修正原調查權數加權後觀察樣本中各組的比例，即得修正係數（ fc ）（見公式（1））。¹⁶最後，此修正係數再進行電訪樣本原調查權數（ d_j^T ）的調整，即得最後修正的PS權數（adjusted weight, $d_j^{T.PSA}$ ）（見公式（2））。

$$fc = \frac{\sum_{k \in (S_c^R)} d_k^R / \sum_{k \in (S^R)} d_k^R}{\sum_{j \in (S_c^T)} d_j^T / \sum_{j \in (S^T)} d_j^T} \equiv \frac{\hat{N}_c^R / \hat{N}^R}{\hat{N}_c^T / \hat{N}^T} \quad (1)$$

$$d_j^{T.PSA} = fcd_j^T = \frac{\hat{N}_c^R / \hat{N}^R}{\hat{N}_c^T / \hat{N}^T} d_j^T \quad (2)$$

S_c^R = 次樣本分組面訪參考樣本， S_c^T = 次樣本分組電訪觀察樣本，
 d = 原調查權數。

¹⁵ 例如分四組時，每一分組包含兩樣本的組數為兩樣本總案數的25%。不過，考慮相同PS值者歸入同一分組比較適當，各分組的總案數多少會有一兩案，非常少數的差異。

¹⁶ 此修正係數（ fc ）亦即PS權數，若屬無原調查權數的修正係數（ fc ），因非本文重點，讀者可參見Lee（2006）一文。

最後是產生上網率的PSA估計值。根據各次樣本分組中最後修正的PS權數則可依據公式(3)調整電訪觀察樣本中「上網率」的估計值。

$$\hat{y}^{T.PSA} = \frac{\sum_c \sum_{j \in (s_c^T)} d_j^{T.PSA} y_j}{\sum_c \sum_{j \in (s_c^T)} d_j^{T.PSA}} \quad (3)$$

四、結果

如文前所述，本研究主要的目的是比較兩種傳統加權法單獨或納入PSA之五種上網率的調整結果。調整的成效是依據擬母體黃金標準（數位落差調查的估計值）與五個估計值的比較結果。同時，本研究也比較PSA不同次樣本分組數之間的估計效果。在說明各種上網率估計方法的成效之前，需先說明PSA步驟中，PS經次樣本分組後，各分組內兩樣本案數的分布以及修正係數的變化，以及納入原調查權數之前與之後的差異。

表2顯示四到十個次樣本分組後原則上兩樣本總案數在各分組之間均需平均分布。依據兩樣本總數為3117（即參考樣本2067案與觀察樣本1050案的總和），分成四分組時每一分組總數平均約為779（即總數的25%），五分組時約為623，依此類推到十分組時約為389。若相同PS值被歸入不同分組，實屬不恰當，本研究因此彈性地調整，將之歸入相同的分組內。另外，次樣本分組是在兩樣本加權之前，依據迴歸預測所得的PS值的結果進行的，因此表2所顯示的觀察樣本均分於各分組的案數在加權前後會有很小的差異是正常的。

表 2 PSA入選機率值的分組與各式修正係數

分組	參考樣本 變遷面訪	觀察樣本					
		前瞻電訪 (未加權)	fc^1	前瞻電訪 (raking)	$rrfc^2$	前瞻電訪 (PoS)	$ppfc^3$
四分組							
1	697	82	4.318	105	4.815	103	4.778
2	562	214	1.334	250	1.34	255	1.327
3	478	307	0.791	305	0.767	293	0.781
4	330	447	0.375	390	0.339	399	0.33
總計	2067	1050		1050		1050	
五分組							
1	575	54	5.409	71	5.944	69	5.944
2	459	150	1.554	177	1.63	182	1.589
3	429	204	1.068	224	1.063	221	1.069
4	361	266	0.689	247	0.654	240	0.665
5	243	376	0.328	331	0.291	338	0.286
總計	2067	1050		1050		1050	
六分組							
1	492	34	7.351	44	7.889	41	7.778
2	408	112	1.851	135	2	137	1.975
3	359	150	1.216	176	1.219	180	1.203
4	339	190	0.906	195	0.905	190	0.925
5	271	242	0.569	218	0.51	212	0.51
6	198	322	0.312	282	0.284	290	0.277
總計	2067	1050		1050		1050	
七分組							
1	436	20	11.074	26	12.667	24	12.667
2	349	87	2.038	107	2.143	106	2.143
3	319	118	1.373	140	1.361	144	1.361
4	304	147	1.051	166	1.063	163	1.063
5	275	178	0.785	172	0.781	167	0.781
6	219	220	0.506	186	0.442	186	0.442
7	165	280	0.299	253	0.271	259	0.271
總計	2067	1050		1050		1049	
八分組							
1	392	12	16.594	16	18.5	14	18
2	306	70	2.221	90	2.44	89	2.44
3	283	97	1.482	111	1.525	116	1.475
4	279	117	1.211	138	1.209	140	1.2

表 2 PSA入選機率值的分組與各式修正係數（續）

分組	參考樣本	觀察樣本					
		變遷面訪	前瞻電訪 (未加權)	fc^1	前瞻電訪 (raking)	$rrfc^2$	前瞻電訪 (PoS)
5	262	135	0.986	141	0.981	136	0.99
6	216	172	0.638	164	0.605	157	0.614
7	188	200	0.478	163	0.44	161	0.433
8	141	247	0.29	227	0.263	238	0.263
總計	2067	1050		1050		1051	
九分組							
1	352	8	22.351	11	22	9	19
2	278	57	2.478	74	2.85	73	2.75
3	270	81	1.693	94	1.714	96	1.714
4	232	99	1.19	119	1.198	115	1.16
5	246	107	1.168	117	1.172	120	1.194
6	220	134	0.834	136	0.832	135	0.842
7	191	155	0.626	140	0.564	133	0.574
8	156	184	0.431	154	0.39	152	0.39
9	122	225	0.275	205	0.244	216	0.244
總計	2067	1050		1050		1049	
十分組							
1	319	4	40.512	5	0	5	0
2	255	50	2.591	66	2.833	64	2.833
3	245	64	1.945	73	2	73	2.05
4	215	86	1.27	104	1.321	109	1.264
5	225	92	1.242	107	1.221	107	1.233
6	204	112	0.925	117	0.92	113	0.932
7	195	122	0.812	114	0.81	111	0.81
8	166	144	0.586	133	0.531	130	0.541
9	133	174	0.388	143	0.355	144	0.345
10	110	202	0.277	188	0.25	193	0.236
總計	2067	1050		1050		1049	

註：1. fc ：入選機率值（PS）經排序分組，觀察樣本未加權時所產生的修正係數（PS權數）。

2. $rrfc$ ：入選機率值（PS）經排序分組，以觀察樣本經raking之原調查權數所調整後的最後修正PS權數（adjusted weight）。

3. $ppfc$ ：入選機率值（PS）經排序分組，以觀察樣本經事後分層（PoS）之原調查權數所調整後的最後修正PS權數（adjusted weight）。

在表2中，未加權以及經raking與事後分層調整後的三種修正係數的計算方式如下步驟。首先，分別將參考樣本該分組內的案數除以參考樣本的總案數，以及觀察樣本的該分組案數除以觀察樣本的總案數。其次，再將前者除以後者。以四分組未加權觀察樣本為例，第一組的修正係數4.318，亦即 $(679/2067) / (82/1050)$ 。在觀察樣本數較少的分組內，修正係數的數值比觀察樣本數較多之分組內的數值還要高，且分組數越多，其數值高出更多。分組數增加時，最少觀察樣本數那一分組內的案數明顯減少，十分組中甚至在加權後的案數等於零。這顯示分組數增加後PSA的成效不見得更好。無論是組間案數的分配或修正係數，在未加權與兩種加權後的差異不大。不過，確定的是加權後的修正係數比較高，且Raking後又比事後分層後還要高。

接下來比較各種上網率的估計值，表3中從左到右的欄位分別顯示分組數、「前瞻電訪」觀察樣本在未加權、多變項反覆加權、事後分層、PSA、Raking 後PSA、以及事後分層後PSA，共六種上網率以及其分別與擬母體黃金標準（數位落差調查估計的上網率55.927%）的差異。首先，「前瞻電訪」的上網率在未加權之前為59.905%，低於數位落差的估計結果近4%。經raking後降為58.008%，相距更小，經事後分層則下降更多（57.624%），也更接近「數位落差調查」估計值。

其次，進一步與三種PSA的估計結果比較則顯示，無論哪一種PSA，分成六個或更多個分組時，估計所得的上網率反而低估，且估計誤差反而比兩個傳統加權法的還要大。分組數越多時估計誤差越大，尤其是十個分組時估計誤差大幅增加。再者，無論是四或五個分組，三種PSA的估計誤差均小於兩種傳統加權約0.3%到1.68%不等，其中又以raking 後PSA的估計誤差最低。但未納入原調查權數之PSA與事後分層後PSA的估計誤差在四個分組與五個分組之間有些差異。四個分組時經

表 3 電訪觀察樣本的y估計值（上網率）：調整前後的比較

參考樣本：變遷面訪						
觀察樣本前 瞻電訪分組	未加權	Raking	PoS ¹	PSA	Raking +PSA ²	PoS +PSA ³
四分組	59.905%	58.008%	57.624%	54.952%	55.048%	54.909%
相差% ⁴	-3.977%	-2.081%	-1.697%	0.975%	0.880%	1.018%
五分組	59.905%	58.008%	57.624%	55.439%	55.524%	55.481%
相差%	-3.977%	-2.081%	-1.697%	0.489%	0.404%	0.446%
六分組	59.905%	58.008%	57.624%	53.714%	53.677%	53.283%
相差%	-3.977%	-2.081%	-1.697%	2.213%	2.250%	2.645%
七分組	59.905%	58.008%	57.624%	51.908%	51.862%	51.668%
相差%	-3.977%	-2.081%	-1.697%	4.019%	4.065%	4.259%
八分組	59.905%	58.008%	57.624%	52.376%	52.637%	51.901%
相差%	-3.977%	-2.081%	-1.697%	3.551%	3.291%	4.026%
九分組	59.905%	58.008%	57.624%	52.050%	51.870%	51.143%
相差%	-3.977%	-2.081%	-1.697%	3.878%	4.058%	4.785%
十分組	59.905%	58.008%	57.624%	50.333%	49.574%	49.762%
相差%	-3.977%	-2.081%	-1.697%	5.594%	6.353%	6.165%

註：1. 事後分層。

2. 經raking之原調查權數修正後的PSA。

3. 經事後分層（PoS）之原調查權數修正後的PSA。

4. 擬母體黃金標準(數位落差樣本經raking後)的上網率55.927%減電訪觀察樣本調整前或後之上網率的結果。

事後分層的PSA估計誤差反而比單獨PSA還要大。但五個分組的結果顯示，納入兩種原調查權數的PSA在上網率的估計上，均比未納入原調查權數之單獨PSA的結果接近於擬母體黃金標準。綜合以上，以PSA經五個分組的估計誤差較小，此結果呼應文獻所建議的分組數（Cochran 1968），而最佳的估計值來自五個分組時用raking後PSA的估計結果。

五、討論與建議

本研究採用文獻推薦且已普遍應用在電訪與網路調查的「入選機率調整法 (propensity scores adjustment, PSA)」，進行包含兩種傳統加權方法與其納入PSA之幾種上網率估計的比較，同時也比較估計值在次樣本分組法中四到十個分組之間的差異。結果發現，首先，單獨raking或事後分層加權的估計誤差均比PSA在四或五個分組的結果還要大，但若超過五個分組時，則估計誤差反而較小。其次，各分組的PSA估計結果中以分成五組的效果最佳，其估計誤差最小，次為四分組。雖然納入與未納入兩種傳統加權的權數於PSA的估計誤差無明顯的差異（不到1%），從估計誤差消滅比例來看，raking後PSA的估計誤差比事後分層後更小。

本研究證實納入原調查權數後在五個次樣本分組的PSA對上網率的估計效果最好，又以raking加權後PSA更好。欲調整電訪估計的準確度，採五分組的入選機率調整法（無論是否納入原調查權數）確實是優於傳統的raking與事後分層加權法。從表3中相差的比例來看，PSA可以修正1.208%到1.677%之間的估計誤差，呼應了前人研究結果，將反事實架構觀點的PSA應用在上網率的估計是可推薦的（Lee and Valliant 2008）；PSA可以比傳統加權法對電訪樣本做更有效的推估（Srinath et al. 2009）且納入傳統加權法的估計成效更好（Duncan and Stasny 2001）。本研究也再次證實採次樣本分組法時，早期文獻所建議，曾經被質疑的五分組仍是最推薦的，主要是其估計誤差消滅的比例在其他分組數是最大的，而且此分組數是經過前人研究利用不同性質資料所建議過的，並非因為純粹本研究在資料或所選之PSA方法上的特殊性（Cochran 1968; Rosenbaum and Rubin 1983, 1984; Rosenbaum 2005; Imai

and Dyk 2004; Bethlehem and Biffignandi 2012b)。因此五分組是後續研究進行PS次樣本分組時可參考的方式。

儘管以上的結論，對於不同方法之調查結果的估計或未來相關研究來說PSA並非完美，使用時仍有些面向需加以注意。首先，本研究證實PSA為不錯的選擇是依據電話調查樣本之上網率的估計誤差，雖然呼應文獻，PSA優於傳統加權法，對於無機率抽樣基礎的調查來說，無納不納入原調查權數於PSA可言，也沒有兩種傳統加權PSA可比較，因此無法推論到如自願性樣本的網路調查樣本的估計上。

第二、資料選擇的確會影響PSA估計的可靠性，但根據PSA的原理，關鍵不在觀察樣本與（擬）黃金母體的選擇，而是參考樣本。並不會因為觀察樣本或（擬）黃金母體的不同，不同估計方法之優劣比較的結果會有差異。若選用其他的參考樣本，估計效果的確可能會不同，但是，本研究選用的參考樣本來自調查執行與資料品質管控嚴密的全國性大型調查樣本，符合理想參考樣本的條件（Lee and Valliant 2008, 2009），因此可提供未來研究或PSA估計的參考，不過，是否還有其他比社會變遷調查樣本還要適合的參考樣本，值得未來繼續探討。再者，未來若對其他觀察樣本之上網率進行PSA估計，將結果也與傳統加權的結果比較，能再次驗證本研究結果會更理想。而是否有更適合之上網率的黃金標準可作為PSA成效的比較基準是可遇不可求的，也有待繼續尋找。

第三、預測PS的迴歸模式中該納入哪些共變項仍有討論的空間。理論上，所選擇的共變項與待估計的結果變項之間的相關性需小甚或無，但文獻中對於與結果變項（如：本研究的上網率）息息相關的共變項是否需納入PS迴歸預測中，說法不一（Lee and Valliant 2009; Austin 2011），有必要繼續驗證。本研究雖沒有態度相關變項，但納入了行為

相關變項（電腦使用能力），若可找到可以考慮納入上網行為的態度變項也許能更加提升對上網率的推估成效，不過這還需要更進一步的實證確定（Schonlau, van Soest, and Kapteyn 2007）。

第四、選擇理想的共變項對於估計的準確度來說是關鍵，只是，PSA也需要有理想的參考樣本加以配合，無論是在抽樣設計或可用的共變項上均符合需求。在現實中，我們常是利用已經蒐集完成的調查樣本，即便抽樣設計上符合要求，此調查資料中不見得有符合PSA需求的共變項。因此，預先規劃好的研究設計（類似早期應用PS法於生物實驗的研究設計）變得很重要。

第五、PSA的成效在大樣本中會比較明顯，但是相對的選用的共變項中，遺漏值可能較多，如何處理遺漏值或選哪一種處理方法（刪除遺漏值或哪一種插補方法）比較適當均有待討論。前人研究中不乏使用熱卡插補或鄰近插補，但是利用文獻所建議之多重插補的研究仍有限（Yanovitzky, Zanutto, and Hornik 2005）。本研究採用多重插補算是拋磚引玉的示範。

第六、在各種PS調整方法中，本研究僅採用較方便於納入原調查權數的次樣本分組法，證明原調查權數有必要一併考慮於PSA中。不過，若採用其他的調整方法，例如配對，PS倒數等時，是否仍會有相同的結論則需進一步探討。

最後，本研究只選擇兩種普遍使用的傳統加權法產生最後修正的PS權數，並未進行其他加權方法的比較。未來研究仍值得針對其他不同且文獻建議更有效的加權方法（例如：GREG），進行原調查權數納入PSA後之估計成效的比較。

作者簡介

杜素豪，現任中央研究院人文社會科學研究中心暨調查研究專題中心副研究員。研究的主題領域為問卷調查的訪員效應與調查結果的估計。近年的研究分別進行樣本分析、加權與推估，社會距離、訪問風格與訪問情境的訪答效應，以及量表的回答模式誤差。

參考書目

- 內政部戶政司，2007，〈內政統計年報——人口年齡分配、十五歲以上人口教育程度〉。內政部統計處網頁，<http://www.moi.gov.tw/stat/year.aspx>。取用日期：2008年3月10日。
- 行政院研究發展考核委員會，2008，〈九十七年數位落差調查報告〉。行政院研考會網頁，<http://www.rdec.gov.tw/>。取用日期：2009年4月1日。
- 杜素豪、羅婉云、洪永泰，2009，〈以入選機率調整法修正調查推估偏差的成效評估〉。《政治科學論叢》41: 151-176。
- 杜素豪、瞿海源、張苙雲，2012，〈抽樣調查研究法〉。頁202-206，收錄於瞿海源、畢恆達、劉長萱、楊國樞編，《社會及行為科學研究法：總論與量化》。台北：東華書局。
- 李隆安、韓成業，2008，〈The study of weights by the raking method〉。「2008年度中國統計學社學術研討會」論文（12月19日）。台北：中國統計學社。
- 洪永泰，1996，《戶中選樣之研究》。台北：時英。
- ，2006，〈台灣地區電話訪問調查資料加權處理之探討〉。「第六屆調查研究方法與應用國際學術研討會」論文（8月30-31日）。台北：中央研究院人文社會科學研究中心暨調查研究專題中心。
- ，2007，〈台灣地區選前電訪民意調查推估偏差的補救〉。「政治思想、政黨與選舉行為—紀念謝延庚教授學術研討會」論文（5月8日）。台北：台北大學公共行政暨政策學系。
- 洪永泰、黃永政，2000，〈台灣地區電話隨機撥號抽樣方法之研究〉。《選舉研究》7(1): 173-185。

- 洪永泰、洪百薰、林宇璇、呂孟穎、許勝懋、吳淑惠、卓仲彥、徐書儀，2014，〈手機使用對台灣地區電話調查涵蓋率之影響評估〉，*《調查研究》* 31: 7-30。
- 侯佩君，2008，〈電話調查的「類拒訪」分析〉。*《調查研究》* 23: 81-118。
- 俞振華、蔡佳泓，2006，〈如何利用全國性民調推估地方民意？多層次貝氏定理估計模型與分層加權的應用〉。*《台灣政治學刊》* 10(1): 5-38。
- 黃紀、張佑宗，2003，〈樣本代表性檢定與最小差異加權：以2001年台灣選舉與民主化調查為例〉。*《選舉研究》* 10(2): 1-35。
- 張苙雲、廖培珊，2008，〈台灣社會變遷基本調查計畫第五期第四次調查計畫執行報告〉。中央研究院社會學研究所。
- 劉從葦、陳光輝，2005，〈Is Weighting a Routine or Something that Needs to be Justified?〉。*《選舉研究》* 12(2): 149-187。
- Austin, Peter C., 2006, "A Comparison of Propensity Score Methods: A Case-study Estimating the Effectiveness of Post-AMI Statin Use." *Statistics in Medicine* 25: 2084-2106.
- , 2011, "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46(3): 399-424.
- , 2014, "A Comparison of 12 Algorithms for Matching on the Propensity Score." *Statistics in Medicine* 33: 1057-1069.
- Bethlehem, Jelke and Silvia Biffignandi, 2012a, "Weighting Adjustment Techniques." Chapter Ten, Pp. 329-384 in *Web Surveys*. New York: Wiley.

- , 2012b, “Use of Response Propensities.” Chapter Eleven, Pp. 385-418 in *Web Surveys*. New York: Wiley.
- Cochran, William G., 1968, “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies.” *Biometrics* 24(2): 295-313.
- Czajka, John L., Sharon M. Hirabayashi, Roderick J. A. Little and Donald B. Rubin, 1992, “Projecting from Advance Data Using Propensity Modeling: An Application to Income and Tax Statistics.” *Journal of Business & Economic Statistics* 10(2): 117-131.
- Duncan, Kristin B. and Elizabeth A. Stasny, 2001, “Using Propensity Scores to Control Coverage Bias in Telephone Surveys.” *Survey Methodology* 27(2): 121-130.
- Harder, Valerie S., Elizabeth A. Stuart, and James C. Anthony, 2010, “Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research.” *Psychological Methods* 15(3): 234-249.
- Harding, David J., 2003, “Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy.” *American Journal of Sociology* 109(3): 676-719.
- Imai, Kosuke and David A. Van Dyk, 2004, “Causal Inference with General Treatment Regimes: Generalizing the Propensity Score.” *Journal of the American Statistical Association* 99(467): 854-866.
- Judkins, David R., David Morganstein, Paul Zador, Andrea Piesse, Brandon Barrett and Pushpal Mukhopadhyay, 2007, “Variable Selection and Raking in Propensity Scoring.” *Statistics in Medicine* 26: 1022-1033.

- Kalton, Graham and Ismael Flores-Cervantes, 2003, "Weighting Methods." *Journal of Official Statistics* 19(2): 81-97.
- Lee, Sunghee, 2006, "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics* 22(2): 329-349.
- Lee, Sunghee and Richard Valliant, 2008, "Weighting Telephone Samples Using Propensity Scores." Pp. 170-183 in *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski et al.. New York: John Wiley & Sons.
- , 2009, "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment." *Sociological Methods & Research* 37(3): 319-343.
- Little, Roderick J. A. and Donald B. Rubin, 2002, *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley.
- Loosveldt, Geert and Nathalie Sonck, 2008, "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2(2): 93-105.
- Luellen, Jason K., William R. Shadish and M. H. Clark, 2005, "Propensity Scores: An Introduction and Experimental Test." *Evaluation Review* 29(6): 530-558
- Rosenbaum, Paul R., 2002, *Observational Studies*. New York: Springer-Verlag.
- , 2005, "Observational Study." Pp. 1451-1462 in *Volume 3 Encyclopedia of Statistics in Behavioral Science*, edited by Brian S. Everitt and David C. Howell. New York: John Wiley & Sons.

- Rosenbaum, Paul R. and Donald B. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41-55.
- , 1984, "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *JASA* 79(387): 516-524.
- Rubin, Donald. B., 1987, *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Särndal, Carl-Erik, 2007, "The Calibration Approach in Survey Theory and Practice." *Survey Methodology* 33(2): 99-119.
- Schonlau, Matthias, Kinga Zapert, Lisa Payne Simon, Katherine Haynes Sanstad, Sue M. Marcus, John Adams, Hongjun Kan, Rachel Turner and Sandra H. Berry, 2004, "A Comparison between Responses from a Propensity-Weighted Web Survey and Identical RDD Survey." *Social Science Computer Review* 22: 128-138.
- Schonlau, Matthias, Arthur van Soest, and Arie Kapteyn, 2007, "Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?" *Survey Research Methods* 1(3): 155-163.
- Schonlau, Matthias, Arthur van Soest, Arie Kapteyn, and Mick Couper, 2009, "Selection Bias in Web Surveys and the Use of Propensity Scores." *Sociological Methods & Research* 37(3): 291-318.
- Sinharay, Sandip, Hal S. Stern, and Daneil Russell, 2001, "The Use of Multiple Imputation for the Analysis of Missing Data." *Psychological Methods* 6(4): 317-329.
- Srinath, K. P., Martin R. Frankel, David C. Hoaglin, and Michael P. Battaglia,

2009, "Compensating for Noncoverage of Nontelephone Households in Random-Digit-Dialing Surveys: A Comparison of Adjustments Based on Propensity Scores and Interruptions in Telephone Service." *Journal of Official Statistics* 25(1): 77-98.

Stuart, Elizabeth A., 2010, "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25(1): 1-21.

Yanovitzky, Itzhak, Elaine Zanutto and Robert Hornik, 2005, "Estimating Causal Effects of Public Health Education Campaigns Using Propensity Score Methodology." *Evaluation and Program Planning* 28: 209-220.

Zanutto, Elaine L., 2006, "A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data." *Journal of Data Science* 4: 67-91.

Zanutto, Elaine L., Bo Lu, and Robert Hornik, 2005, "Using Propensity Score Subclassification for Multiple Treatment Does to Evaluate a National Antidrug Media Campaign." *Journal of Educational and Behavioral Statistics* 30(1): 59-73.