

台灣 2012 年總統選舉 隨機作答模式調查探討*

謝淑惠** 杜素豪*** 李榮銘**** 王千文*****

摘要

在問卷調查中若有涉及個人隱私或非法行為的敏感性問題時，問卷的設計以直接詢問 (Direct Inquiry) 經常會造成拒答，或即使願意回答也難以確保內容的真實性。因此，為了減少受訪者拒絕回答、逃避回答、或不實回答等現象，可利用隨機作答方式 (Randomized Response Technique, 簡稱 RRT) 蒐集正確問卷資料，降低統計推論的誤差。RRT 可以讓受訪者在不洩露自己的身分且不知道問題的真正意圖情況下，誠實地回答敏感性問題，但目前國內學界鮮少於實地調查中使用 RRT 來蒐集敏感性問題，僅有楊文山 (1994) 於大型面訪調查中應用不相關問題模型的隨機作答方式，估算台灣地區選舉的賄

-
- * 本文初稿發表於 2014 年 12 月 5 日中央研究院社會學研究所舉辦的「國家認同：台灣社會變遷基本調查第二十三次研討會」，承蒙政治大學選舉研究中心游清鑫教授等多位學術先進，對於本文提出的寶貴評論與建議，作者受益良多，特此致謝。
 - ** 謝淑惠為通訊作者，中央研究院人文社會科學研究中心調查研究專題中心助研究員，E-mail: shhsieh@gate.sinica.edu.tw。
 - *** 杜素豪為中央研究院人文社會科學研究中心調查研究專題中心副研究員。
 - **** 李榮銘為逢甲大學統計學系教授。
 - ***** 王千文為中央研究院人文社會科學研究中心調查研究專題中心博士後研究人員。

選比例。本文利用 Hsieh et al. (2013) 在 2012 年至 2013 年參與的「台灣社會變遷基本調查計畫」中所提出的多元類別隨機作答 (Randomized Response Technique in Multi-Category)，蒐集 2012 年總統大選投票行為的資料，估計各候選人的得票率，並與直接詢問法及中央選舉委員會公布的總統候選人之得票率進行有效性比較。

關鍵詞：隨機作答、多元類別隨機作答、敏感性問題、總統選舉

Application of the Randomized Response Technique in the 2012 Presidential Election of Taiwan

Shu-Hui Hsieh^{*}, Su-Hao Tu^{}, Shen-Ming Lee^{***},
Chain-Wen Wang^{****}**

ABSTRACT

Direct inquiry in social surveys of sensitive questions such as personal privacy or illegal behaviors often encounters refusal or untrue response. In order to collect reliable sensitive data, reduce the unwillingness of respondents, and reduce bias in estimation, Randomized Response Technique (RRT) is one of the important methods. RRT has been commonly used to avoid biased answers in surveys on sensitive issues by ensuring respondents' privacy, but are rarely adopted in face-to-face interview surveys in Taiwan, except for Yang (1994) in his

* Assistant Research Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, 128 Academia Rd., Section 2, Nankang, Taipei 115, Taiwan. Email: shhsieh@gate.sinica.edu.tw

** Associate Research Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica. Email: shuo@gate.sinica.edu.tw

*** Professor, Department of Statistics, Feng Chia University, Taiwan. Email: smlee@fcu.edu.tw

**** Postdoctoral Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica. Email: chianwen@gate.sinica.edu.tw

study estimating election bribery by using the unrelated question RRT of Greenberg et al. (1969). We use Randomized Response Technique in Multiple Categories (RRTMC) developed by Hsieh et al. (2013) in the Taiwan Social Change Survey (TSCS) conducted in 2012 and 2013 to collect two waves of 2012 presidential election data to validate RRTMC based by comparing the estimates with the responses to direct inquiry collected in the same survey and official voting records.

Keywords: Randomized response technique, Randomized Response Technique in Multiple Categories, sensitive question, presidential election

一、前言

在 1996 年之前台灣的總統是透過國民代表大會來間接選舉產生，自 1996 年開放公民直選總統開始，透過人民參與的民主體制對於國家未來發展具正面的價值，在經濟事務、政治制度與選舉事務也形成更健全體制。1996 年至 2012 年的五次公民直選總統中，1996 年、2000 年及 2012 年，分別有三位或三位以上候選人同時競爭，使得策略性投票在選舉過程中蘊釀，而有「棄保效應」投票行為的發酵。如 2012 年的總統大選，因牽涉台灣政權的爭奪戰，同時也牽動兩岸與國際關係等議題的激烈攻防，選情變化莫測。在該次選舉中主要的競爭對手為馬英九、蔡英文與宋楚瑜三人，最後國民黨的馬英九與吳敦義獲得 51.60% 得票率，相對於民進黨的蔡英文與蘇嘉全 45.63% 得票率，兩組得票率相差了 6% 約近 80 萬票。然而在實際投票結果與選前最後一次公布民調相比，宋楚瑜（2.77% 得票率）的選票的確是少很多，這

意味著選舉過程中部分選民有棄保效應存在，或尚有其他左右選情的因素等，這也提供選舉研究更寬廣的研究空間。

台灣政治的法定選舉採行了「無記名投票」的方法，目的不在於維護選舉結果的正確，而是在於保障投票人的投票自由。在刑法第 148 條（妨害投票秘密罪）：於無記名之投票，刺探票載之內容者，處三百元以下罰金。所以在台灣目前的選舉生態下，透露個人的投票對象，有時會招致不必要的糾紛或麻煩。而研究者若欲研究此類敏感性議題，以直接詢問（Direct Inquiry，簡稱 DI）方式調查，易因拒訪或無反應率太高的資料產生選舉預測上偏誤的推估。在社會調查研究中常有些問題涉及到個人隱私或非法行為，與社會的一般規範或期許有所衝突，如投票意向、吸毒、墮胎、性傾向等敏感性問題。

為了減少受訪者拒絕回答或逃避回答所產生的誤差，Warner 在 1965 年提出隨機作答（Randomized Response Technique，簡稱 RRT）的調查方法來蒐集敏感性問題的資料，期待藉由 RRT 可以保護受訪者隱私或降低不誠實回答之機率的優點，獲得可靠的資料以推論敏感性問題的比例。此後許多學者也提出其他改良 Warner（1965）模式的 RRT 程序，其最主要的目的皆在於能更準確的估計敏感性問題的效率。如 Greenberg et al.（1969）又將 Warner 模式加以改善，提出了不相關問題（unrelated-question）的 RRT；Franklin（1989）提出假設隨機裝置（random device）為常態分配，利用最大概似估計法來推估二分類母體的比例；Kuk（1990）提出不直接作答的觀念；Mangat and Singh（1990）提出了二階段隨機作答的觀念等，使得受訪者的隱私保護程度提高不少，隨機作答法之精確度也更加提高。另外，Christofides（2003）延伸 Warner（1965）模式提出新的 RRT，讓受訪者不必直接回答「是」與「否」，僅須回答兩數字的差值，來降低不誠實回答的

機率。綜合上述，多位學者所提出之不同的 RRT 調查方式都是爲了消除受訪者心中的疑慮，讓他們可以誠實回答亦可減少拒絕回答或逃避回答的現象。

本文將選舉調查中訪問受訪者投票對象的問題視爲敏感性議題。因 2012 年總統大選中有三組候選人，對其得票率的推估乃屬於社會調查常須估計之多分類母體的比例。Abul-Ela et al. (1967) 推廣 Warner 模型於多類別型態；Chang and Liang (1996) 的隨機作答可處理多屬性 (multiattribute) 的情況，即同一位受訪者能同時具有一個以上的敏感性特質；Greenberg et al. (1971) 修改不相關隨機作答模式，來估計敏感性問題的量化特性。另外，Chen and Singh (2009) 使用 Franklin (1989) 提出的隨機作答模型，並將其推廣至可以估計三項式敏感特性 (A、B、C) 的隨機作答方式。若受訪者具有 A 特性，則依甲隨機裝置作答；若受訪者具有 B 特性，則依乙隨機裝置作答；否則依丙隨機裝置作答。實務上，訪員須準備三個隨機裝置，且其分配皆爲已知。若受訪者具有 A 特性，則從甲隨機裝置抽取一個數字並告訴訪員。而訪員僅知道數字，但無法知道受訪者回答的數字來自於甲、乙、丙哪個隨機裝置，能確實的保護受訪者隱私。然而，此隨機作答的執行條件，受訪者自己應誠實且明確清楚自己所歸屬的敏感性特性，否則將會造成選錯隨機裝置，這種調查操作上的誤差，會使得資料不準確而導致推論的偏誤，也易使受訪者感受到不同隨機裝置可能歸屬爲某敏感性特質，此設計因此在受訪者隱私保護上容易受到質疑。

爲了改進前人隨機作答模型的缺點，並提高多分類 RRT 模型在實務執行上的可行性，同時降低調查操作上的誤差，本文因此引用 Hsieh et al. (2013) 所提出的多元類別隨機作答 (Randomized Response Technique in Multi-Category, 簡稱 RRTMC)。此方法主要延伸自 Christo-

fides (2003) 所提出之二分類 RRT 的概念，僅須使用一個隨機裝置方式，即可降低來自受訪者於隨機裝置操作所產生的變異。主要目的在於利用 2012 年與 2013 年執行的「台灣社會變遷基本調查計畫」（以下簡稱「變遷調查」）中針對 2012 年總統大選投票對象議題所進行的 RRTMC，以與 RRTMC 並列同一份問卷上的 DI 結果，以及中央選舉委員會（以下簡稱中選會）所公告的總統候選人得票率為比較基準，驗證 RRTMC 的有效性。目前在台灣地區的全國性抽樣調查中鮮少使用隨機作答設計，僅楊文山（1994）應用不相關問題模型的隨機作答方式，估計台灣地區選舉的賄選問題。相信本文的研究結果對於文獻中關於敏感性回答品質、RRT 於實地調查的應用，以及 RRT 改進等的相關議題，具有參考價值。

二、文獻探討

隨著台灣的民主化，幾乎年年舉辦的大小型選舉活動，不只是民眾日常生活的一部分，投票更成為政治參與中最重要的一環。由於近十年來國內政治生態急遽變化，面對越來越大的時代衝擊、選舉壓力與選情瞬息萬變，影響選舉的因素越來越繁雜，但在掌握選民投票意向的選舉研究上則提供更寬廣的研究空間。選舉預測是選舉研究者極感興趣的議題之一，過去台灣對於選舉預測的研究大多集中在以民意調查所得到資料作為分析主體，並透過個體資料或整體資料的呈現，在選舉前予以評估預測，或在事後尋求更具解釋力的模型，不僅在學術領域中有助於了解選民的投票行為，在實務上也可以幫助競選策略擬定及競爭態勢研判。在台灣受到選罷法規定選前十天禁止發布民調結果，使得民意受到短期競選事件影響後，採用發布之民調結果進行

選舉預測工作具有局限性。

在台灣地區的選舉調查，不論是電話訪問、面對面訪問或網路訪問等調查方式，已實行十多年且累積了相當多的研究資料。選前的民意調查會在無意間刺激受訪者對於選舉產生興趣或促進其積極參與選舉活動和投票，而選後的民意調查則容易引起受訪者在接受訪問時產生當時社會所讚賞或期許的回答行爲。因此，不論是在選舉前的預測或者選舉後的結果估計上，我們往往無法蒐集到投票行爲有效與可信的資料。

此外，在問卷調查中若受訪者認爲投票對象、投票行爲等議題是涉及個人隱私的敏感性問題時，受訪者往往會有拒絕回答或不誠實回答的傾向，造成推論之投票率或得票率有誤差的現象。爲了解決此一問題，Warner（1965）提出 RRT，其構想在於透過某種隨機裝置（如數字卡、擲骰子）的設計，受訪者可以隨機選答敏感性問題或該敏感性問題的反向問題。透過隨機裝置，受訪者須利用研究者設計好的隨機裝置來選擇回答題目 A 或題目 B，其選擇的機率分別爲 p 和 $1-p$ ：

題目 A：我具有 A 特性嗎？

題目 B：我具有 A^c 特性嗎？

實際執行時，受訪者僅須告訴訪員他的回答爲「是」或「否」，而訪員無法知道受訪者是回答題目 A 或 B，藉此設計希望受訪者了解其所提供的答案並不會在不認識的人面前曝光，因而願意提供正確的回答。假設具有敏感性 A 特性的比例爲 θ ，其餘具 A^c 者的比例爲 $1-\theta$ 。在受訪者皆誠實作答的情形下，受訪者回答「是」的機率爲：

$$\pi = \theta p + (1 - \theta)(1 - p)$$

假設 n 為樣本總數， n_1 為回答「是」的人數，則 π 的不偏估計量為 $\frac{n_1}{n}$ 。所以 θ 的不偏估計量與變異數為：

$$\hat{\theta}_w = \frac{1}{2p-1} \left(p - 1 + \frac{n_1}{n} \right)$$

$$\text{Var}(\hat{\theta}_w) = \frac{\theta(1-\theta)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

此處 $p \neq \frac{1}{2}$ 。由於研究者可以先行設計隨機裝置的機率 p ，然而 p 的大小會決定變異數的大小，也會影響到受訪者與訪員之間的合作程度。當 $p=1$ 或 0 時變異數會最小， $\hat{\theta}_w$ 會得到較高的效率。但這樣的設計就是所有的受訪者應回答題目 A ($p=1$)，或所有的受訪者應回答題目 B ($p=0$)，就相當於 DI 法，則受訪者誠實回答率應該不高。

為了改善 Warner (1965) 模式中題目 A 和題目 B 皆為敏感性問題的疑慮，Greenberg et al. (1969) 提出了不相關問題的隨機作答模型，令 Y 與 A 不相關的特性，期望藉由具有 Y 特性的題目 C 來增加受訪者的信任感，以降低不誠實回答的比例。Greenberg et al. (1969) 的執行步驟與 Warner (1965) 相同，透過隨機裝置來選擇題目 A 與題目 C，其機率分別為 p_1 和 $1-p_1$ ：

題目 A：我具有 A 特性嗎？

題目 C：我具有 Y 特性嗎？

在此，假設 γ 為題目 C 回答是（具有 Y 特性）的比例，則受訪者皆誠實作答的情形下，受訪者回答「是」的機率為：

$$\pi = \theta p_1 + \gamma(1-p_1)$$

若假設 γ 已知時，以 θ 的不偏估計量與變異數為：

$$\hat{\theta}_g = \frac{\frac{n_1}{n} - (1-p_1)\gamma}{p_1}$$

$$\text{Var}(\hat{\theta}_g) = \frac{n(1-\pi)}{np_1^2}$$

但若 γ 未知時，我們需要兩個獨立的樣本才能來估計 θ 和 γ 。

Abul-Ela 等人（1967）延伸 Warner 模型的兩類回答選項（是與否）為多類別型態。假設敏感特性之回答選項可分為 m 個互斥的類別 A_1, A_2, \dots, A_m ，則其所對應參數為 $\theta_1, \theta_2, \dots, \theta_m$ 。其原理是選取 $s=m-1$ 樣本，每一組內的樣本數為 n_1, n_2, \dots, n_s ，因此須準備 s 組隨機裝置，其對應的機率為 $p_{ij}(1 \leq i \leq s, 1 \leq j \leq m)$ ，則 $\sum_{j=1}^m p_{ij} = 1$ 。因此，第 i 組受訪者被要求回答第 i 個隨機裝置中所提及之特性，則回答「是」的機率為：

$$\pi_i = \sum_{j=1}^m p_{ij} \theta_j$$

在多類別設計中，若問題的回答類別有 m 個，則需 $m-1$ 組獨立樣本，實務上將付出較高的調查成本。為解決此缺點，Bourke and Dalenius（1976）提出僅須選取一組樣本的 RRT 模型，每一受訪者據實陳述其類別的機率為 p ，而按指示回答 j 類別的機率為 $p_j(j=1, 2, \dots, m)$ 其中 $p + \sum_{j=1}^m p_j = 1$ 。因此，回答類別 j 的機率為：

$$\phi_j = p\pi_j + p_j\pi_1 + \dots + p_j\pi_j + \dots + p_j\pi_m$$

此處 $j=1, 2, \dots, m$ ，亦利用矩陣型式則可得到 $\theta_1, \theta_2, \dots, \theta_m$ 的不偏估

計量與變異矩陣。

之後，許多學者爲了讓受訪者誠實作答，一直不斷的研究如何獲得較準確的估計值，如 Kuk (1990)、Christofides (2003)、Kim and Warde (2004) 等。其中 Christofides (2003) 提出讓受訪者不必直接回答「是」和「否」，僅須回答是兩個數字的差值，藉以降低不誠實回答的機率。假設具有敏感特性 A 的比例爲 θ ，Christofides (2003) 提供一個隨機裝置給每一位受訪者使用，其中隨機裝置爲整數 $1, 2, \dots, L$ ，且其對應的機率爲 P_1, P_2, \dots, P_L ，則定義每一個受訪者使用隨機裝置之後得到的數字爲 Y 。若受訪者具有 A 特性，則回答 Y 到 $L+1$ 的距離；反之，則回答 Y 到 0 的距離。因此，我們可以定義若受訪者具有特性 A，則 $X=L+1$ ，反之則 $X=0$ ，所以 $P(X=L+1)=\theta$ 與 $P(X=0)=1-\theta$ 。

實際執行時，訪員僅知道受訪者回答兩數差的距離 $D=|X-Y|$ ，完全不會知道受訪者的 X 和 Y 分別爲數字多少，藉此可提高受訪者願意誠實回答的比例。因此，每第 i 位受訪者回答數字爲 k 的機率爲：

$$P(D_i=k)=(1-\theta)P_k+\theta P_{L+1-k}$$

此處 $i=1, 2, \dots, n$ 和 $k=1, 2, \dots, L$ 。令 \bar{D} 爲 D_1, D_2, \dots, D_n 的樣本平均數，因此可得 θ 的估計式爲：

$$\hat{\theta}_c = \frac{\bar{D} - E(Y)}{L+1 - 2E(Y)}, \quad L+1 - 2E(Y) \neq 0$$

則 $\hat{\theta}_c$ 的變異數爲：

$$\text{Var}(\hat{\theta}_c) = \frac{\theta(1-\theta)}{n} + \frac{\text{Var}(y)}{n[L+1-2E(Y)]^2}$$

綜合以上，應用隨機作答法於敏感性問題的調查，可以降低非抽

樣誤差。然而，在調查設計、訪問費等方面的成本較一般調查為高。除了調查成本外，隨機作答設計方法應用在社會科學大型面訪調查時，在操作程序的可行性與受訪者的隱私保護措施等相關問題上也需嚴密的考量與設計。

三、研究設計

民意調查在台灣的政治選舉文化中占有很重要的一環，但如何蒐集準確性高的資料則是一件相當困難的工作。為了使龐大的樣本在不同年齡或教育程度皆能順利完成，問卷設計內容的完整性與複雜性就成為設計問卷相當重要的考慮因素。一般而言，以 2012 年總統選舉為例，最常見的問卷設計如下：

題目 A：2012 年一月的總統選舉，請問您有沒有去投票？

- (01) 有，投給馬英九 (02) 有，投給蔡英文
 (03) 有，投給宋楚瑜 (04) 有，投廢票
 (05) 沒有去投票 (06) 年滿 20 歲但沒有總統投票權
 (07) 未滿 20 歲

在此，若以題目 A 的設計方式直接詢問獲得投票對象的資料，我們稱之為 DI。本文以 2012 年總統選舉為例，引用可以更簡化又準確估計，由 Hsieh 等人（2013）所提出的 RRTMC 與執行流程。接著，說明分析資料來源。

（一）多元類別隨機作答（簡稱 RRTMC）

為了提高實務執行上的可行性，確實保護受訪者的隱私，Hsieh et

al. (2013) 所提出的 RRTMC 主要是延伸 Christofides (2003) 二類別的隨機作答設計概念，並延伸為多元類別的作答方式，其所設計的題目如下：

題目 B1：請問在 2012 年一月份總統選舉時，您有沒有去投票？

(01)有去投票 (02)沒有去投票 (03)沒有投票權 (04)投廢票 (98)拒答

題目 B2：請問您投票時所選的總統候選人是數字 7、6 或 0，但請不要告訴我您選的數字，只需要告訴我這數字跟您現在所抽到的數字差多少就好了。

(由於您回答的是『數字的差』，我們完全不會知道您真正是投票給哪一組總統候選人，所以請您放心回答。)

數字 7：宋楚瑜

數字 6：蔡英文

數字 0：馬英九

那麼，請問您抽到的數字跟提示卡上所選的數字，大減小後差多少？

在 RRTMC 的調查執行程序中，詢問所有的受訪者題目 B1，若受訪者回答有去投票(01)，則再進一步回答題目 B2。在回答題目 B2 之前，受訪者需由隨機裝置中抽取 1 張數字卡（其中，數字為 1 者有 8 張；4 張數字 2；8 張數字 3；16 張數字 4，與 4 張數字 5，共 40 張），並假設受訪者所抽中的數字卡為變數 Y 。接下來，設變數 X 為受訪者回想 2012 年投票時每位總統候選人所代表的數字，則分別為 $X=0$ 代表馬英九， $X=6$ 代表蔡英文， $X=7$ 代表宋楚瑜，因此， $P(X=0)=\theta_1$ ， $P(X=6)=\theta_2$ ， $P(X=7)=\theta_3$ ，且 $\theta_1+\theta_2+\theta_3=1$ 。由於受訪者不須讓訪員知道自己抽中的數字為 Y 與投票對象所代表數字 X ，僅須回答 X 和 Y 兩

數字的差值 $D=|X-Y|$ ，則 D 的範圍為 1 至 6。舉例說明，若受訪者投票對象為馬英九則 $X=0$ ，抽中數字 $Y=3$ ，則其機率 $P(X=0, Y=3)=\theta_1 P_3$ 。因此，在 RRTMC 的設計下，當訪員可得受訪者回答 $D=3$ ，其中可能來自於 $(X=0, Y=3)$ 、 $(X=6, Y=3)$ 和 $(X=7, Y=4)$ ，則其機率為：

$$P(D=3)=P(X=0, Y=3)+P(X=6, Y=3)+P(X=7, Y=4)=(\theta_1+\theta_2)P_3+\theta_3P_4$$

接著，由表 1 隨機裝置數字卡 Y 與受訪者回答數字 D 的聯合機率分配，亦可清楚得到 $P(D=k)$ ， $k=1, 2, \dots, 6$ 的機率。而在參數 θ_1 、 θ_2 、 θ_3 估計方面，Hsieh et al. (2013) 提出最大概似法來得到不偏估計量及變異數。

整體而言，RRTMC 僅能得到受訪者回答數字 D 的資料，因此，只能做「整體意見的推估，不會知道個人的意見」，藉此希望可以消弭受訪者害怕自己身分隱私被曝光的疑慮，以避免受訪者的無反應及提供錯誤的回答等。

表 1 Y 與 D 的聯合機率分配

D \ Y	1	2	3	4	5	$P(D=k)$
1	$\theta_1 P_1$				$\theta_2 P_5$	$\theta_1 P_1 + \theta_2 P_5$
2		$\theta_1 P_2$		$\theta_2 P_4$	$\theta_3 P_5$	$\theta_1 P_2 + \theta_2 P_4 + \theta_3 P_5$
3			$(\theta_1 + \theta_2) P_3$	$\theta_3 P_4$		$(\theta_1 + \theta_2) P_3 + \theta_3 P_4$
4		$\theta_2 P_2$	$\theta_3 P_3$	$\theta_1 P_4$		$\theta_2 P_2 + \theta_3 P_3 + \theta_1 P_4$
5	$\theta_2 P_1$	$\theta_3 P_2$			$\theta_1 P_5$	$\theta_2 P_1 + \theta_3 P_2 + \theta_1 P_5$
6	$\theta_3 P_1$					$\theta_3 P_1$

(二) 分析資料

本研究所使用的資料來自 2012 年和 2013 年由中央研究院社會學研究所執行的「變遷調查」(章英華等人 2012)。此調查以台灣地區(不含福建省金門縣和連江縣)具有本國國籍且設有戶籍,年齡在 18 歲(含)以上的民眾為抽樣母體,以戶籍資料為抽樣清冊(sampling frame)。樣本的選取採分層三階段等比例機率抽樣,各層內採用抽取率與單位大小成比例(Probability Proportional to Size Sampling, PPS)等距抽樣法,逐步抽取「鄉鎮市區」、「村里」、「人」。實際調查訪問時,並不包括軍事單位、醫院、療養院、學校、職訓中心、宿舍、監獄等機構內之居民及通緝犯;採用電腦輔助面對面訪問方法。

本文利用「變遷調查」2012 年的社會階層組和 2013 年的國家認同組中針對 2012 年總統選舉投票對象議題,於同一份問卷中設計 DI 的題目 A 與 RRTMC 的題目 B2 (見附錄)。因考量到先問 RRTMC 的題目 B2 後,受訪者可能會覺得 DI 的題目 A 敏感性過高,或是 RRTMC 的題目 B2 操作執执行程序複雜,因而造成拒絕回答或逃避回答 DI 的題目 A 比例過高,形成資料可靠性降低的問題。所以 DI 的題目 A 置於前,並間隔一些題目後,於問卷最後設置 RRTMC 的題目 B2。

在題目設計方法上,2012 年與 2013 年的 RRTMC 的篩選題目 B1 有些不同。2013 年的問卷中並沒有篩選題目 B1,僅利用 DI 的題目 A 作為篩選題,若受訪者有確切回答出投票對象、有去投票但不願回答或忘記投給誰、投廢票者,皆歸類成有去投票者,亦須回答題目 B2。但是為了整體一致性的比較分析,最後資料分析時決定將 2012 年的資料也調整為與 2013 年相同,皆以 DI 的題目 A 作為篩選進入 RRTMC 的題目 B2 的依據。

四、研究結果

本文利用 2012 年及 2013 年的「變遷調查」資料，希望了解選民真正的投票意向，以及受訪者回答是否有社會讚許或期許的傾向。

（一）調查方法對回答投票對象意願是否有影響

表 2 列出 2012 年和 2013 年「變遷調查」實際完成的樣本數，分別為 2134 和 1952，因本研究所需，經過篩選後僅保留 20 歲以上且具總統投票權的受訪者，分別為 2044 和 1839。針對有關投票對象的問題，資料重新整理為一個新的變數。如 DI 的題目 A，若受訪者確切回答出投票對象則歸為「願意回答」，反之為「不願意回答」；而 RRTMC 的題目 B2，若受訪者確切回答出兩數值差值則歸為「願意回答」，反之為「不願意回答」。因此，我們就可以比較同一年度不同調查方法 DI 和 RRTMC 的願意回答投票對象比例之間是否存在差異。

表 2 願意回答投票對象比例 DI 和 RRTMC 的比較

調查年度和主題	2012 年社會階層		2013 年國家認同	
完成樣本數	2134		1952	
篩選後樣本數	2044		1839	
調查方法	DI	RRTMC	DI	RRTMC
願意回答投票對象	1335 (65.31%)	1499 (73.34%)	1283 (69.77%)	1366 (74.23%)

註：1. DI 為直接詢問法。

2. RRTMC 為 Hsieh et al. (2013) 提出的多元類別隨機作答模式。

在表 2 中，利用 DI 所蒐集資料而估算之願意回答投票對象比例，2012 年和 2013 年分別為 65.31% 和 69.77%。在 RRTMC 中，所估算願意回答投票對象比例為 73.34% 和 74.23%，應該是有去投票且願意回答的比例，與中選會公布的投票率 74.3% 相當接近。就 2012 年的比較，受訪者願意回答投票在 RRTMC 相對於 DI 高估 8.03%，而於 2013 年則高估 4.46%，呈現 RRTMC 在隱私保護上比 DI 還要理想，所以受訪者願意回答的比例較高。

（二）調查方法對得票率估算是否有影響

爲了了解有多少受訪者對 RRTMC 有操作上問題，或多少受訪者願意直接回答投票對象等，本研究設計了兩類回答選項，用以區分成受訪者自行計算和訪員計算兩類，其中，只要受訪者告訴訪員兩數之任一個或以上歸類爲訪員計算。如表 3，就受訪者計算部分，先利用卡方檢定值爲 1.18，且 P 值爲 0.95，表示 2012 年和 2013 年資料於各類別選項比例差異不顯著。在 2012 年受訪者計算的比例約 80.39%，與其相比較 2013 年則減少約 4.91%，也相對呈現訪員計算會多出此部分。

在得票率方面，表 4 爲 DI 和 RRTMC 於 2012 年總統選舉中各投票對象得票率的分析結果。就 DI 而言，馬英九的得票率在 2012 年和 2013 年分別爲 56.40% 和 54.79%，相對於中選會公布結果高估 4.80% 和 3.19%。呈現在讚賞或期許的導向下，選後調查時受訪者是有可能會偏向於回答已獲得勝選的總統候選人，因而有可能造成該候選人的得票率高估的現象。相對的，已落敗的候選人，極有可能在得票率的估計上會較爲低落，如表 4 蔡英文與宋楚瑜兩位候選人得票率與中選會公布結果相比，分別爲平均低估約 4.48%，與平均高估 0.48%。

表 3 RRTMC 的資料次數分配

	2012 年社會階層		2013 年國家認同	
	受訪者計算	訪員計算	受訪者計算	訪員計算
兩數差 1	182	40	154	57
兩數差 2	270	70	227	75
兩數差 3	284	73	228	72
兩數差 4	277	70	246	78
兩數差 5	169	38	155	50
兩數差 6	23	3	21	3
合 計	1205	294	1031	335
百分比	80.39	19.61	75.48	24.52
卡方檢定	X ² =1.18, P-value=0.95			

表 4 得票率 DI 和 RRTMC 的比較

總統候選人	中選會 公布結果	2012 年社會階層		2013 年國家認同	
		DI	RRTMC	DI	RRTMC
1. 馬英九	51.60	56.40 (1.40)	46.31 (2.98)	54.79 (1.36)	48.03 (3.08)
2. 蔡英文	45.63	40.22 (1.39)	44.31 (2.94)	42.09 (1.35)	43.48 (3.04)
3. 宋楚瑜	2.77	3.37 (1.25)	9.38 (1.69)	3.12 (1.27)	8.50 (1.56)
願意回答投票 對象的樣本總數		1335	1499	1283	1366

註：表格中數字為百分比，括弧中數字為估計標準差。

在 RRTMC 蔡英文的得票率在 2012 年和 2013 年分別為 44.31% 和 43.48%，與中選會公布結果相比低估 1.32% 和 2.15%，相較於 DI 方法精確性高。而馬英九得票率分別為 46.31% 和 48.03%，與中選會公布結果相比分別低估 5.29% 和 3.57%，其中差距比例與 DI 差不多，但 RRTMC 皆呈現低估的現象。然而，蔡英文或馬英九在 2012 年，與其相對於 2013 年 RRTMC 方法的得票率分別減少 0.83% 和增加 1.72%，或許是因調查年度與實際投票選舉日期距離較遠，受訪者回答比較有社會讚許或期許從眾效果，所以馬英九推估得票率 2013 年會較 2012 年高。但在宋楚瑜得票率分別為 9.38% 和 8.50%，都相較於中選會公布結果高估許多，反而 DI 和中選會公布結果較接近。

（三）不同受訪者特質在回答投票對象意願的差異

由於選民的投票行為本來就是一種心理狀態的陳述，在調查時只能依受訪者當時自我的表達來判斷。本研究的設計是將 DI 的題目 A 與 RRTMC 的題目 B2 並列同一份問卷，所以我們可依受訪者回答 DI 與 RRTMC 的情況，來區分成問卷設計上的「隱密性」和「複雜性」兩種效果。其中「隱密性」效果意指受訪者不願意回答在 DI 的題目 A，但願意回答在 RRTMC 的題目 B2，如同表 5 所示，RRTMC 中願意回答投票對象的比例於 2012 年和 2013 分別為 176 (8.61%) 和 106 (5.76%)。「複雜性」效果方面，則為受訪者願意回答在 DI 題目 A 的投票對象，但不願意回答在 RRTMC 的題目 B2，占有願意回答投票對象的比例於 2012 年和 2013 年分別為 12 (0.59%) 和 23 (1.25%)。整體而言，RRTMC 的隱密性設計效果較佳，所得到的願意回答投票對象的比例相較於 DI 高，主要來自於 DI 有去投票者中，有去投票但不願回答或忘記投給誰、投廢票的受訪者，在使用 RRTMC 設計時比

表 5 DI 與 RRTMC 問卷設計效果的比較

調查年度和主題	2012 年社會階層		2013 年國家認同	
調查方法	DI	RRTMC	DI	RRTMC
隱密性僅回答 RRTMC	—	176 (8.61)	—	106 (5.76)
複雜性僅回答 DI	12 (0.59)	—	23 (1.25)	—
DI 和 RRTMC 皆有回答	1323		1260	
總樣本數	1335	1499	1283	1366

註：表格中數字為人數，括號中數字為百分比。

較願意回答投票對象。

進一步，在隱密性設計效果中探討 RRTMC 對哪些人具有效用，能提高其回答率。針對影響受訪者回答意願的主要屬性，我們檢視受訪者的性別、年齡、教育程度、政黨支持與主觀社經地位等五個變數在回答上的差異，結果如表 6。就性別而言，以 2013 年有去投票且願意回答投票對象者的女性為最高（55.66%，相對於男性的 44.34%）。整體而言，女性願意回答投票對象的比例較男性高 11.32%。在年齡方面，2012 年資料顯示 40 歲到 59 歲的選民，最不願意透露投票對象（45.45%），其次則為 20 歲到 39 歲的 28.98%。在教育程度方面，兩年資料均呈現教育程度高中以下的受訪者，越傾向告訴訪員他們的投票對象，於 2012 年和 2013 年分別為 63.07% 和 69.81%。在政黨支持上，都沒有／都支持的選民比較傾向告訴訪員他們的投票對象，其中 2012 年的比例為 87.25%，相較於 2013 年（81.11%）高出約 6.14%。最後，主觀社經地位方面，呈現主觀社經地位高的人最不願意回答投

表 6 分析隱密性效果的受訪者特質

調查年度和主題		2012 年社會階層	2013 年國家認同
性 別	男性	82 (46.59)	47 (44.34)
	女性	94 (53.41)	59 (55.66)
年 齡	20-39	51 (28.98)	33 (31.13)
	40-59	80 (45.45)	39 (36.79)
	60 以上	45 (25.57)	34 (32.08)
教育程度	高中以下	111 (63.07)	74 (69.81)
	專科以上	65 (36.93)	32 (30.19)
政黨支持	泛藍	9 (6.04)	6 (6.67)
	泛綠	10 (6.71)	11 (12.22)
	都沒有／都支持	130 (87.25)	73 (81.11)
主觀社經地位	低	60 (35.09)	30 (29.70)
	中	98 (57.31)	59 (58.42)
	高	13 (7.60)	12 (11.88)

註：1. 若資料有缺失值採刪除後再進行計算。

2. 括號內數字為百分比。

票對象，以 2012 年而言只占 7.60%，其次為主觀社經地位低。整體而言，女性、年齡 40 歲到 59 歲、教育程度高中以下、政黨支持都沒有／都支持、主觀社經地位中者，若應用 RRTMC 法調查時，會提高願意回答投票對象的比例。

五、結論與討論

在「變遷調查」中，我們於 2012 年的社會階層組與 2013 年國家認同組加入調查 2012 年總統選舉中受訪者的投票對象議題的隨機作答方法研究。我們將「直接詢問法」(DI) 的題目與「多元類別隨機作答」(RRTMC) 的題目並列同一份問卷，先以 DI 的題目詢問受訪者，在問卷最後才採 RRTMC 的題目，希望能提高受訪者兩種方法都回答的比例，以進行 DI 與 RRTMC 的差異性與有效性的比較。就訪員訓練方面，爲了提高受訪者願意接受訪問的機率，除了培訓訪員如何說服受訪者的技巧外，並告知本研究僅能做「整體意見的推估，不會知道個人的意見」，在保護資料提供者的隱私權之下，提供研究者較爲可靠之資料，此乃爲隨機作答法主要用途之一。

本研究發現 RRTMC 對於願意回答 2012 年總統選舉投票對象的比例估算結果，於調查年度 2012 年與 2013 年分別爲 73.34% 和 74.23%，或許民眾覺得在隱私保護上比 DI 更受到尊重，所以願意回答投票對象的比例較 DI 高。在投票對象得票率的估算方面，蔡英文於 2012 年和 2013 年調查估算的得票率分別爲 44.31% 和 43.48%，與中選會公布結果低估 1.32% 和 2.15%，且此差距都在一倍標準差以內，此結果與 DI 相比較都更接近實際得票率 45.63%。而馬英九估計的得票率與中選會公布結果分別低估 5.29% 和 3.57%，而此差距大於一倍標準差，反而 DI 估計的得票率與中選會公布結果分別高估 4.8% 和 3.19%，差距低於 RRTMC。由 DI 的結果可以清楚看出選後調查會有社會讚許或期許的情況，但在 RRTMC 下則看不到。最後，RRTMC 所估算宋楚瑜得票率分別爲 9.38% 和 8.50%，與實際得票率 2.77% 差距頗大，卻

與多數媒體於選前最後一次公布的民意調查介於 7% 至 10% 的方向相同，反而 DI 法的估計結果與中選會較接近。

我們也進一步探討 RRTMC 對哪些人具有提高其回答率的效果。針對影響受訪者回答意願的主要屬性，我們檢視受訪者的性別、年齡、教育程度、政黨支持與主觀社經地位等。最後發現，女性、年齡 40 歲到 59 歲、教育程度高中以下、政黨支持都沒有／都支持、主觀社經地位中者，若應用 RRTMC 來調查時，願意回答投票對象的比例較高。整體而言，在 RRTMC 下，願意回答投票對象的比例應該是有去投票且願意回答者的比例，而其估算的比例與中選會公布的投票率相當接近。但由於台灣目前沒有其他相關文獻可以參考，需要未來多做一些實證分析才能有更清楚的證據來說明。

綜合上述，在選後調查中利用 RRTMC 所估算各候選人的得票率，僅蔡英文得票率比較準確，低估了馬英九得票率，卻高估了宋楚瑜得票率。然而，如文獻假設及前人研究的發現，隨機作答法比較能測量或反映民眾內心的看法或傾向等，但或許政治民主化後，認為調查投票對象的問題屬於高度敏感性問題的受訪者人數減少或不誠實回答太高，而使得此 RRTMC 對於得票率的估算結果與預期效果不同。由於上述結論仍缺乏相關變項可以檢驗，有待未來進行更深入的研究設計來處理此問題。

最後，本文仍有不足與局限之處。本研究 RRTMC 所估算的得票率與真實得票率有差距，卻與多家媒體在選前禁止公布前最後一次民調的結果相同（見附錄），這是一個值得深入探討的問題。若想利用 RRTMC 了解是否有策略性投票的選舉效應，從單一時間點的橫斷面調查資料亦有其局限，理想上若能納入選前民調調查資料進行跨時比較分析，或許更能看出趨勢中的效應，才能用以解釋其投票選擇。然

而，由於未能針對相同樣本進行追蹤調查，無從得知這些選民在大選中的實際投票選擇，而形成驗證此調查方法可行性的主要困境。不過，本文以選後調查所得之研究發現仍可提供後續總統大選進行調查研究時，思考與規劃可行性的選前樣本並進行選後追蹤調查，以建構比較完整跨時的調查方法研究。另外，本文是在「變遷調查」中利用 Hsieh et al. (2013) 提出的多元類別隨機作答模型，是僅次於楊文山 (1994) 應用不相關問題模型的隨機作答方法調查選舉賄選問題後，以台灣樣本在面對面的訪問中利用隨機作答方法進行調查，這在研究方法、訪問技術或問卷設計上都是未來可以再努力與改進的研究方向。

附錄 1 問卷

2012 年社會階層

A. 直接詢問 (DI) 題目

96. 今年一月的總統選舉，請問您有沒有去投票？
- (01)有，投給馬英九 (02)有，投給蔡英文
(03)有，投給宋楚瑜 (04)有，投廢票 (05)沒有去投票
(06)年滿 20 歲但沒有總統投票權 (07)未滿 20 歲

B. RRTMC 題組

112. 請問在今年一月份總統選舉時，您有沒有去投票？
- (01)有去投票 (02)沒有去投票 (03)沒有投票權
(04)投廢票 (98)拒答
113. 請問您投票時所選的總統候選人姓名前是 7、6 或 0 的哪個數字？
(提示卡 38) 請不要說出您所選的數字，記在心裡就好。

數字 7：宋楚瑜

數字 6：蔡英文

數字 0：馬英九

那麼，請您抽一張數字卡，然後告訴我數字卡上的數字跟提示卡上所選的數字，大減小後差多少？

受訪者自己算的：

- (01)01 (02)02 (03)03 (04)04
 (05)05 (06)06

訪員幫忙算的：

- (07)01 (08)02 (09)03 (10)04
 (11)05 (12)06

2013 年國家認同

A. 直接詢問 (DI) 題目

92. 去年一月的總統選舉，請問您有沒有去投票？投給誰？

- (01)有，馬英九 (02)有，蔡英文 (03)有，宋楚瑜
 (04)有，投廢票 (05)有，但不願意回答或忘記投給誰
 (06)有，但拒領總統選舉票 (07)沒有去投票
 (08)當時年滿 20 歲但沒有總統投票權 (09)當時未滿 20 歲

B. RRTMC 題組

112. 請問在去年一月的總統選舉，您投票時所選的總統候選人是數字 7、6 或 0？但請不要告訴我您選的數字，只需告訴我這數字跟您現在所抽到的數字差多少就好了。(由於您回答的是『數字的差』，我們完全不會知道您真正是投票給哪一組總統候選人，所以請您放心回答。)

數字 7：宋楚瑜

數字 6：蔡英文

數字 0：馬英九

那麼，請問您提示卡與抽到的數字，大減小後差多少？

受訪者自己算的：

- (01)01 (02)02 (03)03 (04)04
 (05)05 (06)06

訪員幫忙算的：

(07)01 (08)02 (09)03 (10)04

(11)05 (12)06

附錄 2

2012 年總統大選前各調查機構最後一次民調的結果

	馬英九		蔡英文		宋楚瑜		未表態
	原始支持度	預測得票率	原始支持度	預測得票率	原始支持度	預測得票率	原始支持度
TVBS (1/2)	45	51.1	37	42.0	6	6.8	12
聯合報 (1/3)	44	50.6	36	41.4	7	8.0	13
蘋果日報 (1/2)	42.2	50.2	35.7	42.4	6.2	7.4	15.9
中國時報 (1/3)	39.5	48.3	36.5	44.6	5.8	7.1	18.2
台灣智庫 (1/1)	38.8	43.9	37.8	42.8	11.8	13.3	11.6
未來事件交易所 (1/2)		42.6		49.8		10.7	

資料來源：<http://blackrain.skycity.cc/2012/01/03/11448.html>

註：括弧內數字為日期。

參考文獻

- 中央選舉委員會，2012，歷屆公職選舉資料。選舉資料庫網站 (<http://db.cec.gov.tw/histMain.jsp?voteSel=20120101A1>，取用日期：2014 年 11 月 5 日)。
- 章英華、杜素豪、廖培珊，2012，《台灣社會變遷基本調查計畫第六期第二次調查計畫執行報告》。台北：中央研究院社會學研究所。
- ，2013，《台灣社會變遷基本調查計畫第六期第三次調查計畫執行報告》。台北：中央研究院社會學研究所。
- 楊文山，1994，〈隨機問答在大型實地調查中的應用：以估計台灣地區賄選為例〉。頁 363-382，收錄於伊慶春主編，《台灣社會的民眾意向：社會科學的分析》。台北：中央研究院中山人文社會科學研究所。
- Abul-Ela, Abdel-Latif A., Bernard G. Greenberg, and Horvitz, Daniel G., 1967, "A multi-Proportion Randomized Response Model." *Journal of the American Statistical Association* 62: 990-1008.
- Bourke, P. D. and T. Dalenius, 1976, "Some New Ideas in the Realm of Randomized Inquiries." *International Statistical Review* 44(2): 219-221.
- Chang, Horng Jinh and Der Hsin Liang, 1996, "A Two-Stage Unrelated Randomized Response Procedure." *Australian Journal of Statistics* 38(1): 43-51.
- Chen, Cheng C. and Sarjinder Singh, 2009, "The Franklin's Randomized Response Model for Two Sensitive Attributes." *Section on Survey Research methods-JSM 2009*: 4171-4185.
- Christofides, Tasos C., 2003, "A Generalized Randomized Response Technique." *Metrika* 57(2): 195-200.
- Franklin, Le Roy A., 1989, "A Comparison of Estimators for Randomized Response Sampling with Continuous Distributions from a Dichotomous Population." *Communication in Statistics Theory and Methods* 18(2): 489-505.
- Greenberg, Bernard G., Abdel-Latif A. Abul-Ela, Walt R. Simmons and Daniel G. Horvitz, 1969, "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of American Statistical Association* 64: 520-539.
- Greenberg, Bernard G., Roy R. Kuebler, Jr., James R. Abernathy and Daniel G. Horvitz, 1971, "Application of the Randomized Response Technique in Obtaining Quantitative Data." *Journal of American Statistical Association* 66: 243-250.
- Hsieh, Shu-Hui, Shen-Ming Lee and Su-Hao Tu, 2013, "A New Randomized Response Model for Multiple Sensitive Attributes." 論文發表於統計學會主辦「統計學年會」。高雄：

高雄大學，2013 年 6 月 28-29 日。

Kim, Jong-Min and William D. Warde, 2004, "A Stratified Warner's Randomized Response Model." *Journal of Statistical Planning Inference* 120: 155-165.

Kuk, Anthony Y. C., 1990, "Asking Sensitive Questions Indirectly." *Biometrika* 77(2): 436-438.

Mangat, N. S. and Ravindra Singh, 1990, "An Alternative Randomized Response Procedure." *Biometrika* 77(2): 439-442.

Warner, Stanley L., 1965, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60: 63-69.