

## 研究論文

# 追蹤調查中 問卷回答品質的混合模式效應\*

杜素豪\*\* 于若蓉\*\*\* 洪琮瑋\*\*\*\*

### 摘要

為了提高母體涵蓋率與問卷回收率，降低調查執行成本，混合模式 (mixed mode) 已普遍應用於大型追蹤調查。若不同調查模式之間的樣本非隨機分配，除了存在因不同模式產生的測量效應 (即模式效應)，同時還會有樣本以何種模式完訪的選樣效應。若擬探討不同模式對受訪者答題行為的影響，首要的挑戰是如何釐清並存的兩種效應。本研究採用可因應樣本非隨機分派的傾向分數配對法，探討大型追蹤調查答題行為的模式效應，比較總體題項無回應、拒答、「不知道」回答，以及態度量表的默許、極端回答風格在面訪模式與網路自填問卷模式之間是否有明顯的差異。分析資料來自「家庭動態調查」

- 
- \* 本研究是在國科會「混合模式調查的模式效應：2018 年家庭動態調查資料的分析」專題研究計畫 (108-2410-H-001-094) 經費補助下完成，特此申謝。此外，作者感謝三位審查人提供寶貴意見。然文責概由作者自負。
- \*\* 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研究員。臺北市南港區研究院路二段 128 號，電話：(02)27898169，電子郵件：suhao@gate.sinica.edu.tw。
- \*\*\* 中央研究院人文社會科學研究中心調查研究專題中心研究員。臺北市南港區研究院路二段 128 號，電話：(02)27898195，電子郵件：Email: yurr@gate.sinica.edu.tw。
- \*\*\*\* 中央研究院人文社會科學研究中心調查研究專題中心研究助理。臺北市南港區研究院路二段 128 號，電話：(02)27898131，電子郵件：tamao@gate.sinica.edu.tw。

在 2018 年以面訪、網路自填問卷模式同步蒐集到的問卷資料。由於面訪模式與網路自填問卷模式之間的樣本數差異大，為了提高傾向分數配對法的有效配對數，擴大估計的涵蓋範圍，以及降低估計誤差，本研究採用以擴大樣本（oversampling）與配對替代（replacement）為原則的兩種配對方法，分別為半徑配對（radius matching）與核函數配對。模式效應的估計結果顯示，網路自填問卷模式發生題項無回應與「不知道」回答的可能性明顯較面訪為高，與既有文獻的發現相近。在平衡型態度量表題組中，面訪的默許風格明顯較網路自填問卷模式嚴重，但網路自填問卷模式的極端回答風格明顯較面訪嚴重。其中，默許風格的發現如同預期，但極端風格的發現則不同於既有研究。儘管本文的研究課題限於題項無回應、態度量表回答風格兩類答題行為，而傾向分數配對法的應用上仍可能存在改善的空間，在調查方法研究、分析方法應用與調查實務上均具參考價值。

關鍵詞：追蹤調查、模式效應、題項無回應、默許、極端回答、傾向分數配對法

## Mode Effects on Response Behaviors in a Mixed-mode Survey

Su-hao Tu\*, Ruoh-Rong Yu\*\*, Tsung-Wei Hung\*\*\*

### ABSTRACT

In order to improve coverage and response rates, and to reduce

---

\* Corresponding Author. Research Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, No. 128 Sec. 2 Academia Road, Nankang 10529, Taipei city, Tel: (02)27898169, email: suhao@gate.sinica.edu.tw.

\*\* Research Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, No. 128 Sec. 2 Academia Road, Nankang 10529, Taipei city, Tel: (02)27898195, email: yurr@gate.sinica.edu.tw.

\*\*\* Research assistant, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, No. 128 Sec. 2 Academia Road, Nankang 10529, Taipei city, Tel: (02)27898131, email: tamao@gate.sinica.edu.tw.

survey cost, mixed modes have been commonly used in large-scale panel surveys. If the assignment of a survey mode is not random, the respondents' answers to a mixed-mode survey might be subject to two kinds of biases. One is measurement bias (i.e., mode effects) evoked by the modes themselves, and the other is sample selection bias, which results from the respondents' non-random assignment to different modes. How to disentangle these two biases is a crucial challenge for the estimation of mode effects. This study adopts propensity score matching, an analytical method which can deal with non-random sample assignment, to examine mode effects on response behaviors in a panel survey with a mixed-mode design of face-to-face and self-administered online modes. The outcome variables analyzed in this study include overall item nonresponse, "refusal" and "don't know" answers, and two response styles in balanced attitude scales, namely the acquiescence and extreme response styles.

Data analyzed in this study are from the Panel Study of Family Dynamics survey conducted in 2018, in which the sample was pre-assigned to face-to-face and self-administered online modes based on whether they provided an email address and finished an online questionnaire previously sent with a festival greeting card. There is a large difference in the numbers of complete questionnaires between the face-to-face and self-administered online modes. In order to improve the estimates for measurement effects, this study uses two matching methods, including the radius and kernel matching methods. These two matching methods are based on an oversampling strategy and matching with replacement.

The results of the two matching methods indicate that the probabilities of occurrence of item nonresponse and "don't know" answers in the self-administered online mode were significantly higher than those in the face-to-face mode, consistent with previous studies. It is also consistent with previous studies in that, regarding responses to the balanced attitudinal scales, respondents who finished the questionnaires

by face-to-face interviews were significantly more likely to provide acquiescent responses than those who filled out online questionnaires by themselves. However, different from previous studies, our findings indicate that respondents who completed self-administered online questionnaires were more likely to provide extreme responses to items on the balanced attitudinal scales than face-to-face interviewees. One other finding worth mentioning is that this study did not find significant mode effects for “refusal” answers.

This study contributes to research on mode effects, applications of propensity score matching, and survey practices. Our findings suggest that, to mitigate mode effects, a mixed-mode survey including face-to-face and self-administered modes should adopt the same design for “don’t know” and “refusal” options between modes, and respondents in the face-to-face mode should be allowed to enter answers to questions with social desirability concerns by themselves. Despite these academic and practical contributions, our study still has its limitations. One is that the response behaviors explored in this study are confined to overall item-nonresponse, “don’t know” and “refusal” answers, and the style of responses to attitude scales. In addition to extending the investigation of mode effects to a broader range of survey questions, future research should endeavor to increase the application of propensity score matching methods in order to disentangle the mode effects and selections effects in mixed-mode surveys. Future directions include but are not confined to the selection of covariates for the logistic model used to predict propensity scores, the methods of imputing missing values, and other matching strategies.

**Keywords:** panel survey, mode effect, item nonresponse, acquiescence, extreme response, propensity score matching

---

## 壹、前言

大型調查採混合模式（mixed mode）蒐集問卷資料已被廣泛應用（Börkan 2010; De Leeuw 2005, 2018; Dillman et al. 2009; Lynn 2013）。混合模式的優點是可以觸及到更多單一調查模式難以找到的受訪者，有助於提高涵蓋率（coverage rate）、完訪率（response rate），也能降低調查執行成本（Klausch et al. 2017; Vannieuwenhuyze et al. 2012; Vannieuwenhuyze and Loosveldt 2013）。國際大型追蹤調查常在單一模式進行到某階段後，改採混合模式。其主要考量是，面訪樣本的居住地點在後續追訪中會因搬遷等因素愈來愈分散，僅採單一模式不僅不利於追蹤完成率，也可能提高調查執行成本。又，受訪者對於調查的信任感常會隨著追訪而提高，因此能蒐集到受訪者多種聯絡資訊（如居住地址、聯絡電話、電子郵件信箱），提高運用混合模式的可行性。

不過，混合模式會不會造成調查結果的負面效應，普遍被討論的是混合模式所產生的測量效應（measurement effect），以下稱為模式效應（mode effect）。模式效應是指同一受訪者的回答純粹因接受的調查模式不同而有差異。如果樣本的完訪模式是隨機決定，即可較精確地估計模式效應。然而，當受訪者接受哪一種調查模式不是隨機決定，而受到樣本屬性（例如，有無電子郵件信箱、電話號碼）、樣本配置方式（例如，將較偏遠的樣本配置給非面訪模式）、受訪者對模式的偏好（例如，不喜好與人接觸的受訪者可能偏好自填問卷模式）等因素的影響時，除了模式效應外，同時還會產生樣本選擇效應（sample selection effect），造成不同模式完訪的樣本具有不同屬性（例如，以網路自填問卷模式完訪者，較可能有電子郵件信箱、較可能上

網，也較可能偏好網路自填問卷）的偏誤。<sup>1</sup>

由於樣本選擇效應常會與模式效應並存，探究混合模式中的模式效應變得較為困難（Biemer and Lyberg 2003; De Leeuw 2005; De Leeuw et al. 2008; Klausch et al. 2017; Kreuter et al. 2008）。近年的相關研究多聚焦在如何將樣本選擇效應、模式效應分離開來，以釐清不同模式本身帶來的模式效應（Klausch et al. 2014; Vannieuwenhuyze and Loosveldt 2013; Vannieuwenhuyze et al. 2014）。研究者提出的估計方法，多奠基於統計、計量等領域學者所提出的分析方法，如傾向分數分析法（propensity score analysis）、工具變項法（instrumental variable method）等。

本研究利用「家庭動態調查」（Panel Study of Family Dynamics, PSFD）於 2018 年採面訪、網路自填問卷混合模式蒐集的資料。由於 PSFD 採用的是同步（concurrent）執行的混合模式，且樣本分配到不同模式並非隨機決定，爲了能控制樣本選擇效應，本研究以傾向分數配對法（propensity score matching, PSM）估計模式效應。本研究關注的結果變項（outcome variable）爲問卷答題結果。首先是題項無回應，包含總體題項無回應（item nonresponse）、拒答（refusal）、「不知道」回答（don't know）。其次是態度量表的回答風格（response style），包含默許（acquiescence）與極端回答（extreme response）。爲檢視分析結果是否具有強韌性（robustness），我們採用兩種配對方法與不同的傾向分數配對指標，驗證配對結果的合理性與一致性。在嚴謹的分析下，希望分析結果不僅有助於瞭解在臺灣的社會、文化脈絡下，面訪、網路自填問卷兩種模式的測量效應有多大，亦有助於與國外相關研究進行對話。

---

1 對於混合模式調查中採用的「網路自填問卷模式」，爲便行文，文中與「網路模式」（online mode）一詞交替使用。

## 貳、文獻回顧

### 一、混合模式調查的模式效應估計

對於採用混合模式的調查，調查方法研究者最關心的問題，在於調查模式對研究者關注的結果變項，是否會造成模式效應及此效應的嚴重程度（Voogt and Saris 2005）。一般的混合模式調查中，個別受訪對象最終以哪類模式完訪，多半不是隨機決定的，因此並存著樣本選擇效應，以致難以釐清模式效應（Schouten et al. 2013; Vannieuwenhuyze et al. 2012; Vannieuwenhuyze and Loosveldt 2013）。樣本選擇效應是指樣本完訪模式並非隨機決定所造成的答題結果差異，不同模式完訪的樣本會有不同的人口特性。若研究者關心的結果變項與受訪者的人口特性相關，透過不同模式蒐集到的變項資料會有選樣誤差（sample selection bias）（Vannieuwenhuyze et al. 2012; Vannieuwenhuyze and Loosveldt 2013）。在有選樣誤差的情況下，不同模式的答題結果差異，無法視為純因模式不同所產生的差異。同一題項在不同模式得到的測量結果不同，除了模式效應之外，可能也並存著完訪模式的自我選擇問題而產生的樣本選擇效應。如何區分「樣本選擇效應」與「模式效應」是調查方法領域近十餘年的重要課題。為便讀者瞭解既有文獻採用的研究方法，以下依照相關文獻的研究方法進行討論。

#### （一）以實驗調查將樣本隨機分配到不同模式

探討模式效應最理想的方式是採樣本隨機分配的實驗設計（如 Jäckle et al. 2010）。Heerwegh（2009）以比利時某所大學的學生為調查對象分為兩群，進行面訪、網路自填問卷兩種模式的實驗研究。其

中一群學生以面訪施測；另一群學生則分兩階段進行調查，第一階段接受面訪，回答部分題項；第二階段上網填答其餘題項。Mariano and Elliott (2017) 的實驗研究將 45 間醫院的病人隨機分派到四種調查模式：郵寄問卷、電訪、互動式語音應答 (interactive voice response)、兼採郵寄問卷與電訪的混合模式。Schouten et al. (2013) 在荷蘭國家統局 2011 年執行的 Crime Victimization Survey (CVS) 中的實驗性調查，是將樣本隨機分配到面訪、網路自填問卷、郵寄自填問卷、電訪四種調查模式。將樣本隨機分派至不同調查模式的調查實驗，可免於樣本選擇誤差，直接觀察模式效應。但在調查實務中，考量調查執行成本、完訪率等因素，此類隨機法極為罕見。

## (二) 以登記資料作為真實答案

另一種探討模式效應的方法是以登記資料 (registration data) 作為「真實答案」，比較特定調查模式蒐集到的資料與登記資料的差異 (Fowler et al. 2002; Kreuter et al. 2010; Link and Mokdad 2006; Sakshaug et al. 2010; Voogt and Saris 2005)。但現實社會中，對研究者關心的結果變項，可能不存在對應的登記資料；或雖存在這類資料，但難以跟調查蒐集到的資料進行比較（如登記資料與調查資料的蒐集時點不一致，或測量不一致）。另外，登記資料本身亦可能存在測量誤差或資料缺漏的問題 (Klausch et al. 2017)。

## (三) 以單一模式或複查資料作為對照或底牌

Vannieuwenhuyze et al. (2010)、Vannieuwenhuyze and Loosveldt (2013) 指出，在對混合模式蒐集到的資料進行模式效應分析時，如另外採用單一模式對相同題項蒐集資料，可將單一模式蒐集到的資料



作為標竿，用以解決模式效應分析上的認定問題。以 Vannieuwenhuyze and Loosveldt 為例，除了混用郵寄問卷、面訪兩種模式外，另抽出一群樣本進行面訪調查，並以此調查結果作為模式效應分析的標竿。Biemer (2001)、Klausch et al. (2014)、Schouten et al. (2013)、Cernat and Revilla (2021) 則是以混合模式調查的複查 (retest) 資料作為標竿，藉由複查資料分析不同模式的模式效應。在調查實務上，要在混合模式之外以單一模式進行調查資料的蒐集，考量調查成本，可行性不高。另外，實際執行複查時，為達成較高的完成率，有可能兼採面訪、電訪或其他模式。複查資料本身也可能因為複查的調查模式的選用存在樣本選擇效應。

#### (四) 以統計方法解決模式效應的認定問題

採用混合模式的調查，不論是橫斷面調查或追蹤調查，由於實務上難以將樣本「隨機」分派到不同模式，或是難以利用登記資料、複查資料或其他方式進行分析，透過統計方法分離樣本選擇效應與模式效應，是較為實際且最重要的研究方式。採用混合模式在針對某個結果變項分析模式效應時，需要蒐集所有受訪者均參與不同模式的觀察值。然而，在一般的問卷調查中，我們通常只蒐集到特定受訪者在某一種模式下的回答資料，而無其他模式的答題結果。因而，在模式效應的估計上，需要藉由反事實分析 (counterfactual analysis) 方法，估計受訪者在其他模式中的潛在答案 (Vannieuwenhuyze et al. 2014)。

在調查模式並非隨機配置的情況下，其中一種分析途徑是借助 Rosenbaum and Rubin (1983)、Rubin (1974) 提出的後門 (back door) 法，找出可能影響完訪模式的一組共變項 (covariate)，藉以估計受訪者以特定模式完訪的傾向分數，再進行模式效應的估計。另一種研究

途徑則是應用 Pearl (1995, 2009) 提出的前門 (front door) 法，找出完訪模式、結果變項可能的中介變項，以分析完訪模式對結果變項的影響。在混合模式的模式效應分析上，後門法的應用較廣 (如 Lugtig et al. 2011; Vannieuwenhuyze et al. 2014)，而前門法只見於極少數的研究 (如 Vannieuwenhuyze et al. 2014)。然而，如 Vannieuwenhuyze et al. (2014) 的討論，不論前門法或後門法，均須遵循一些前提假設。

以混合模式完成資料蒐集的 2018 年家庭動態調查，由於樣本的完訪模式並非隨機決定，且本文關心的結果變項欠缺可供參照的登記資料或其他底牌，僅能藉由統計方法進行模式效應的估計。另外，由於家庭動態調查執行的過程中，蒐集了相當豐富的周邊資料，有助於找出可能關係最終完訪模式的影響因素。本文因此採用後門法，以傾向分數配對法分析模式效應。

## 二、不同調查模式的作答差異

由於本文所觀察的是面訪與網路自填問卷兩模式間受訪者的答題行為，包含整體問卷的題項無回應、「不知道」回答與拒答，以及態度題組的默許風格、極端回答風格，以下的文獻探討，將聚焦於這幾類答題行為在訪員訪談與受訪者自填問卷兩大類模式之間的研究結果比較。<sup>2</sup>

調查模式可依照有沒有訪員介入，大致分為兩類，一類是由訪員以一問一答的方式，向受訪者進行問卷資料蒐集 (如面訪、電訪)。另一類則是透過受訪者自填問卷的方式，進行問卷資料蒐集 (如網路

---

2 在此所探討的相關文獻，以已經正式出版的學術期刊論文與專書論文為主，不包含技術報告 (technical report) 或在機構內流通的工作論文 (working paper)。

自填問卷、紙筆自填問卷)。在訪員操作模式 (interviewer-administered mode) 中, 訪員通常需要依循調查單位的作業規範, 進行提問、追問、答案確認與記錄等。受訪者自填模式 (self-administered mode) 則仰賴受訪者自行閱讀題目、解讀題意、萃取回答所需的參考資訊、記錄答案, 相較由訪員操作的模式, 受訪者的認知負荷較大。此外, 在受訪者自填模式下, 受訪者在填答問卷時可能處於多工狀態, 且沒有訪員協助確認或追問答案。因而, 研究者指出, 在認知負荷較重且欠缺填答動力的情況下, 由受訪者自填問卷的模式, 受訪者在答題時常只會盡到最低滿意 (satisficing) 的認知努力 (Krosnick 1991; Fricker et al. 2005)。

基於前述的最低滿意理論, 調查方法研究者多主張, 相對於訪員操作模式, 題項無回應的問題在受訪者自填模式中會較為嚴重 (Fricker et al. 2005; Heerwegh and Loosveldt 2008)。Heerwegh (2009)、Heerwegh and Loosveldt (2008)、Díaz de Rada (2022) 對面訪、網路自填問卷的實驗研究均發現, 網路自填問卷模式的題項無回應、「不知道」回答的比例均顯著較面訪為高。而 Jones et al. (2016) 與 Jiang et al. (2017) 分別針對敏感性疾病及廣電相關題目的研究則發現, 電訪的 (總) 題項無反應比網路自填問卷明顯的還要多。不過, 仍有少數研究得到不同的結論, 如 Fricker et al. (2005) 針對電訪、網路自填問卷模式所做的調查實驗發現, 相較於電訪, 網路自填問卷模式的題項無回應較多, 但電訪的「無意見」回答則較多 (表 1)。

除題項無回應之外, 部分研究者針對量表或態度題組的回答風格, 進行不同模式的比較。以默許風格而言, 由於最低滿意作答行為在網路調查中較為嚴重, 自填問卷模式的默許風格可能較訪員操作模式來得嚴重。但另一方面, 在訪員操作模式之下, 受訪者為迎合訪員

表 1 題項無回應、量表回答風格的模式效應：相關文獻

	題項無回應或不知道						態度量表的回答風格						
							默許風格				極端風格		
研究來源	Fricker et al. (2005) <sup>1</sup>	Heerwegh & Loosveldt (2008) <sup>2</sup>	Heerwegh (2009) <sup>3</sup>	Jones et al. (2016) <sup>4</sup>	Jiang et al. (2017) <sup>5</sup>	Diaz de Rada (2022) <sup>6</sup>	Weijters et al. (2008) <sup>7</sup>	Heerwegh (2009) <sup>3</sup>	Heerwegh & Loosveldt (2011) <sup>8</sup>	Liu et al. (2017) <sup>9</sup>	Weijters et al. (2008) <sup>7</sup>	Heerwegh & Loosveldt (2008) <sup>2</sup>	Liu et al. (2017) <sup>9</sup>
網vs電(面)	>	>	>	>	>	>	<	=	=	<	<	=	<
研究對象	18 歲以上	大一新生	大一 大二生	18 歲以上	電視台會員	15 歲以上	一般民眾	大一 大二生	一般民眾	一般民眾	一般民眾	大一新生	一般民眾
樣本數	2,352	3,255	444	2,085	102	7,255	1,941	444	6,000	5,510	1,941	3,255	5,510
分析方法	卡方檢定	迴歸模型	t 檢定	迴歸模型	t 檢定	迴歸模型	結構方程	t 檢定	結構方程	迴歸模型	迴歸模型	結構方程	結構方程

註：表中的「>」、「=」、「<」符號分別標示「顯著高於」、「無顯著差異」、「顯著低於」。

資料來源：

1. Schulman, Ronca & Bucuvalas, Inc. 於 2003 年執行的 Practicum Survey。電訪 RDD 機率抽樣，對完訪樣本中有網路的受訪者 (N=2,352)，依隨機分配原則分派給電訪與網路自填問卷模式。
2. 比利時 Katholieke Universiteit Leuven 執行的社會調查。機率樣本 (N=3,255)，再隨機分配其中 3,000 案給網路自填問卷模式，255 案進行面訪。
3. 比利時 Katholieke Universiteit Leuven 執行的社會調查。機率樣本 (N=444)，隨機分配其中 186 案給網路自填問卷模式 (控制組)，其他的 258 案則讓受訪者選擇面訪或網路自填問卷模式 (實驗組)。
4. 依據民調公司提供的名單，針對加拿大英語及法語區的民眾，進行兩階段隨機抽樣法。
5. 針對美國中西部的小規模市場調查。對電視台共 2,073 位會員進行比例抽樣。
6. Survey of Linguistic Uses of the Population of Catalonia. 機率樣本，3,212 案以網路自填完成問卷，3,229 案以電訪完訪，其餘以面訪完訪。
7. 496 案電訪樣本來自一般人口資料，1,445 案網路樣本來自網路民調公司的 email 名單。兩樣本非隨機分配的結果。
8. 2007 年於比利時進行的實驗調查，自每兩年一次的 Security Monitor 全國性調查，以不成比例分層抽樣原則，分別對其中有住宅電話者抽出 50%的樣本 (3,000 案) 進行電訪，並對沒有住宅電話者抽出 50%的樣本 (3,000 案) 進行網路調查。
9. 面訪樣本是以 18 歲以上成人為母體，採分層多階段方法抽出樣本，於 2012 年執行調查的美國選舉研究調查完訪樣本 (1,929 案)；網路調查樣本則是具有母體代表性且具機率樣本特性的 GfK Knowledge Panel 完訪樣本 3,581 案。

的提問，可能會做出肯定的表態（yea saying），選擇正向的回答選項（例如「同意」、「贊成」、「符合」等），而在回答上呈現較強烈的默許風格（Heerwegh 2009; Liu et al. 2017）。如 Weijters et al.（2008）與 Liu et al.（2017）分別針對電訪與網路調查，以及面訪與網路調查進行比較，發現電訪與面訪的默許風格均較網路調查嚴重。然而，Heerwegh（2009）對面訪、網路調查的研究，以及 Heerwegh and Loosveldt（2011）對電訪、郵寄紙本問卷的研究，均未發現明顯差異（表 1）。

至於極端回答風格，研究者指出，由於調查單位通常期望受訪者對態度題組能夠明確表態，在訪員操作的模式下，受訪者可能偏向選擇極端的選項，以顯示對訪員提問的配合；自填問卷模式則因為沒有訪員進行確認與追問，而讓受訪者較可能選擇偏中間的選項（Liu et al. 2017）。此外，由於受訪者自填模式仰賴「視覺」閱讀題目，而訪員操作模式依靠「聽覺」，在前者以視覺為主的情境下，受訪者較有可能選擇偏中間的選項（Heerwegh and Loosveldt 2008）。Weijters et al.（2008）的實證研究發現，電訪的極端回答風格較網路調查明顯；Liu et al.（2017）亦發現，面訪的極端回答風格較網路調查明顯。但 Heerwegh and Loosveldt（2008）則發現面訪、網路調查的極端回答情況相近（表 1）。

對訪員操作模式、受訪者自填模式，研究者指出，儘管這兩大類的模式在資料品質、回答型態上存在較大的差異，同一大類的不同模式（如同樣由訪員以問卷訪談蒐集資料的面訪、電訪，或同樣由受訪者自行填答問卷的郵寄、網路自填問卷模式）的差異較小（Schouten et al. 2013）。儘管如此，仍有不少研究者針對同一大類的不同模式進行資料品質、作答行為上的研究，並發現模式之間的差異。如 Holbrook et al.（2003）對面訪、電訪的研究指出，相較於面訪，電訪的受訪者

較可能提供最低滿意的答案，而出現沒有意見（no opinion）、默許等類型的答案。

本研究擬以「家庭動態調查」這項追蹤調查計畫於 2018 年由全面訪模式轉換為面訪、網路自填問卷混合模式所蒐集到的問卷資料，分析兩種調查模式問卷答題行為的差異。對於整體問卷的題項無回應、「不知道」回答與拒答，以及態度量表的默許風格、極端回答風格，綜合前人的研究結果（表 1），本研究預期，網路自填問卷模式的總體題項無回應、「不知道」回答、拒答較面訪模式嚴重。在態度量表回答風格方面則預期，面訪模式的默許風格與極端回答風格相對較網路自填問卷模式嚴重。

## 參、資料與研究方法說明

在追蹤調查的後續追訪中，基於調查執行成本的考量，多採用次序性（sequential）模式，先以成本較低的模式（如網路自填問卷、電訪）啟動調查；對該模式執行期間未能完訪的樣本，再以成本較高的模式（如面訪）進行問卷資料蒐集（Klausch et al. 2017）。這樣的次序性混合模式在進行模式效應分析時，會面臨一些問題。首先，以「先行」模式完訪的受訪者，多為配合意願較高、較容易接觸到的樣本；採用「後行」模式完訪的受訪者，則可能是「先行」模式中不願受訪或未能接觸到的樣本。在不同模式之間，受訪對象的自我選擇問題可能較為嚴重。此外，由於不同模式的執行有先後順序，在時間因素的干擾下，會使得模式效應的估計更為困難。本研究利用「同步」混合模式所蒐集的調查資料，並採用可控制選樣誤差的傾向分數分析方法，以避免「次序性」模式額外衍生的模式運用順序誤差問題，在控

制樣本自我選擇誤差下，期能更精確地估計問卷答題行為的模式效應。以下說明資料來源、分析方法，及變項測量。

## 一、資料來源

本研究的分析資料來自 2018 年家庭動態調查。家庭動態調查自 1999 年開始執行追蹤調查，主樣本為滿 25 歲以上的人口。首波主樣本（1953-64 年次）的面訪調查（RI 問卷）以全臺灣為範圍，隨機抽出訪問對象。後續分別於 2000、2003、2009、2016 年進行新抽主樣本（分別為 1935-54、1964-76、1977-83、1984-91 年次）的首波面訪。在 2012 年之前，主樣本追蹤調查（RR 問卷）採逐年進行；自 2012 年起，改為每兩年追訪一次。自 2000 年起進行子女問卷（C 問卷）面訪，將訪問對象擴展到主樣本的 16-24 歲子女。其後每兩年進行一次的子女問卷面訪對象，均納入主樣本年屆 16 歲的子女。子女樣本在年滿 25 歲以前參與 C 問卷訪問；在滿 25 歲時則視同主樣本，以結構與 RI 問卷類似的 RCI 問卷進行訪問；在完成 RCI 問卷之後，則連同主樣本一併以 RR 問卷進行追訪。

家庭動態調查在 2018 年調查之前，除了極少數的個案外，幾乎全數都採面訪調查。在成本的考量下，2018 年調查有部分樣本改採網路自填問卷模式，以電子郵件邀請樣本上網自填問卷；其他樣本則仍採面訪調查。在該年調查中，採用面訪、網路自填問卷混合模式的樣本，僅有適用主樣本追蹤問卷（RR2018 問卷）的樣本。未滿 25 歲的子女樣本問卷（C2018 問卷）與子女樣本轉主樣本後的首波問卷（RCI2018 問卷）則仍採面訪。因而，本研究的分析對象設定為主樣本追蹤問卷的完訪樣本。

2018 年調查在前置作業階段即依樣本屬性及過往填答狀況，將樣



本分配到網路自填問卷模式或面訪模式。分派至網路自填問卷模式的樣本，除了需要有電子郵件信箱外，還須填答過 2017 年新年賀卡附帶的網路問卷，或於收到中秋、新年賀卡時上網填寫過回函。此外，若樣本所屬家戶中有其他成員亦為家庭動態調查的受訪對象，為避免同一家戶的樣本分採不同模式可能帶來的困擾，同一家戶的所有訪問對象均需要符合前述條件，才會分派到網路模式；否則該家戶的受訪者仍以面訪為原則。除了既定的非隨機樣本分派原則，在調查執行過程中，可能會因為實際的接觸情況或受訪者的意願而轉換調查模式（例如：原本分派至面訪，但經訪員多次探訪仍無法完訪的受訪者，會邀請上網填寫問卷）。本研究的分析樣本為 2018 年調查中，完成 RR 問卷的 6,287 位受訪者（于若蓉 2019）。

## 二、分析方法與模型設定

### （一）分析方法

要瞭解面訪或網路自填問卷模式之間是否存在模式效應，最理想的方式是進行混合模式的調查實驗，將樣本隨機分派到面訪、網路自填問卷模式，進而利用迴歸模型等方式分析調查模式對問卷答題行為的影響。採用迴歸模型分析的前提假設之一是調查模式為外生的（exogenous）變項。若分析資料不符合此前提假設，仍採迴歸分析方法，即使在模型中控制了其他可能的自變項，由於未考慮調查模式的樣本自我選擇問題，會得到偏誤的估計結果（Berhman et al. 2004）。本研究分析的追蹤調查資料，樣本並非隨機分配到不同的調查模式，由於調查模式並非外生變項，而可能受到其他自變項（如：人口特徵）或其他無觀測的因素的影響，為了避免迴歸模型對問卷答題行為的模式效應產生估計偏誤，以及樣本選擇效應和模式效應並存的問題，本



研究採傾向分數配對法，在控制樣本選擇效應下分析模式效應。<sup>3</sup>

本研究以「網路自填問卷」為目標組 (target group)，「面訪」為對照組 (non-target group)，並以虛擬變項 (dummy variable)  $W$  表示分析對象屬於目標組 ( $W=1$ ) 或對照組 ( $W=0$ )。 $Y$  為結果變項，亦即問卷答題結果。目標組 ( $W=1$ )、對照組 ( $W=0$ ) 對應的結果變項分別以  $Y_1$ 、 $Y_0$  表示，其中的下標 1、0 分別代表目標組、對照組。對目標組，其結果變項的期望值為  $E(Y_1|W=1)$ ；對照組結果變項的期望值為  $E(Y_0|W=0)$ 。本研究關心的平均處理效果 (average treatment effect, ATE) 為前者與後者的差異，如以下公式所示 (Guo and Fraser 2015: 25)。

$$ATE = E(Y_1|W=1) - E(Y_0|W=0)$$

## 1. 傾向分數配對法：原理與配對方法

傾向分數分配對法的原理，是利用一組與處理變項 (treatment variable) 或結果變項息息相關的共變項 ( $x_i$ ) 估計出傾向分數，進行目標組與對照組樣本的配對，再針對結果變項計算 ATE 的估計值 (Rosenbaum and Rubin 1985)。在本研究的架構下，處理變項為網路模式完訪與否的虛擬變項，傾向分數則是針對處理變項，在給定的共變項之下，以 logistic 迴歸模型預測的條件機率 (conditional probability) 值。<sup>4</sup> 此機率值又稱傾向分數，是進行兩組觀察樣本的「配對」(matching)

---

3 基於審查人的意見，本研究除 PSM 外，也曾以多元迴歸分析調查模式對問卷答題行為的影響，結果發現估計值會高估模式效應。此外，也利用 Stata 中的 KHB 法 (Kohler et al. 2011)，分析調查模式與本研究納入的所有共變項對結果變項是否存在中介影響效果，結果發現中介效果並不存在。在篇幅考量下，相關結果未列於文中，有興趣的讀者可向作者索取分析內容。

4 文獻指出，雖可採用 probit 迴歸模型等其他估計方法，logistic 迴歸模型仍是最常被採用的方法 (Harris and Horst 2016)。因而，本研究亦採用 logistic 迴歸模型。

的依據，配對後即可減少無法觀測的選擇誤差（unobserved selection bias）。

在各種傾向分數配對法中，常用的包括最鄰近配對法（nearest neighbor matching）、半徑（radius）配對法、核函數（kernel）加權配對法等。最鄰近配對法是以傾向分數為依據，在對照組中尋找傾向分數最接近的個案，以進行目標組、對照組樣本的配對（Harris and Horst 2016）。此法較容易產生配對品質不良的問題。例如，當傾向分數高的個案在目標組中很多，在對照組中卻很少時，會產生高傾向分數的個案配對到低傾向分數者的不良配對結果（Caliendo and Kopeinig 2008; Austin 2011; Harris and Horst 2016）。在配對方式上，如果採用一對一且「不可替代」的配對（pair matching without replacement），很容易產生大量無法配對的樣本數，對配對平衡性（balance）不利，並可能帶來估計誤差（King and Nielson 2019）。修正的辦法是改以「可替代」（with replacement）、卡鉗調整（caliper adjustment）、擴大取樣（oversampling）、加權擴大取樣（weights for oversampling）的一對一或一對多配對（Caliendo and Kopeinig 2008）。在對照組的樣本數相對較小時，文獻多建議採「可替代」的配對法以提高配對的品質；若對照組的樣本數相對大，則多建議一對多配對法（Baser 2006）。針對小樣本，前人研究顯示，一對一且「不可替代」的配對結果，樣本平衡性較佳；但採一對多配對時，相較「不可替代」配對，「可替代」配對具有較佳的樣本平衡性及較小的估計變異（Bottigliengo et al. 2021a）。

不論是「可替代」配對或擴大取樣（一對多）配對，都面臨替代或多重配對的次數可容許的上限該是多少的問題。若無上限，同一個案被多次配對會因此降低配對品質，增加估計變異。解決方法之一是採卡鉗調整，透過傾向分數距離的設定，在可接受的距離範圍內進行

配對；範圍之外無法配對者則予以排除。若設定的卡鉗距離越小，可有效配對的樣本數會越少，估計結果的母體涵蓋率（coverage）會越低（Harris and Horst 2016）。但若搭配文獻中普遍採用的「可替代」配對法，可提高有效的配對數（Austin 2009; Baser 2006; Ripollone et al. 2018）。Rosenbaum and Rubin（1985）將卡鉗距離設定在 0.25 倍標準差，以期能降低至少百分之九十的估計誤差。Austin（2009）則依循 Cochran and Rubin（1973），以降低百分之九十八的估計誤差為原則，進行一對一「不可替代」的配對，並比較兩種卡鉗距離的估計結果，發現卡鉗距離設在 0.20 的估計誤差比設在 0.60 還要低。不過，卡鉗距離的選擇仍有討論的空間（Austin 2009; Austin 2011; Caliendo and Kopeinig 2008; Harris and Horst 2016）。

由於本研究中對照組的樣本數比目標組的樣本數大很多，兩組的傾向分數分布可能有很大的差異，如果採用一對一配對，恐因此降低有效的配對數，增加模式效應的估計誤差，降低估計的母體涵蓋率。為了進行傾向分數配對時能在對照組中找到可配對的個案，我們綜合考量上述各種修正法，在「可替代」的原則下，採用兩種擴大取樣的配對方法—半徑配對法（以下簡稱半徑法）與核函數加權配對法（以下簡稱核函數法）。半徑法主要利用卡鉗來進行半徑的設定，在此半徑距離內找傾向分數最接近的對照組個案來配對。本研究的卡鉗設定為「傾向分數標準差」的 0.25 倍。半徑法具有最鄰近配對法輔以擴大取樣及卡鉗調整的優勢，在合理配對之半徑距離範圍下，並採用「可替代」原則，理論上會有較好的配對品質（Baser 2006; Caliendo and Kopeinig 2008; Huber et al. 2015）。核函數法是透過某一寬帶（bandwidth）的設定，對於對照組個案進行「加權」，讓每個目標組觀察值和（多）個對照組觀察值進行配對。若配對中的兩者距離相距較遠，給較大的權

重；若距離相近，則給較小的權重。寬帶的設定本身可以在函數中發揮調節變異數（variance）的作用。此法適用於樣本數在目標組與對照組之間呈現不均勻或不對稱分布時，尤其是傾向分數分布呈現不連續的斷層，能有效降低估計誤差（Baser 2006; Caliendo and Kopeinig 2008; Galdo et al. 2008; Hazlett 2020）。在實際的估計上，本研究是透過 Stata 統計軟體的 psmatch2 套件來進行傾向分數配對的操作。對核函數法，本研究是採用 biweight 函數作為核函數的設定。<sup>5</sup>

## 2. 傾向分數配對法：配對結果的評估

採用傾向分數配對法進行目標組、對照組樣本配對之前，本研究先確認分析資料符合三個假設：（1）配對後結果變項的數值與處理變項無相關性（ignorability）。（2）每個配對後的傾向分數同時包含目標組與對照組的觀察值，此一條件稱為 common support 或 overlapping。（3）配對後的每個共變項的分配在目標組與對照組之間不存在任何關聯性，亦即，應無顯著的差異而具有平衡性。所謂的平衡性是指，在真實的傾向分數下，樣本屬於「目標組」或「對照組」與用來產生傾向分數的共變項（ $x_i$ ）無關（亦即，獨立於  $x_i$ ）。換言之，具有相同傾向分數的目標組、對照組個案應該在共變項上有相近的分配（Rosenbaum 2002; Rosenbaum and Rubin 1983）。如果傾向分數配對結果無法通過平衡性檢測，表示所採用的傾向分數估計模型可能存在設定上的問題。本研究參酌相關文獻，採用以下幾項平衡性測量指標，協助判

---

5 本研究亦曾嘗試過 Epanechnikov kernel 與 Gaussian kernel 此兩類 kernel 函數，但在後續的傾向分數配對檢測中，因部分結果變項的配對效果不佳，而未呈現相關分析結果。至於 tricube 函數，由於估計結果與 biweight 函數估計結果相近，未將結果呈現於文中。有興趣的讀者，可向作者索取相關結果。

斷配對後樣本的平衡性。

(1) 標準化偏差 (standardized mean bias)

此為共變項在目標組與對照組之間的「平均值差」(difference in means)除以「變異數平均值的平方根」(square root of the average of variance)後得到的數值。標準化偏差的數值越小，表示共變項在兩組間的差異越小，平衡性的表現越佳。Caliendo and Kopeinig (2008)認為可接受的數值在 3% 至 5% 以下，而 Austin (2011, 2013, 2014)則以 10% 以下作為標準。

(2) 虛擬判定係數 (pseudo R-squared)

虛擬判定係數為處理變項 logistic 迴歸模型的解釋能力指標，越接近 1 表示模型變項與預測變項間的相關性越高。在進行傾向分數配對後，若虛擬判定係數明顯降低，則表示共變項與處理變項的關聯性有所縮減。

(3) logistic 迴歸模型卡方值 (LR chi-squared)

卡方值的作用與前述虛擬判定係數相似，可用於判斷共變項與處理變項之間是否存在關聯性。透過卡方值對應的 p 值，可比較配對前後的卡方值是否有顯著變化。

(4) B 值與 R 值

B 值是傾向分數在「目標組」與「對照組」的差異，再經標準化處理後的數值 (standardized difference of the propensity score)。Rubin (2001)建議，經過傾向分數配對後得到的 B 值應小於 25%。R 值是指「目標組」與「對照組」傾向分數變異數的比值 (ratio of variance of the propensity score)。在經過傾向分數配對後，理想的 R 值應接近 1。Rubin (2001)建議，可接受的 R 值應介於 0.5 與 2 之間。

除前述平衡性測量指標外，Rosenbaum (2002) 等研究指出，可藉由敏感度分析 (sensitivity analysis)，檢視 logistic 迴歸模式納入的共變項是否適切。對於連續的結果變項，本研究依循 Rosenbaum (2002)，採用以下的 Gamma 統計量作為敏感度指標。

$$\frac{1}{\Gamma} \leq \frac{\frac{\pi_j}{(1-\pi_j)}}{\frac{\pi_k}{(1-\pi_k)}} = \frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)} \leq \Gamma$$

其中， $j$ 、 $k$  代表目標組、對照組樣本， $\pi_j$ 、 $\pi_k$  代表  $j$ 、 $k$  兩組樣本分配到目標組的機率，亦即傾向分數，而 Gamma 值則為兩組樣本相對傾向分數的比值。若 Gamma 值等於 1，表示樣本是隨機分配到目標組與對照組。假設 Gamma 值為 2，表示分配到某一組的相對機率是另一組的兩倍，顯示偏離隨機配置，可能存在著未觀測到的共變項 (Rosenbaum 2005)。對於二元 (binary) 結果變項，Aakivk (2001) 建議，可改用 Mantel-Haenszel Test，對應的統計量如後：

$$Q_{MH} = \frac{|Y_1 - \sum_{s=1}^S E(Y_{1s})| - 0.5}{\sqrt{\sum_{s=1}^S Var(Y_{1s})}}$$

其中， $Y_1$  表示配對後目標組的樣本數， $Y_{1s}$  表示成功配對後目標組的樣本數。

文獻建議以敏感度指標搭配顯著性檢定結果，來判斷模型估計的穩定性。若 Gamma 數值大於 1 且檢定結果未達顯著水準，表示模型設定存在較大的異質性，可能高估或低估處理效應。即便 Gamma 值等於 1，若檢定結果顯著，表示模型設定有不穩定的疑慮。文獻中對於敏感度指標臨界值的合理範圍尚無定論，不過實務上通常建議不要超過 5 (Becker and Caliendo 2007)。

## （二）變項、模型設定與建構

本研究的結果變項包含問卷的總體題項無回應、拒答、「不知道」回答，以及態度量表題組的兩種回答風格，包含默許與極端回答。其中，總體題項無回應、拒答、「不知道」回答這三個結果變項均採虛擬變項設定，如果有一題或一題以上的無回應、拒答、「不知道」回答，則設數值為 1，否則為 0。回答風格是依據經過平衡設計問法的兩個態度量表（其中有近半數的題目採反向概念或敘述），分別是婚姻態度與性別角色態度（附錄 1），採用的測量是這兩個態度量表的 16 個題項中出現默許、極端回答風格的題項比例。處理變項是以網路模式完訪與否的虛擬變項。

用以產生傾向分數的共變項是經由下列步驟篩選、決定。首先，依據相關理論及前人研究初步篩選可能影響調查模式選擇、問卷答題品質的變項（Datta and Terrell 2002; Heerwegh and Loosveldt 2008; Heerwegh and Loosveldt 2011; Hibbing et al. 2019; Liu et al. 2017; Valentino et al. 2020）。由於 PSM 的共變項不可受到處理變項（調查模式選擇）與結果變項（問卷答題行為）的影響，初步篩選出來的變項需符合此條件，第二個步驟乃針對初步篩選出來的變項進行多重相關分析與中介效應分析，以確定處理變項是否為獨立的自變項以及共變項的獨立性（Garrido et al. 2014）。其中，本研究是利用 Stata 的 KHB 法（Kohler et al. 2011）進行中介效應分析，檢驗初步篩選出來的共變項是否為調查模式選擇與結果變項間的中介變項，若否才進一步以 logistic 迴歸模型估計傾向分數。<sup>6</sup> 第三個步驟是在執行 PSM 的每個階段中進行傾向

---

6 有興趣的讀者可向作者索取關於 KHB 法的分析結果。



分數的重疊性、配對後共變項的樣本平衡性，與模式敏感度的檢定，反覆地檢視共變項的獨立性（Caliendo and Kopeinig 2008; Yanovitsky et al. 2005），並嘗試不同的共變項組合或不同的變項操作型定義，如將連續變項改為類別變項、類別變項重新分類。本研究所呈現的結果並未納入文獻建議可以考慮的交互項（Garrido et al. 2014），主要有兩個理由，其一是不含交互項的共變項已符合應有的獨立性；此外，雖然調查模式與某些人口特徵或人格特質可能具有交互作用，但尚無前人分析研究結果可供參考。

本研究依據前述方式，對可能影響完訪調查模式、問卷答題行為的變項進行篩選。最後篩選確定的共變項，依其測量的概念層次區分為三大類：（1）社會人口基本特徵、（2）調查參與行為、與（3）人格特質。第一類包含基本人口特徵、社經特徵、家戶特徵與社會網絡等相關變項，有受訪者性別、年齡、教育程度、婚姻狀態、居住地區鄉村化程度、工作狀態、工作收入、同一家戶有無其他受訪者、每日接觸人數。第二類是調查參與行為變項。納入這類變項的主要考慮是，受訪者在上一波調查的參與行為可能會影響下一波調查的模式選擇與問卷答題行為。此類變項來自上一波（2016 年）問卷完訪後由訪員對受訪者行為的觀察紀錄，包含受訪者在訪問過程有沒有表達過拒絕受訪與答題耐心程度。第三類是受訪者的人格特質。有研究者指出，受訪者的人格特質可能左右其調查模式的偏好並影響題項無回應與回答風格（Hibbing et al. 2019; Valentino et al. 2020）。本研究採用大五人格量表中的親和性（agreeableness）、開放性（openness）及外向性（extroversion）三個面向的人格特質（附錄 2）。依據上述三類共變項，本研究在不同的變項組合下，設定三個 logistic 迴歸模型，以估計傾向分數。模型一包含第一類共變項，模型二包含第一類與第二類



共變項，模型三則將三類共變項都納入。設定三個模型的主要目的是藉由每新增一類共變項，觀察調查模式效應大小與顯著程度的變化，不僅作為共變項選取的參考，也透過共變項的樣本平衡性、預測傾向分數之模型的穩定性、及模式效應的估計結果，比較不同模型的調查模式效應估計結果。模式中的結果變項、處理變項與共變項的詳細操作定義，請參照附錄 3。

對於模型中選取的共變項，完訪問卷中難免存在資料缺漏的情況，本研究採兩階段方式回填資料。首先，以前一波（2016 年，包括 RI2016、RCI2016、RR2016 三份問卷）資料中比較不會因時間而大幅變動的出生年次、性別、教育程度、居住地區、薪資、工作狀態等，進行 2018 年對應資料的回填。另外，用以測量人格特質的「大五人格」量表，在 2018 年調查中，由於僅有 RR2018、RCI2016 與 RI2016 問卷的樣本群需要回答，對於其他樣本群的缺漏資料，我們以 2009 年 RR 問卷的大五人格量表資料回填；但由於 2009 年新抽的主樣本群並未詢問過大五人格量表，無法回填。<sup>7</sup> 其次，針對第一階段資料回填後仍缺漏的薪資，利用年齡、性別與教育進行多重插補（multiple imputation by chained equations, MICE）（Bottigliengo et al. 2021b; Choi et al. 2019），產生的 15 個插補集，估計標準誤為 0.261。最後，對於無法回填又不適合插補的其他共變項，則以 listwise deletion 方式刪除缺漏值。

最後實際分析的樣本數，會隨著採用的模型與分析的結果變項有所差異。在題項無回應相關結果變項的分析上，由於模型三納入人格特質後會刪除無法回填缺漏值的 2009 年新抽主樣本群，觀察值明顯

---

7 詳細的「家庭動態調查」問卷資料架構，可以參考于若蓉、黃奕嘉（2018）的說明。

比模型一與模型二少。另外，因為不是所有樣本都有在 2018 年調查中回答婚姻態度與性別角色態度量表，分析回答風格的模式效應時，無論哪一個模型，樣本數均不到一千八百。不同結果變項、模型對應的分析樣本數，請參見附錄 4。

## 肆、分析結果

### 一、各變項的描述統計分析結果

如前所述，我們採用三種迴歸模型進行傾向分數配對分析，模型一與模型二有相同的樣本數，而模型三的樣本數較少，在呈現描述統計分析結果的表格與附錄中，為求簡化且便於讀者了解，基於模型一與模型二採用同樣的樣本，兩個模型的描述統計數據併為同一欄；模型三的統計數據則以單獨的欄位呈現。針對描述統計數據，首先，在題項無回應方面，「至少有一個題項無回應」的樣本數接近一半，只比「至少有一個題項回答不知道」的樣本數多一些；「至少有一個題項拒答」的樣本占比不到 6%。兩種回答風格的情況也不嚴重，默許、極端回答風格兩個結果變項的平均值分別為 0.31、0.20（附錄 4）。其次，在 2018 年調查的完訪樣本中，以網路自填問卷完訪的比例，在模型一／二、模型三間有些微差異，分別為 12% 與 11% 左右（附錄 5）。

再者，關於本研究的共變項，在模型一／二與模型三中的樣本，平均年齡分別為 42.07、44.67 歲，無論模型一／二或模型三的樣本，男性、已婚者、具大專或以上的教育程度者、居住於工商市鎮或新興市鎮者均超過半數；有正式工作者超過七成，且工作者的薪資多介於兩萬五到五萬之間。家中有其他受訪者的情況約三成，而每日接觸人數達 10 人或以上者達七成五；關於前一波（2016 年）的調查參與行

爲，只有不到 5%、15% 的樣本在訪問過程中表示過拒訪、有過不耐煩。而三個面向的人格特質中，親和性的平均分數高於外向性與開放性（附錄 5）。進一步比較目標組（網路模式）與對照組（面訪模式）的樣本分配，無論模型一／二或模型三，網路模式樣本的平均年齡均較低，居住地區的鄉村化程度較低，有較高比例的樣本是大專院校以上教育程度、有正式工作、從未結過婚、薪資高於兩萬五千元、每日接觸人數小於 20、2016 年受訪時未表示過拒訪、2016 年受訪時從未有不耐煩的情形（表 2）。依據模型三的兩組間樣本分配，網路模式樣本的人格特質較具開放性，但較不具親和性。再進一步從標準化平均差（standardized mean difference, SMD）來看樣本平衡性，若依據 Austin（2009）以 0.1 以下爲理想標準，表 2 顯示除了模型一／二的性別、已婚者與拒訪傾向以及模型三的薪資在 0-2.5 萬元、每日接觸人數在 10-49 人、與外向性人格特質之外，其他共變項的樣本平衡性均不佳。綜上顯示，本研究採用兩種配對法能否改進平衡性，值得探討與驗證，詳見本文後續的分析結果。

**表 2 共變項的描述統計及標準化平均差（SMD）（缺漏資料回填後）**

	模型一／模型二			模型三		
	面訪 (n=5,457)	網調 (n=744)	SMD	面訪 (n=3,567)	網調 (n=467)	SMD
年齡	42.95 (15.06)	35.61 (7.26)	-0.62	45.90 (17.14)	34.58 (8.92)	-0.83
性別 (男性 = 1；女性 = 0)	0.53 (0.50)	0.48 (0.50)	-0.10	0.53 (0.50)	0.47 (0.50)	-0.12
教育程度						
國(初)中或以下	0.18 (0.38)	0.01 (0.09)	-0.61	0.24 (0.43)	0.02 (0.11)	-0.69

表2 共變項的描述統計及標準化平均差(SMD)(缺漏資料回填後)(續)

	模型一／模型二			模型三		
	面訪 (n=5,457)	網調 (n=744)	SMD	面訪 (n=3,567)	網調 (n=467)	SMD
高中職或五專	0.27 (0.44)	0.08 (0.28)	-0.52	0.26 (0.44)	0.08 (0.28)	-0.49
大專院校	0.44 (0.50)	0.64 (0.48)	0.40	0.40 (0.49)	0.64 (0.48)	0.49
研究所(或)以上	0.11 (0.32)	0.27 (0.44)	0.42	0.10 (0.30)	0.26 (0.44)	0.43
家戶中有無其他受訪者	0.33 (0.47)	0.15 (0.36)	-0.43	0.32 (0.47)	0.14 (0.35)	-0.44
目前有無「正式」工作 (有=1)	0.77 (0.42)	0.89 (0.31)	0.32	0.71 (0.45)	0.88 (0.33)	0.43
婚姻狀態						
從未結婚	0.36 (0.48)	0.46 (0.50)	0.20	0.36 (0.48)	0.55 (0.50)	0.39
已婚	0.54 (0.50)	0.52 (0.50)	-0.04	0.53 (0.50)	0.43 (0.50)	-0.20
分居、離婚或喪偶	0.10 (0.30)	0.02 (0.14)	-0.34	0.11 (0.32)	0.02 (0.15)	-0.37
工作薪資						
沒有正式工作或 0	0.24 (0.43)	0.12 (0.32)	-0.32	0.30 (0.46)	0.14 (0.35)	-0.39
>0 且≤25K	0.13 (0.34)	0.09 (0.28)	-0.13	0.13 (0.34)	0.10 (0.30)	-0.09
>25K 且≤50K	0.42 (0.49)	0.50 (0.50)	0.16	0.39 (0.49)	0.52 (0.50)	0.26
>50K	0.21 (0.41)	0.29 (0.45)	0.19	0.18 (0.39)	0.24 (0.43)	0.15

表2 共變項的描述統計及標準化平均差(SMD)(缺漏資料回填後)(續)

	模型一／模型二			模型三		
	面訪 (n=5,457)	網調 (n=744)	SMD	面訪 (n=3,567)	網調 (n=467)	SMD
每日接觸人數						
0-9 人	0.22 (0.41)	0.31 (0.46)	0.20	0.24 (0.43)	0.31 (0.46)	0.16
10-19 人	0.31 (0.46)	0.36 (0.48)	0.11	0.32 (0.47)	0.36 (0.48)	0.08
20-49 人	0.28 (0.45)	0.22 (0.42)	-0.14	0.26 (0.44)	0.22 (0.41)	-0.09
50 人以上	0.19 (0.39)	0.11 (0.31)	-0.23	0.18 (0.38)	0.11 (0.31)	-0.20
居住地區鄉村化程度						
都會核心	0.10 (0.30)	0.14 (0.34)	0.12	0.10 (0.30)	0.16 (0.37)	0.18
工商市鎮	0.29 (0.45)	0.42 (0.49)	0.27	0.28 (0.45)	0.39 (0.49)	0.23
新興市鎮	0.24 (0.43)	0.22 (0.41)	-0.05	0.22 (0.42)	0.22 (0.41)	0.00
傳統產業市鎮	0.09 (0.29)	0.09 (0.28)	0.00	0.11 (0.31)	0.10 (0.30)	-0.03
低度發展鄉鎮	0.18 (0.38)	0.11 (0.31)	-0.20	0.19 (0.39)	0.11 (0.32)	-0.23
高齡化鄉鎮	0.10 (0.30)	0.03 (0.17)	-0.29	0.10 (0.30)	0.03 (0.16)	-0.29
偏遠鄉鎮	0.00 (0.03)	0.00 (0.06)	0.00	0.00 (0.03)	0.00 (0.00)	0.00
2016 年受訪時是否中途表示拒訪						
沒有表示拒訪	0.95 (0.23)	0.96 (0.20)	0.05	0.94 (0.23)	0.97 (0.18)	0.15

表2 共變項的描述統計及標準化平均差(SMD)(缺漏資料回填後)(續)

	模型一／模型二			模型三		
	面訪 (n=5,457)	網調 (n=744)	SMD	面訪 (n=3,567)	網調 (n=467)	SMD
曾表示拒訪	0.04 (0.20)	0.03 (0.17)	-0.05	0.05 (0.22)	0.03 (0.18)	-0.10
2016 年未完訪	0.01 (0.11)	0.01 (0.10)	0.00	0.01 (0.08)	0.00 (0.05)	-0.14
2016 年受訪時不耐煩	0.24 (0.69)	0.15 (0.58)	-0.14	0.25 (0.65)	0.14 (0.50)	-0.19
從未	0.86 (0.35)	0.91 (0.28)	0.16	0.84 (0.37)	0.90 (0.30)	0.18
很少	0.08 (0.27)	0.05 (0.22)	-0.12	0.10 (0.29)	0.07 (0.25)	-0.11
有時	0.04 (0.20)	0.02 (0.14)	-0.12	0.05 (0.21)	0.02 (0.14)	-0.16
經常	0.01 (0.10)	0.01 (0.07)	0.00	0.01 (0.10)	0.01 (0.10)	0.00
2016 未受(完)訪	0.01 (0.11)	0.01 (0.10)	0.00	0.01 (0.08)	0.00 (0.05)	-0.14
大五人格						
外向性 (E)	—	—	—	3.20 (1.03)	3.20 (1.03)	0.00
親和性 (A)	—	—	—	3.79 (0.71)	3.56 (0.71)	-0.32
開放性 (O)	—	—	—	3.13 (0.93)	3.42 (0.86)	0.32

1. 年齡、居住地區鄉村化程度、2016 年受訪時不耐煩程度與大五人格為平均數，其他則為比例，括號內為標準差。
2. 模型一、模型二由於分析樣本相同，變項統計量呈現於同一欄；模型三因額外納入人格特質作為共變項，分析樣本數較少，而以另一欄呈現。

## 二、調查模式參與的 logistic 迴歸模型分析結果

本研究依循 Rosenbaum and Rubin (1985) 建議，使用 logistic 迴歸模型分析產生處理變項的傾向分數（以網路模式完訪的機率）。根據表 3，模型一／二的分析結果顯示，教育程度愈高、有正式工作與已婚有偶的受訪者以網路模式（相對於面訪）完訪的機率比較高；而年齡較高、男性、每日接觸到人數越多、家中有其他受訪者、居住地區鄉村化程度越高、薪資大於 0 的受訪者，以網路模式完訪的機率較小。前一波（2016 年）調查曾表示拒訪或未受（完）訪，以及較沒有耐心的樣本，在 2018 年追蹤調查中以網路模式完訪的機率較低。新增人格特質變項後，模型三的分析結果顯示，個性較開放或親和性較低的受訪者，以網路模式完訪的機率比較高；外向性對於是否以網路模式完訪，則無顯著影響。相較模型一／二，在模型三中，顯著程度下降的共變項有工作狀態、薪資水準、2016 年訪問中途拒訪傾向與不耐煩程度等。

**表 3 以網調查完訪與否的 logistic 迴歸模型分析結果**

共變項	模型一	模型二	模型三
年齡	-0.0218*** (0.0062)	-0.0220*** (0.0062)	-0.0287*** (0.0073)
性別（男性）	-0.3964*** (0.0877)	-0.3811*** (0.0879)	-0.4634*** (0.1160)
教育程度（參照組：國中（或）以下）			
高中職或五專	1.4781** (0.4508)	1.4598** (0.4508)	1.2518* (0.5094)
大專院校	2.8888*** (0.4448)	2.8662*** (0.4449)	2.5142*** (0.5019)

表 3 以網調查完訪與否的 logistic 迴歸模型分析結果（續）

共變項	模型一	模型二	模型三
研究所（或）以上	3.3631*** (0.4548)	3.3347*** (0.4550)	2.9884*** (0.5175)
目前有無「正式」工作	1.1943** (0.4243)	1.1944** (0.4260)	0.9544† (0.5286)
家戶中有無其他的受訪者	-0.9754*** (0.1118)	-0.9814*** (0.1120)	-0.8732*** (0.1523)
居住地區鄉村化程度	-0.1389*** (0.0314)	-0.1443*** (0.0314)	-0.1734*** (0.0415)
婚姻狀態（參照組：從未結婚）			
已婚	0.2788** (0.0949)	0.2861** (0.0952)	0.3867** (0.1312)
分居、離婚或喪偶	-0.1792 (0.2939)	-0.1451 (0.2946)	0.0844 (0.3690)
工作薪資（參照組：沒有正式工作或 0 元）			
>0 且 ≤25K	-0.8672* (0.4280)	-0.8542* (0.4296)	-0.6659 (0.5357)
>25K 且 ≤50K	-0.6579 (0.4109)	-0.6443 (0.4127)	-0.5153 (0.5145)
>50K	-0.6951† (0.4179)	-0.6850 (0.4197)	-0.5239 (0.5258)
每日接觸人數（參照組：0-9 人）			
10-19 人	-0.6494*** (0.1105)	-0.6579*** (0.1109)	-0.6266*** (0.1454)
20-49 人	-1.1943*** (0.1229)	-1.2025*** (0.1233)	-1.1168*** (0.1629)
50 人（或）以上	-1.5572*** (0.1481)	-1.5727*** (0.1485)	-1.4702*** (0.1975)



表 3 以網調查完訪與否的 logistic 迴歸模型分析結果（續）

共變項	模型一	模型二	模型三
2016 年受訪時是否中途表示拒訪（參照組：沒有表示拒訪）			
曾表示拒訪	—	0.1473 (0.2612)	-0.1545 (0.3316)
2016 未受（完）訪	—	1.3375* (0.5943)	0.3152 (1.2082)
2016 年受訪時不耐煩			
	—	-0.3554** (0.1093)	-0.2441† (0.1306)
外向性（E）			
	—	—	-0.0504 (0.0563)
開放性（O）			
	—	—	0.2122** (0.0649)
親和性（A）			
	—	—	-0.2352** (0.0806)
常數項	-2.6925*** (0.5690)	-2.6159*** (0.5701)	-1.6788* (0.7441)
Log likelihood	-1,883.08	-1,876.61	-1,101.85
Pseudo R-squared	0.172	0.175	0.199
樣本數	6,201	6,201	4,002

1. 括號內的數值為估計標準誤（standard error）。

2. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.1$ 。

### 三、傾向分數配對後的樣本平衡性與模型敏感度分析

表 4、表 5 分別呈現半徑法、核函數法配對後樣本平衡性與各模型敏感度的分析結果。經「半徑法」配對後的各平衡性指標中，無論哪一種答題行為或哪一個模型的配對結果，logistic 迴歸模型的卡方值均不顯著（對應的  $p$  值在 0.9 至 1.0 之間），平均的標準化偏差都小於 5%（約在 0.9% 與 4.6% 之間），而 B 值均小於 25%（且模型三的

表 4 配對後共變項的平衡性與敏感度分析結果——半徑法

總體題項無回應	虛擬 判定 係數	logistic 迴歸 卡方值	標準化 偏差 (平均值)	標準化 偏差 (中位數)	B 值	R 值	敏感度 指標
模型一	0.002	3.58	0.9	0.7	9.7	0.46*	4.3(+)
模型二	0.002	3.46	0.9	0.9	9.5	0.47*	4.3(+)
模型三	0.002	2.03	1.1	1.1	9.6	0.68	4.3(+)
題項無回應—不知道							
模型一	0.002	3.58	0.9	0.7	9.7	0.46*	3.7(+)
模型二	0.002	3.46	0.9	0.9	9.5	0.47*	3.7(+)
模型三	0.002	2.03	1.1	1.1	9.6	0.68	3.6(+)
題項無回應—拒答							
模型一	0.002	3.58	0.9	0.7	9.7	0.46*	7.7(+)
模型二	0.002	3.46	0.9	0.9	9.5	0.47*	7.7(+)
模型三	0.002	2.03	1.1	1.1	9.6	0.68	7.0(+)
默許風格							
模型一	0.009	8.25	4.6	3.8	22.6	1.01	1.6(-)
模型二	0.010	9.4	4.3	3.3	22.3	0.56	1.6(-)
模型三	0.002	1.47	1.5	1.2	9.6	1.22	1.6(-)
極端回答風格							
模型一	0.009	8.25	4.6	3.8	22.6	1.01	1.0
模型二	0.010	9.40	4.3	3.3	22.3	0.56	1.0
模型三	0.002	1.47	1.5	1.2	9.6	1.22	1.0

1. \*：R<0.5 或 R>2.0。  
2. \*：B>25。  
3. (+)：模式效應被高估，(-)：模式效應被低估。

表 5 配對後共變項的平衡性與敏感度分析結果——核函數法

總體題項無回應	虛擬 判定 係數	logistic 迴歸 卡方值	標準化 偏差 (平均值)	標準化 偏差 (中位數)	B 值	R 值	敏感度 指標
模型一	0.002	4.96	1.1	0.6	11.4	0.40*	4.3(+)
模型二	0.002	4.98	1.1	0.8	11.4	0.41*	4.3(+)
模型三	0.002	2.89	1.2	1.1	11.4	0.57	4.3(+)
題項無回應—不知道							
模型一	0.002	4.96	1.1	0.6	11.4	0.40*	3.7(+)
模型二	0.002	4.98	1.1	0.8	11.4	0.41*	3.7(+)
模型三	0.002	2.89	1.2	1.1	11.4	0.57	3.6(+)
題項無回應—拒答							
模型一	0.002	4.96	1.1	0.6	11.4	0.40*	7.7(+)
模型二	0.002	4.98	1.1	0.8	11.4	0.41*	7.7(+)
模型三	0.002	2.89	1.2	1.1	11.4	0.57	7.0(+)
默許風格							
模型一	0.001	0.68	1.1	0.8	6.5	1.41	1.7(-)
模型二	0.001	0.94	1.3	1.3	7.6	1.50	1.7(-)
模型三	0.001	1.18	1.4	1.4	8.5	1.42	1.7(-)
極端回答風格							
模型一	0.001	0.68	1.1	0.8	6.5	1.41	1.0
模型二	0.001	0.94	1.3	1.3	7.6	1.50	1.0
模型三	0.001	1.18	1.4	1.4	8.5	1.42	1.0

1. \*：R&lt;0.5 或 R&gt;2.0。

2. \*：B&gt;25。

3. (+)：模式效應被高估，(-)：模式效應被低估。

B 值均小於 10%) (表 4)。但是, R 值僅在依據模型三進行配對的結果下符合 Rubin (2001) 所建議的臨界值標準。表 5「核函數法」的樣本平衡性分析結果與「半徑法」類似, logistic 迴歸模型的卡方值均不顯著, 標準化偏差都小於 5%, 而 B 值也都小於 12%。依模型一、二配對後的 R 值不夠理想 ( $R < 0.5$ ), 但依據模型三配對後的 R 值則顯示出理想的樣本平衡性。

表 4、表 5 最右一欄顯示模型敏感度分析的結果。首先, 總體題項無回應與「不知道」回答的模式效應分析, 在三個模型下, 無論採哪一種配對方式, 其模型穩定度均在可接受的範圍內 (敏感度指標的臨界值 5 以下) (Becker and Caliendo 2007)。雖然如此, 值得注意的是, 數值旁括號內的正號顯示還是有高估的嫌疑, 顯示所設定的模型均存在著不同程度的不穩定性, 亦即模型多少有些不夠適切, 表示納入的共變項還可做增減的調整。在拒答之模式效應方面, 三個模型的穩定度均不佳 (敏感度超過文獻建議的臨界值 5), 也存在高估的問題。

綜上, 三種題項無回應在配對所依據的模型中到底要包含哪些共變項, 仍有討論的空間。高估雖然表示可能需要刪除某些共變項, 但仍不能忽視有些重要的共變項可能沒考慮到。其次, 回答風格的 PSM 結果均顯示, 模型穩定性相對於題項無回應好很多。默許風格存在些許低估, 而極端回答風格的敏感度指標最為理想, 表示該結果變項的三個模型所納入的共變項都是適切的。

#### 四、問卷答題行為的模式效應

根據表 6, 不同問卷答題行為在不同模型的設定下, 經半徑法、核函數法所得到的平均處理效果 (ATE) 估計值, 均呈現顯著的模式效應。

表 6 比較面訪及網路自填問卷的調查模式效應

結果變項	配對前 平均值差異	模型一 (n=6,201)		模型二 (n=6,201)		模型三 (n=4,002)	
		半徑法 平均值差異	核函數法 平均值差異	半徑法 平均值差異	核函數法 平均值差異	半徑法 平均值差異	核函數法 平均值差異
總題項無回應	0.3559*** (0.0190)	0.3715*** (0.0250)	0.3683*** (0.0247)	0.3711*** (0.0260)	0.3678*** (0.0260)	0.3359*** (0.0440)	0.3361*** (0.0405)
配對後樣本數	—	6,198	6,198	6,197	6,197	3,994	3,994
題項無回應— 不知道	0.3339*** (0.0190)	0.3368*** (0.0292)	0.3328*** (0.0290)	0.3358*** (0.0301)	0.3319*** (0.0298)	0.2639*** (0.0556)	0.2708*** (0.0525)
配對後樣本數	—	6,198	6,198	6,197	6,197	3,994	3,994
題項無回應— 拒答	0.1998*** (0.0086)	0.2111*** (0.0247)	0.2116*** (0.0242)	0.2218*** (0.0258)	0.2215*** (0.0255)	0.2375*** (0.0377)	0.2234*** (0.0341)
配對後樣本數	—	6,198	6,198	6,197	6,197	3,994	3,994
默許風格	-0.0523*** (0.0100)	-0.0475*** (0.0125)	-0.0477*** (0.0131)	-0.0493*** (0.0125)	-0.0470*** (0.0131)	-0.0430** (0.0148)	-0.0435** (0.0142)
配對後樣本數	—	1,628	1,653	1,730	1,662	1,645	1,646
極端回答風格	0.0520*** (0.0128)	0.0457* (0.0193)	0.0466** (0.0177)	0.0396* (0.0185)	0.0429* (0.0175)	0.0433* (0.0199)	0.0439* (0.0182)
配對後樣本數	—	1,628	1,653	1,730	1,662	1,645	1,646

1. 括號內的數值為估計標準誤。

2. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ 。

3. 配對後樣本數：各模型實際分析的觀察值數，可能會因為「成功配對」（on-supported）樣本數或結果變項缺漏值的差異而有不同。

### （一）題項無回應的模式效應

首先，表 6 顯示，在傾向分數配對前，亦即未排除樣本選擇效應前，網路模式樣本的總體題項無回應的平均值較面訪樣本高出 0.3559

倍 ( $p < 0.001$ )。經兩種方法配對後，不論依據哪一個模型，ATE 估計結果均顯示，在控制樣本選擇效應後，總體題項無回應在網路模式樣本的平均值也都顯著高於面訪樣本 ( $p < 0.001$ )。不過，相較於配對前的平均值差異，依據模型一與模型二配對所估計的 ATE 均相對較大，依據模型三估計的 ATE 則相對較小。

進一步將題項無回應再細分為「不知道」回答與拒答，配對前的平均值差異也顯示，網路模式樣本在「不知道」回答與拒答的平均值均明顯高於面訪 ( $p < 0.001$ )。類似總體無回應的配對結果，相較於配對前的平均值差異，經兩種方法配對後，除了模型一與模型二的半徑法，模型三的半徑法及三個模型的核函數法所估得的 ATE，均比配對前的平均值差異為低，尤其模型三的兩種配對法估計的 ATE 更低 ( $p < 0.001$ )。相反的，在拒答方面，各模型兩種配對法的估計結果 (ATE) 均顯示，控制樣本選擇效應後的 ATE 估計值，均比配對前的平均值差異高，且模型三的兩種配對法估計的 ATE 更高 ( $p < 0.001$ )。此外，ATE 會隨著每增加一組概念層次的共變項而提高。

綜上，「不知道」回答與總體題項無回應的模式效應有類似的發現，在同一配對方法中，模型一與模型二的 ATE 非常接近，較配對前的平均值差異及模型三的 ATE 還要大。唯一的差異是模型三的兩種配對結果中，「不知道」回答的 ATE 遠低於配對前的平均值差異，但總體項目無回應的 ATE 與配對前的平均值差異幾乎一樣。此發現呼應 PSM 原理，增加共變項後的樣本平衡性應該會更好，調查模式的影響程度則會下降。另外，對於上述兩類題項無回應的模式效應估計來說，最理想的 logistic 迴歸模型設定是包含人格特質且通過所有平衡性指標的模型三，表示若忽略受訪者的人格特質，對配對結論會有影響。然而，拒答的分析結果則與前述總體題項無回應、「不知道」

回答的結果不同，無論採用的共變項組合、配對法為何，模型穩定度均偏低（敏感度指標超過 5），表示估計結果較不可信，共變項仍有調整的空間。

## （二）態度量表回答風格的模式效應

表 6 顯示，無論傾向分數配對前、後，網路模式樣本的默許風格顯著低於面訪樣本（ $p < 0.001$  或  $p < 0.01$ ），表示相較於網路模式，以面訪蒐集問卷資料時，受訪者較容易提供「同意」的回答。另外，以絕對值來看，無論哪一配對方法，三個模型的 ATE 較配對調整前的平均值差異來得小；不過，模型三配對估計的統計顯著性比模型一與二還要低，也是呼應共變項越多，樣本平衡性越好，模式效應顯著性下降的原理。不同於默許風格，在極端回答風格方面，無論有無經過傾向分數配對調整，相較於面訪，網路模式樣本更容易出現極端回答（「非常不同意」或「非常同意」）（ $p < 0.01$  或  $p < 0.05$ ）。此外，三個模型的 ATE 均比配對前的平均值差異為低。

綜上，傾向分數配對法的估計結果，並配合三種模型共變項的樣本平衡性與模型敏感度的分析結果，不論依據哪一種模型均可適切地利用 PSM 估計兩種回答風格的調查模式效應。不過，模型三還是最佳選擇，表示有必要納入人格特質。對於主觀態度題，兩種回答風格都存在顯著的模式效應。相較於由受訪者自行填答問卷的網路模式，在面訪由訪員提問的情況下，受訪者可能會為了迎合提問，而提供默許的答案，但較不會選擇極端的回答選項。

## 伍、結論

本研究運用家庭動態調查於 2018 年利用混合模式（面訪和網路自填問卷模式）所蒐集的調查資料，依據三組不同概念層次的共變項，設定了三個 logistic 迴歸模型，每一個模型採半徑法與核函數法兩種傾向分數配對。在三種模型設定、兩種配對法的組合下進行問卷答題行為的模式效應分析。所關注的問卷答題行為包含「題項無回應」與「態度題組回答風格」兩大類。本研究在進行傾向分數配對前，先確定處理變項與共變項的獨立性，在配對後則檢視傾向分數的重疊性、共變項的樣本平衡性與 logistic 模型的穩定性。結果顯示，在不同問卷答題行為的模式效應分析上，大多數的模型有通過平衡性、敏感度指標的考驗，對於回答風格來說，依據三種模型進行 PSM 模式效應分析是相當理想的。對總題項無回應與「不知道」回答則是以包含人格特質的模型三比較理想。

至於模式效應的分析結果，首先，在題項無回應方面，總體題項無回應、拒答與「不知道」回答似乎均存在顯著的模式效應，網路模式出現這三類題項無回應的可能性均高於面訪模式。不過，考慮到共變項的樣本平衡性與 logistic 迴歸模型的穩定度，依據模型三估計的總體題項無回應與「不知道」回答的模式效應比較可信賴。這兩類題項無回應的分析結果均符合作者的預期，也與 Heerwegh (2009)、Heerwegh and Loosveldt (2008) 的發現相近。我們的分析結果也顯示，儘管家庭動態調查的題項無回應、「不知道」回答不嚴重，面訪、網路模式之間仍存在明顯的差異，表示網路模式的題項無回應問題仍有改進的空間。



其次，態度量表回答風格的分析結果顯示，在默許與極端回答風格具有顯著的模式效應。相較於網路模式，面訪模式較容易出現「同意」回答，但較不容易出現極端回答（非常不同意或非常同意）。容易出現「同意」回答的分析結果符合預期，也與 Liu et al. (2017) 的發現相近；不容易出現極端回答則與預期不符，既不同於現有文獻的論點，也與 Liu et al. (2017) 的發現相左。對於面訪模式較不容易出現極端回答的結果，我們在此提出可能的解釋。既有文獻指出，相較集體主義（collectivism）社會，在個人主義（individualism）社會中，受訪者在面對問卷量表時，較可能選擇極端答案（Chen et al. 1995）。Chen et al. (1995) 即發現，所分析的學生樣本中，來自臺灣、日本的學生，相較美國、加拿大來的學生，較不傾向選擇量表中的極端選項。以本研究分析的混合模式而言，在面訪的情境下，基於集體主義的社會環境使然，受訪者可能會顧慮訪員的想法，而不傾向選擇極端答案。相對的，在網路自填問卷的情況下，由於不需要顧慮他人的想法，較可能選擇極端答案。這與本研究發現面訪模式較容易出現「同意」的默許答案，究其背後，都同樣可能受到文化背景因素的左右。

本研究除了「拒答」的分析結果較不確定之外，其他四種問卷答題行為均存在明顯的調查模式效應，以不同模式完訪的受訪者在總體題項無回應、「不知道」回答、默許風格與極端回答風格上，均有明顯的差異，表示須謹慎進行兩種模式資料的整合分析（Cernat and Revilla 2021）。對已經蒐集完成的調查資料，能解決題項無回應的方法有限（如刪除、插補），但應小心處理。在進行態度量表的資料分析時，應考慮不同模式的回答風格差異會造成分析結果估計的誤差。

為減低混合模式調查的模式效應問題，在這類調查的事前規劃上，本研究的分析結果具有多項實務意涵。首先，在調查成本的限制

下，對於混合模式調查，可藉由調查模式組合、樣本分配原則，以及混合模式的調查執行方式（如樣本轉換模式的時機與原則）等方面的規劃，盡量降低模式效應。其次，在問卷的設計上，對於可能產生模式效應的題目或題組，可藉由問卷題目說明、選項的調整，以減少模式效應。再者，對敏感或有社會期許（social desirability）顧慮的題目，在面訪模式下，可採用受訪者自行填答的方式，以減少訪員在受訪者作答上的影響。

此外，在網路自填問卷中，可適度提供專注作答的警醒或誘因機制，以減少敷衍作答或不專注的情況。以家庭動態調查而言，未來如果繼續採行面訪、網路自填問卷的混合模式，對於以面訪進行問卷資料蒐集的樣本，除了原本就以受訪者自填方式進行的一些敏感題項之外，可考慮適度納入主觀態度題組，以減少訪員的影響效果。另一方面，對網路自填問卷，提供給受訪者的提示與指引可適度增加，對部分題項亦可在受訪者無法提供答案的情況下增加追問機制，並逐題考量是否放置不知道、拒答等選項，以讓網路模式樣本在作答時面對的情境，與面訪有訪員提示、追問的情境相類似。

儘管本研究具有前述學理、實務上的意涵與重要性，仍有其限制存在，以下從兩個面向來討論。在研究的主題及架構方面，首先，由於家庭動態調查的問卷內容以家庭為主，敏感題目不多，在模式效應的分析結果上，可能不同於議題敏感的調查。如果有政治議題、犯罪議題等主題敏感的混合模式調查資料可進行分析，當有助於瞭解這類調查在問卷答題結果上的模式效應。第二，本研究探討的題項無回應，以整體問卷的題項無回應為主，對特定敏感性題目（如收入、政黨傾向、投票行為、犯罪行為）的回答結果或題項無回應，其模式效應仍有待探究。第三，本研究討論的態度題組的回答風格包括默許、

極端回答兩種，並未擴及其他作答行為，例如中立回答風格（middle response style）、直線回答（straight-lining）、初始效應（primacy effect）、新近效應（recency effect）等問題，<sup>8</sup> 相關課題仍有待進一步研究。第四，本研究主要探討問卷答題行為受調查模式本身的影響，而非其他變項或這些變項與調查模式交互作用的影響效果。對於其他變項的可能影響及其與調查模式的交互影響，迄今為止，在題項無回應、態度題組回答風格方面，仍未見相關研究。針對主觀福祉的模式效應，則有少數研究探討人口特徵（如性別、年齡與教育程度）與模式之交互作用（Holford and Pudney 2015; Piccitto et al. 2022）。在這些文獻的基礎上，未來可針對題項無回應、態度題組回答風格兩類答題行為，進一步探究人口特徵及其與模式交互作用的影響。最後，本研究以臺灣蒐集到的混合模式調查資料進行分析，如果將來有不同社會蒐集到的混合模式調查資料可進行比較研究，當有助於進一步解答模式效應是否存在跨文化的差異。

在 PSM 的應用方面，影響模式效應估計準確性的因素很多，本研究的分析雖已對 PSM 的前提假設做了嚴謹的檢測，但配對前後的執行過程仍有值得檢討之處。首先，如前所述，本研究並未深入探討模式效應之外其他形式的影響關係，如中介效果與調節效果。中介效果已在篩選本研究架構下的共變項時被排除，不過，尚有其他共變項（含被排除的共變項）仍可能具有中介的影響效果。由於不含交互項的共變項已符合 PSM 要求的獨立性，本研究並未考慮調節效果。雖然在題項無回應、態度題組答題風格方面尚無文獻可供參考，如前所

---

8 中立回答風格是指受訪者傾向選擇中間選項（詳 Moors 2008），而直線作答是指傾向選擇同一選項的問題（詳 Zhang and Conrad 2014）。初始效應、新近效應則分別指第一個選項、最後一個選項被選擇的可能性較高的問題（參見 Krosnick and Alwin 1987）。

述，未來可參考主觀福祉的模式效應研究，在進行模式效應分析時納入交互項，尤其是調查模式與某些人口特徵（如年齡與教育程度）或人格特質（如本研究發現對模式決定具有顯著影響的親和性與開放性人格特質）可能的交互作用，會是值得探討的方向。

第二，在應用 PSM 分析時，相關研究在進行分析資料處理時，多將資料處理為沒有任何缺漏值的完整資料。本研究也依循此一原則，盡力藉由前一波資料進行一些重要變項資料的回填，包含社會人口基本資料；但部分變項（如人格特質與主觀態度題目）資料則因不適合、無所依據等原因而無法回填或插補。在進行模型三的傾向分數配對時，只好採 listwise deletion，致使樣本數大量減少，殊為可惜。不過，也因此衍生一些問題，諸如缺漏資料到底需不需要回填或插補？僅採 pairwise deletion 可以嗎？若需要插補，又該如何插補？根據作者有限的瞭解，採 pairwise deletion 跟 listwise deletion 的結果是一樣的。由於被刪除的樣本資料無法納入 PSM 中，會造成有效樣本的損失，若採取不替代的配對原則（without replacement）會進一步降低有效的配對結果，PSM 也因此受到 King and Nielson（2019）的批評。不過，近期研究已針對 PSM 分析資料的缺漏值處理方法進行探討。本研究對薪資缺漏值採多重插補處理，即為相關文獻建議的方式之一，儘管如此，未來仍可嘗試其他的插補方法（Bottigliengo et al. 2021b; Choi et al. 2019）。

第三，本研究採用的配對方法是擴大取樣的一對多（或稱多對一）配對，並以可替代配對為原則，以期增加可配對數與共變項樣本的平衡性，並降低樣本自我選擇誤差。不過，擴大取樣仍有可能提高相對誤差（誤差相對於估計變異）與處理效應的估計變異，或面臨無法配對的風險（Bottigliengo et al. 2021a）。在本研究採用的兩種配對方法之外，full matching 也許是可進一步嘗試的方法。Full matching

是將傾向分數分群，且考慮各分群內目標組每一個案與對照組每一個案之間距離的平均加權後，進行傾向分數配對的方法（Bottigliengo et al. 2021b; Rosenbaum 2002）。

第四，文獻對 PSM 的一項批評是，配對結果可能產生不完整配對（incomplete matching）的問題（Bottigliengo et al. 2021a; Bottigliengo et al. 2021b）。由於本研究的無效配對數不多，前述問題並不嚴重。不過，如何提高傾向分數跨兩組之樣本重疊性，以及有效配對數，藉以提高估計結果可涵蓋的母體範圍與樣本代表性，仍是未來需要繼續研討的議題。最後，本研究針對題項無回應的模式效應進行 PSM 分析時，依據模型一與模型二得到的 R 值不在可接受範圍內，「拒答」的 PSM 結果則顯示三個模型的穩定度不佳，顯示共變項仍有調整的空間。這也表示在進行不同結果變項的模式效應 PSM 分析時，可因應結果變項的差異採用不同的共變項，而非採用同樣的共變項。

總結來說，儘管有以上的研究限制與未來研究方向建議，由於本研究使用的樣本數不小，在審慎選擇適當的一對多配對方法並採用可替代配對的原則下，具有可減低有效樣本損失與無效配對的優勢，利於提高配對品質。在 PSM 的應用上，對於 King and Nielson（2019）指出若採一對一不可替代配對會大量損失有效樣本，造成大量無效配對的批評，本研究的分析顯示，若能因應樣本特性調整 PSM 的配對方法，可緩解無效配對的問題。在調查方法層面，本研究以 PSM 方法估計混合模式調查在答題行為上的模式效應，並與既有研究的發現進行比較，具有學理上的貢獻。在調查實務上，本研究對採用混合模式的追蹤調查計畫，對於如何減緩模式效應，提供問卷設計、調查規劃等方面的建議，兼具實務方面的貢獻。

## 附錄 1 用於檢視「默許風格」與「極端回答風格」 的平衡型量表題組（RR2018 問卷）

H2. 以下是一些有關婚姻態度的題目，請問您同不同意下列看法？

（1 表示「非常不同意」；5 表示「非常同意」）

1. 結婚後不一定要有小孩
2. 一對情侶住在一起，即使沒有結婚的打算也沒什麼關係
3. 就算是不好的婚姻，還是比單身來得好
4. 就算沒有結婚，還是可以生小孩
5. 只有少數的婚姻是幸福的
6. 結婚的人比離婚的人快樂
7. 結婚的人比沒結過婚的人快樂

H3. 以下是有關女性、家庭和工作的一些說法，請問您的看法是同意或不同意？（1 表示「非常不同意」；5 表示「非常同意」）

1. 職業婦女和家庭主婦一樣可以和她的小孩建立溫暖且安穩的關係
2. 如果母親有工作，對還沒上小學的小孩可能不好
3. 整體而言，妻子有份全職的工作對家庭生活有不好的影響
4. 有個工作也不錯，但大部分的女人真正想要的還是一個家和小孩
5. 當個家庭主婦與有一份有收入的工作，一樣令人覺得充實
6. 女人要獨立的最好方式就是有份工作
7. 如果夫妻都有工作，應該要平均分擔家事
8. 男人不出外工作，料理家務也很適合
9. 在經濟不景氣時，女性員工應比男性員工先被解雇

## 附錄 2 大五人格量表題項 (RR2018、RR2009)

您認為下面這些特徵符不符合您自己的個性？(「1」表示很符合；「5」表示很不符合)

- a. 不太愛說話 (E)
- b. 很有同情心 (A)
- c. 事情做到完為止 (C)
- d. 懂得放鬆而且會處理壓力 (N)
- e. 對多花腦筋思考的事物很有興趣 (O)
- f. 外向、會和人交際 (E)
- g. 比較會挑別人毛病 (A)
- h. 粗心大意 (C)
- i. 容易緊張 (N)
- j. 藝術方面的興趣比較少 (O)

※反向題以底線標示

依個別題項意義可區分成：「外向性」(extraversion)，以 E 表示；「親和性」(agreeableness)，以 A 表示；「謹慎性」(conscientiousness)，以 C 表示；「神經質傾向」(neuroticism)，以 N 表示；「開放性」(openness)，以 O 表示。

附錄 3 各變項的操作性定義

變項名稱	操作性定義
處理變項：調查模式	
	RR2018 問卷中以「網路自填問卷模式」完訪者，視為目標組，編碼為 1；以面訪完訪者，則歸為對照組，編碼為 0。
結果變項	
總體題項無回應	在 RR2018 問卷的必答題項（無法跳答的題項）中，若受訪者有一個或一個以上的題項回答「不知道」或「拒答」，編碼為 1；沒有任何題項無回應則為 0。
題項無回應——不知道	在 RR2018 問卷的必答題項（無法跳答的題項）中，若受訪者有一個或一個以上的題項回答「不知道」，編碼為 1；沒有「不知道」回答者則為 0。
題項無回應——拒答	在 RR2018 問卷的必答題項（無法跳答的題項）中，若受訪者有一個或一個以上的題項「拒答」；編碼為 1；沒有「拒答」者則為 0。
默許風格 （僅 RCI / RI2016 樣本群回答）	選取 RR2018 問卷中的兩個平衡型量表：H2 題組（婚姻態度 7 題）和 H3 題組（性別角色 9 題），以受訪者對這 16 題回答「同意」的題數比例作為測量。例如，若 16 題中有 2 題回答同意，其數值為 0.125（=2/16）。數值愈大，表示默許風格愈強。相關題項內容可參見附錄 2。
極端回答風格 （僅 RCI / RI2016 樣本群回答）	選取 RR2018 問卷中的兩個平衡型量表：H2 題組（婚姻態度 7 題）和 H3 題組（性別角色 9 題），以受訪者對這 16 題回答「非常同意」或「非常不同意」的題數比例作為測量。例如，若 16 題中有 2 題回答「非常不同意」、3 題回答「非常同意」，其數值為 0.3125（=5/16）。數值愈大表示極端回答風格愈強。相關題項內容可參見附錄 2。



### 附錄 3 各變項的操作性定義（續）

變項名稱	操作性定義
共變項	
年齡	自受訪者所回答的出生年份轉換而來。
教育程度	依受訪者所回答的「最高教育程度」（含畢業、肄業或就學中），區分為四類：「國中（或）以下」、「高中職或五專」、「大專院校」及「研究所（或）以上」。
家戶中有無其他的受訪者	依據「受訪者編號」的前 6 碼（戶號）、第 7 碼（戶內成員編號），判斷同一家戶是否有其他成員受訪，若有，則編碼為 1，否則為 0。
目前有無「正式」工作	依問卷題目「請問您目前有工作嗎」的回答結果，若「從事有收入的工作」或「從事無酬家屬工作，而且每週工作 15 小時以上，或每天工作 3 小時以上」編碼為 1，表示有「正式」工作；「從事無酬家屬工作，而且每週工作 15 小時以下，或每天工作 3 小時以下」或「沒有工作」，則編碼為 0，表示無「正式」工作。
婚姻狀態	依受訪者所回答的目前婚姻狀態，區分為三類：「從未結婚」、「已婚」和「分居、離婚或喪偶」。
工作薪資	將受訪所回答的工作薪資，分成四個類別：「沒有正式工作 或 0 元」、「>0 且 ≤ 25K」、「>25K 且 ≤ 50K」及「>50K」。
每日接觸人數	依據問卷題目「您平常一天裡面，大概跟多少人接觸」的回答結果，重新分為四個類別：「0-9 人」、「10-19 人」、「20-49 人」及「50 人（或）以上」。

附錄 3 各變項的操作性定義（續）

變項名稱	操作性定義
居住地區鄉村化程度	<p>依據侯佩君等（2008）所歸類出臺灣的鄉鎮市區發展分層類別，將受訪者的現居地區分成「都會核心」、「工商或新興市鎮」、「傳統產業市鎮」、「低度發展鄉鎮」、「高齡化鄉鎮」或「偏遠鄉鎮」等七個層級，編碼為 1 到 7，以分數越高，鄉村化程度越高表示。</p> <p>若現居地位於離島地區（連江或金門），則歸入「偏遠鄉鎮」此一層級之中；居住於外國地區者，則不會納入分析樣本之中。</p>
2016 年受訪時是否中途表示拒訪	取自 2016 年調查中訪員紀錄的訪問狀況資訊，經分類後包含：「沒有表示拒訪」、「曾表示拒訪」及「2016 年未受訪或未完訪」。
2016 年受訪時不耐煩	取自 2016 年調查中訪員紀錄的訪問狀況資訊，受訪時表現的不耐煩程度由低至高，分別編碼為「從未=0」、「很少=1」、「有時=2」與「經常=3」；若受訪者在 2016 年未受訪或未完訪，則另編碼為 4，視為極為不耐煩。
外向性（E） 親和性（A） 開放性（O）	<p>此三面向的人格特質取自 RR2018 年問卷中的「大五人格」量表。由於此題組僅限 RCI2016 和 RI2016 的受訪者回答，其他未作答的樣本群以 RR2009 問卷的資訊進行回填插補。但 2009 年的新抽樣主樣本群（RI2009）資訊則仍缺漏。</p> <p>在變項的測量上，是先將反向題反向計分，再將各面向的兩題分數取其平均值。量表題項可參見附錄 1。</p>

## 附錄 4 各結果變項的描述統計

結果變項	模型一／模型二		模型三	
	平均值	樣本數	平均值	樣本數
總體題項無回應		<b>6,201</b>		<b>4,002</b>
沒有	51.75 %		49.95 %	
有一題（或）以上	48.25 %		50.05 %	
題項無回應——不知道		<b>6,201</b>		<b>4,002</b>
沒有	53.44 %			
有一題（或）以上	46.56 %		48.33 %	
題項無回應——拒答		<b>6,201</b>		<b>4,002</b>
沒有	94.47 %		94.25 %	
有一題（或）以上	5.53 %		5.75 %	
默許風格	0.31 (0.16)	<b>1,762</b>	0.31 (0.16)	<b>1,759</b>
極端回答風格	0.20 (0.21)	<b>1,762</b>	0.20 (0.21)	<b>1,759</b>

1. 表中括號內的數值為標準差。
2. 模型一、模型二由於分析樣本相同，變項統計量呈現於同一欄；模型三因額外納入人格特質作為共變項，分析樣本數較少，而以另一欄呈現。

附錄 5 處理變項與共變項的描述統計  
(缺漏資料回填後)

	模型一／模型二	模型三
處理變項：(網路模式=1)	12.00%	10.87%
共變項：		
年齡	<b>42.07 (14.55)</b>	<b>44.67 (16.82)</b>
性別 (男性=1；女性=0)	52.75%	52.05%
教育程度		
國 (初) 中或以下	15.95%	21.66%
高中職或五專	24.77%	24.11%
大專院校	46.17%	42.63%
研究所 (或) 以上	13.11%	11.59%
家戶中有無其他的受訪者	30.41%	30.38%
目前有無「正式」工作 (有=1)	78.15%	73.09%
婚姻狀態		
從未結婚	37.78%	37.71%
已婚	53.44%	51.65%
分居、離婚或喪偶	8.77%	10.64%
工作薪資		
沒有正式工作或 0	23.01%	28.31%
>0 且 ≤25K	12.51%	12.82%
>25K 且 ≤50K	42.56%	39.91%
>50K	21.82%	18.97%

## 附錄 5 處理變項與共變項的描述統計 (缺漏資料回填後)(續)

	模型一／模型二	模型三
每日接觸人數		
0-9 人	22.84%	24.86%
10-19 人	31.75%	32.21%
20-49 人	27.08%	25.86%
50 人以上	18.34%	17.07%
居住地區鄉村化程度	<b>3.19 (1.51)</b>	<b>3.25 (1.53)</b>
都會核心	10.48%	10.72%
工商市鎮	30.45%	28.89%
新興市鎮	23.61%	22.19%
傳統產業市鎮	9.30%	10.59%
低度發展鄉鎮	17.13%	18.42%
高齡化鄉鎮	8.89%	9.10%
偏遠鄉鎮	0.15%	0.10%
2016 年受訪時是否中途表示拒訪		
沒有表示拒訪	94.57%	94.45%
曾表示拒訪	4.23%	4.93%
2016 未受（完）訪	1.21%	0.62%
2016 年受訪時不耐煩（0-4）	<b>0.23 (0.68)</b>	<b>0.24 (0.64)</b>
從未	86.34%	84.63%
很少	7.60%	9.25%
有時	3.77%	4.45%
經常	1.00%	1.05%
2016 未受（完）訪	1.29%	0.62%

附錄 5 處理變項與共變項的描述統計  
(缺漏資料回填後)(續)

	模型一／模型二	模型三
大五人格		
外向性 (E)	—	<b>3.20 (1.03)</b>
親和性 (A)	—	<b>3.77 (0.71)</b>
開放性 (O)	—	<b>3.16 (0.93)</b>
樣本數	<b>6,201</b>	<b>4,002</b>

1. 年齡、居住地區鄉村化程度、2016 年受訪時不耐煩程度與三面向人格變項，表中呈現的是平均數，括號內則為標準差。
2. 模型一、模型二由於分析樣本相同，變項統計量呈現於同一欄；模型三因額外納入人格特質作為共變項，分析樣本數較少，而以另一欄呈現。

## 參考文獻

- 于若蓉, 2019, 家庭動態資料庫的建立: 第十七年計畫 (RR2018) (C00333\_1), 取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫。(Yu Ruoh-Rong, 2019, Panel Study of Family Dynamics: RR2018 (C00333\_1) [data file]. Available from Survey Research Data Archive, Academia Sinica. [https://doi.org/10.6141/TW-SRDA-C00333\\_1-1](https://doi.org/10.6141/TW-SRDA-C00333_1-1))
- 于若蓉、黃奕嘉, 2018, 〈家庭動態調查: 樣本結構、問卷內容、資料外釋與應用〉。《中國統計學報》56(4): 98-115。(Yu, Ruoh-Rong, and I-Chia Huang, 2018, “The Panel Study of Family Dynamics: Samples, Questionnaires, Data Dissemination and Applications.” *Journal of the Chinese Statistical Association* 56(4): 98-115.)
- 侯佩君、杜素豪、廖培珊、洪永泰、章英華, 2008, 〈台灣鄉鎮市區類型之研究: 『台灣社會變遷基本調查』第五期計畫之抽樣分層效果分析〉。《調查研究—方法與應用》23: 7-32。(Hou, Pei-Chun, Su-Hao Tu, Pei-Shan Liao, Yung-Tai Hung, and Ying-Wha Chang, 2008, “Typology of Townships in Taiwan: The Analysis of Sampling Stratification of the 2005-2006 ‘Taiwan Social Change Survey’.” *Survey Research—Method and Application* 23: 7-32.)
- Aakvik, Arild, 2001, “Bounding a matching estimator: the case of a Norwegian training program.” *Oxford Bulletin of Economics and Statistics* 63(1): 115-143.
- Austin, Perter C., 2009, “Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo Simulations.” *Biometrical Journal* 59(1): 171-184.
- , 2011, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.” *Multivariate Behavioral Research* 46(3): 399-424.
- , 2013, “The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios.” *Statistics in Medicine* 32(16): 2837-2849.
- , 2014, “A Comparison of 12 Algorithms for Matching on the Propensity Score.” *Statistics in Medicine* 33(6): 1057-1069.
- Baser, Onur, 2006, “Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching.” *Value in Health* 9(6): 377-385.
- Becker, Sascha O., and Marco Caliendo, 2007, “Sensitivity Analysis for Average Treatment Effects.” *The Stata Journal* 7(1): 71-83.
- Berhman, J. R., Y. Cheng, and P. E. Todd, 2004, “Evaluating Preschool Programs When

- Length of Exposure to the Program Varies: A Nonparametric Approach.” *Review of Economics and Statistics* 86(1): 108–132.
- Biemer, Paul P., 2001, “Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing.” *Journal of Official Statistics* 17(2): 295–320.
- Biemer, Paul P., and Lars E. Lyberg, 2003, *Introduction to Survey Quality*. NJ: John Wiley & Sons.
- Börkan, Bengü, 2010, “The Mode Effect in Mixed-mode Surveys: Mail and Web Surveys.” *Social Science Computer Review* 28(3): 371–380.
- Bottigliengo, D., G. Lorenzoni, H. Ocagli, M. Martinato, P. Berchialla, and D. Gregori, 2021b, “Propensity Score Analysis with Partially Observed Baseline Covariates: A Practical Comparison of Methods for Handling Missing Data.” *International Journal of Environmental Research and Public Health* 18(6694): 1–17.
- Bottigliengo, D., I. Bald, C. Lanera, G. Lorenzoni, J. Bejko, T. Bottio, V. Tarzia, M. Carrozzini, G. Gerosa, P. Berchialla, and D. Gregori, 2021a, “Oversampling and Replacement Strategies in Propensity Score Matching: A Critical Review Focused on Small Sample Size in Clinical Settings.” *BMC Medical Research Methodology* 21(256): 1–16.
- Caliendo, Macro, and Sabine Kopeinig, 2008, “Some Practical Guidance for the Implementation of Propensity Score Matching.” *Journal of Economic Surveys* 22(1): 31–72.
- Cernat, Alexandru, and Melanie Revilla, 2021, “Moving from Face-to-face to a Web Panel: Impacts on Measurement Quality.” *Journal of Survey Statistics and Methodology* 9(4): 745–763.
- Chen, Chuansheng, Shin-ying Lee, and Harold W. Stevenson, 1995, “Response Style and Cross- cultural Comparisons of Rating Scales among East Asian and North American Students.” *Psychological Science* 6(3): 170–175.
- Choi, J., O. M. Dekkers, and S. L. Cessie, 2019, “A Comparison of Different Methods to Handle Missing Data in the Context of Propensity Score Analysis.” *European Journal of Epidemiology* 34: 23–36.
- Cochran, W. G., and D. B. Rubin, 1973, “Controlling Bias in Observational Studies: A Review.” *Sankhya Series A* 35: 417–446.
- Datta, P., K. Walsh, and D. Terrell, 2002, “The Impact Demographics on Choice of Survey Modes.” *Communications of the Association for Information Systems* 9(13): 223–240.
- De Leeuw, Edith D., 2005, “To Mix or Not to Mix Data Collection Modes in Surveys.” *Journal of Official Statistics* 21(5): 233–255.
- , 2018, “Mixed-mode: Past, Present, and Future.” *Survey Research Methods* 12(2): 75–89.



- De Leeuw, Edith D., Don A. Dillman, and Joop J. Hox, 2008, "Mixed Mode Surveys: When and Why." Pp. 299-316 in *International Handbook of Survey Methodology*, edited by Edith D. De Leeuw, Don A. Dillman, and Joop J. Hox. NY: Lawrence Erlbaum.
- Díaz de Rada, Vidal, 2022, "Concurrent Mixed Modes: Response Quality, Speed, and Cost." *Field Methods* 34(3): 191-205.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian, 2009, *Internet, Mail, and Mixed- Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan, 2005, "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69(3): 370-392.
- Fowler, Floyd Jackson, P. M. Gallagher, V. L. Stringfellow, A. M. Zaslavsky, J. W. Thompson, and P. D. Cleary, 2002, "Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members." *Medical Care* 40(3): 190-200.
- Galdo, Jose C., Jeffrey Smith, and Dan Black, 2008, "Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data." *Annales d'Économie et de Statistique* 91(92): 189-216.
- Garrido, M. M., A. S. Kelley, J. Paris, K. Roza, D. E. Meier, R. S. Morrison, and M. D. Aldridge, 2014, "Methods for Constructing and Assessing Propensity Scores." *Health Research and Educational Trust* 49(5): 1701-1718.
- Guo, Shenyang, and Mark W. Fraser, 2015, *Propensity Score Analysis* (2<sup>nd</sup> edition). LA: Sage.
- Harris, Heather, and S. Jeanne Horst, 2016, "A Brief Guide to Decisions at Each Step of the Propensity Score Matching Process." *Practical Assessment, Research, and Evaluation* 21(4): 1-11.
- Hazlett, Chad, 2020, "Kernel Balancing: A Flexible Non-parametric Weighting Procedure for Estimating Causal Effects." *Statistica Sinica* 30: 1155-1189.
- Heerwegh, Dirk, 2009, "Mode Differences Between Face-to-face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects." *International Journal of Public Opinion Research* 21(1): 111-121.
- Heerwegh, Dirk, and Geert Loosveldt, 2008, "Face-to-face versus Web Surveying in a High-internet-coverage Population: Differences in Response Quality." *Public Opinion Quarterly* 72(5): 836-846.
- , 2011, "Assessing Mode Effects in a National Crime Victimization Survey Using Structural Equation Models: Social Desirability Bias and Acquiescence." *Journal of Official Statistics* 27(1): 49-63.
- Hibbing, Matthew V., Matthew Cawvey, Raman Deol, Andrew J. Bloeser., and Jeffery J. Mondak, 2019, "The Relationship between Personality and Response Patterns on Public

- Opinion Surveys: The Big Five, Extreme Response Style, and Acquiescence Response Style.” *International Journal of Public Opinion Research* 31(1): 161-177.
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick, 2003, “Telephone versus Face-to-face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias.” *Public Opinion Quarterly* 67(1): 79-125.
- Holford, Angus J., and Stephen Pudney, 2015, “Survey Design and the Determinants of Subjective Wellbeing: An Experimental Analysis.” *IZA Discussion Papers, No. 8760*. Bonn: Institute for the Study of Labor (IZA).
- Huber, M., M. Lechner, and A. Steinmayr, 2015, “Radius Matching on the Propensity Score with Bias Adjustment: Tuning Parameters and Finite Sample Behavior.” *Empirical Economics* 49: 1-31.
- Jäckle, Annette, Caroline Roberts, and Peter Lynn, 2010, “Assessing the Effect of Data Collection Mode on Measurement.” *International Statistical Review* 78(1): 3-20.
- Jiang, W., L. Ha, M. Abuljadail, and S. A. Alsulaiman, 2017, “Item Non-response of Different Question Types and Formats in Mixed-mode Surveys: A Case Study of a Public Broadcasting TV Station’s Members.” *Journal of Communication and Media Research* 9(1): 173-184.
- Jones, M. K., L. Calzavara, D. Allman, C. A. Worthington, M. Tyndall, and J. Iveniuk, 2016, “A Comparison of Web and Telephone Responses from a National HIV and AIDS Survey.” *JMIR Public Health Surveill* 2(2): 1-15.
- King, Gary, and Richard Nielson, 2019, “Why Propensity Scores Should Not Be Used for Matching.” *Political Analysis* 27(4): 435-454.
- Klausch, L. Thomas, Barry Schouten, and Joop J. Hox, 2014, “The Use of Within-subject Experiments for Estimating Measurement Effects in Mixed-mode Surveys.” *Discussion Paper*. Statistics Netherlands.
- , 2017, “Evaluating Bias of Sequential Mixed-mode Designs Against Benchmark Surveys.” *Sociological Methods & Research* 46(3): 456-489.
- Kohler, U., K. B. Karlson, and A. Holm, 2011, Comparing Coefficients of Nested Nonlinear Probability Models. *The Stata Journal* 11(3): 420-438.
- Kreuter, Frauke, Gerrit Müller, and Mark Trappmann, 2010, “Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data.” *Public Opinion Quarterly* 74(5): 880-906.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau, 2008, “Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity.” *Public*

- Opinion Quarterly* 72(5): 847–865.
- Krosnick, Jon A., 1991, “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology* 5(3): 213–236.
- Krosnick, Jon A., and Duane F. Alwin, 1987, “An Evaluation of a Cognitive Theory of Response-order Effects in Survey Measurement.” *Public Opinion Quarterly* 51(2): 201–219.
- Link, Michael W., and Ali Mokdad, 2006, “Can Web and Mail Survey Modes Improve Participation in an RDD-based National Health Surveillance?” *Journal of Official Statistics* 22(2): 293–312.
- Liu, Mingnan, Frederick G. Conrad, and Sunghee Lee, 2017, “Comparing Acquiescent and Extreme Response Styles in Face-to-face and Web Surveys.” *Quality & Quantity* 51: 941–958.
- Lugtig, Peter, Gerty J. L. M. Lensvelt-Mulders, Remco Frerichs, and Assyn Greven, 2011, “Estimating Nonresponse Bias and Mode Effects in a Mixed-mode Survey.” *International Journal of Market Research* 53: 669–686.
- Lynn, Peter, 2013, “Alternative Sequential Mixed-mode Designs: Effects on Attrition Rates, Attrition Bias, and Costs.” *Journal of Survey Statistics and Methodology* 1(2): 183–205.
- Mariano, Louis T., and Marc N. Elliott, 2017, “An Item Response Theory Approach to Estimating Survey Mode Effects: Analysis of Data from a Randomized Mode Experiment.” *Journal of Survey Statistics and Methodology* 5(2): 233–253.
- Moors, Guy, 2008, “Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement.” *Quality & Quantity* 42: 779–794.
- Pearl, Judea, 1995, “Causal Diagrams for Empirical Research.” *Biometrika* 82: 669–688.
- , 2009, *Causality: Models, Reasoning and Inference* (2<sup>nd</sup> edition). NY: Cambridge University Press.
- Piccitto, G., A. C. Liefbroer, and T. Emery, 2022, “Does the Survey Mode Affect the Association between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode.” *Journal of Happiness Studies* 23: 3441–3461. <https://doi.org/10.1007/s10902-022-00553-y>.
- Ripollone, J. E., K. F. Huybrechts, K. J. Rothman, R. E. Ferguson, and J. M. Franklin, 2018, “Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology.” *American Journal of Epidemiology* 187(9): 1951–1961.
- Rosenbaum, Paul R., 2002, *Observational Studies* (2<sup>nd</sup> edition). NY: Springer.
- , 2005, “Observational Study.” Pp. 1451–1462 in *Encyclopedia of Statistics in Behavioral Science* (Vol. 3), edited by Brian S. Everitt, and David C. Howell. NJ: 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.
- , 1985, "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39(1): 33-38.
- Rubin, Donald B., 1974, "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies." *Journal of Educational Psychology* 66: 688-701.
- , 2001, "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services & Outcomes Research Methodology* 2(3): 169-188.
- Sakshaug, Joseph W., Ting Yan, and Roger Tourangeau, 2010, "Nonresponse Error, Measurement Error, and Mode of Data Collection: Tradeoffs in a Multi-mode Survey of Sensitive and Non-sensitive Items." *Public Opinion Quarterly* 74(5): 907-933.
- Schouten, Barry, Jan van den Brakel, Bart Buelens, Jan van der Laan, and Thomas Klausch, 2013, "Disentangling Mode-specific Selection and Measurement Bias in Social Surveys." *Social Science Research* 42(6): 1555-1570.
- Valentino, Nicholas A., Kirill Zhirkov, D. Sunshine Hillygus, and Brian Guay, 2020, "The Consequences of Personality Biases in Online Panels for Measuring Public Opinion." *Public Opinion Quarterly* 84(2): 446-468.
- Vannieuwenhuyze, Jorre T. A., and Geert Loosveldt, 2013, "Evaluating Relative Mode Effects in Mixed-mode Surveys: Three Methods to Disentangle Selection and Measurement Effects." *Sociological Methods & Research* 42(1): 82-104.
- Vannieuwenhuyze, Jorre T. A., Geert Loosveldt, and Geert Molenberghs, 2010, "A Method for Evaluating Mode Effects in Mixed-mode Surveys." *Public Opinion Quarterly* 74(5): 1027-1045.
- , 2012, "A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys." *International Statistical Review* 80(2): 306-322.
- , 2014, "Evaluating Mode Effects in Mixed-mode Survey Data Using Covariate Adjustment Models." *Journal of Official Statistics* 30(1): 1-21.
- Voogt, Robert J. J., and Willem E. Saris, 2005, "Mixed Mode Designs: Finding the Balance between Nonresponse Bias and Mode Effects." *Journal of Official Statistics* 21(3): 367-387.
- Weijters, B., N. Schillewaert, and M. Geuens, 2008, "Assessing Response Styles across Modes of Data Collection." *Journal of the Academy of Marketing Science* 36: 409-422.
- Yanovitsky, I., E. L. Zanutto, and R. Hornik, 2005, "Estimating Causal Effects of Public

Health Education Campaigns Using Propensity Score Methodology.” *Evaluation and Program Planning* 28(2): 209-220.

Zhang, Chan, and Frederick Conrad, 2014, “Speeding in Web Surveys: The Tendency to Answer Very Fast and Its Association with Straightling.” *Survey Research Methods* 8(2): 127-135.