

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Reducing Postal Survey Nonresponse Bias by Sample Selection Incorporating Noncontact Propensity

A thesis presented in partial fulfilment of the requirements of
the degree of Doctor of Philosophy at Massey University.

Benjamin John Healey

2008

Abstract

Noncontact, the failure of a postal survey sample member to receive a survey request, is a potential source of nonresponse bias that has largely been ignored. This is due to the difficulty of separating the components of nonresponse in postal surveys when nothing is heard from potential respondents. Yet, the need to understand postal nonresponse is increasing as more studies move to mixed mode designs incorporating a postal element, and technological, resource and societal changes increase the attractiveness of self-administered surveys. Thus, this research sought to estimate the level of noncontact in postal surveys, to identify the direction and magnitude of bias due to it, and to investigate targeted in-field mechanisms for reducing this bias. A series of empirical studies involving New Zealand postal surveys fielded between 2001 and 2006 were undertaken to meet these aims.

Noncontact was found to relate to survey-independent demographic variables (e.g., age, household composition). Furthermore, its incidence was estimated to be as much as 400% higher than indicated by 'gone, no address' (GNA) returns, although an envelope message tested as part of the research was able to increase levels of GNA reporting significantly. Thus, noncontact was established as a nontrivial source of nonresponse in the surveys examined.

As far as bias is concerned, noncontacts had a different profile compared to refusers and ineligibles, and were estimated to account for up to 40% of net nonresponse error for some of the variables in the surveys examined. Accordingly, there appears to be a clear opportunity for methods targeted at reducing noncontact bias to improve final survey estimates for a range of items.

A number of potential methods for reducing noncontact bias were explored, but only one had both a compelling theoretical foundation and potential for wide applicability; the noncontact propensity sampling (NPS) scheme. In a resampling simulation study a prototype of the scheme, which increases the selection probabilities for sample units with a higher predicted propensity for noncontact, consistently improved the demographic profile of valid postal survey returns compared to a simple random

sample (SRS). Furthermore, the scheme reduced nonresponse bias by an average of 28% as measured against a range of frame variables (e.g., age, gender) and 17% as measured against survey variables for which census parameters were known (e.g., religiosity, household size, qualifications, income and marital status).

Although the prototype NPS procedure increased the standard deviation of simulated point estimates for a given sample size (1,500 in this research), the effect was small; an average of 4% for frame variables and 2% for survey variables. Furthermore, the scheme had almost no impact on reported cooperation rates and is likely to be cost effective compared to other potential targeted in-field mechanisms, particularly in situations where researchers regularly survey a specific population.

Pairing the scheme with three common post-survey adjustment methods (frame or census age/sex cell weighing, and response wave extrapolation) did not lead to consistently better estimates than an unweighted SRS. But this was largely due to the shortcomings of these methods because in many cases combining them with either sampling scheme (SRS or NPS) actually degraded estimates. This reinforces the idea that researchers should expend effort minimising bias during the field period rather than relying on post-survey weighting to deal with the issue.

Finally, since the NPS scheme aims to reduce error due to noncontact but is not expected to affect error due to other components (e.g., refusal, ineligibility), it presents an opportunity for researchers to begin decomposing the various facets of postal survey nonresponse bias, an important precursor to the development of other targeted bias reduction interventions. Thus, as a methodological tool, the NPS scheme may serve a dual role as both a bias reduction and decomposition mechanism.

In addition to their implications for postal survey research, the methods developed and insights into noncontact established in this research are likely to have applications in other domains. In particular, they will inform activities such as research into online survey nonresponse, organisational database management cost reduction and list procurement.

Acknowledgements

My principal supervisor, Professor Philip Gendall, has made a significant positive contribution to this research and, more generally, to my career. In addition to offering advice on methodology and structure, Phil provided access to critical survey data and spurred me along during periods of low motivation with horse racing analogies. This project would not have been possible without his generosity and I am very grateful for his guidance and support.

I would also like to thank Professor Stephen Haslett, who's co-supervisory advice on the statistical aspects of the research was invaluable. I appreciated his encouragement throughout the process, as well as the patience and good humour he demonstrated by not wincing visibly at some of my ideas.

Several of my colleagues at Massey University contributed their time, data or support to this research: Professor Janet Hoek and Claire Matthews provided access to survey response data; Pru Robbie, Tanya Banks, Vivien Michaels and Fiona Huston helped administer a number of the surveys on which the research is based; while Roseanne MacGillivray greased the wheels of the University bureaucracy and was a constant source of enthusiasm. My thanks go to them for their assistance.

My partner and fellow PhD candidate, Ninya Maubach, has shared the highs and lows of the project with me, sometimes whether she liked it or not. She was a sounding board when I needed it, endured my grumbling during write-up with a smile, and continued to encourage me even as the pressures of her thesis grew. Ninya also introduced me to Newton's three laws of graduation (Cham, 2001) which were strategically placed above my computer to help maintain my focus on completion. I am fortunate to have a mate so understanding, intelligent and embroiled in research!

Finally, I would be remiss to ignore the mainly silent contribution Oscar the Irish Terrier made to this process. The inevitable intrusion of research commitments into home life meant his pleas for a game of tug sometimes fell on deaf ears. He took this stoically, but I suspect I have a lot of playing to do to make up for lost time.

Contents

- List of Tables vii**
- List of Figures..... viii**
- List of Equations viii**

- 1. Background and Objectives 1**
 - 1.1. Introduction 1
 - 1.2. Nonresponse as an Important Error Source 2
 - 1.3. Postal Survey Nonresponse and its Components 12
 - 1.4. Project Structure and Objectives 24

- 2. The Nature and Extent of Postal Survey Noncontact 27**
 - 2.1. Introduction 27
 - 2.2. An Underexamined and Underreported Phenomenon 27
 - 2.3. Exploring Noncontact Reporting using Frame Change Data 30
 - 2.4. Characteristics of Sample Units that Changed Details 34
 - 2.5. Characteristics of Third Parties Reporting Noncontact..... 39
 - 2.6. Effect of Envelope Messages and Follow-Ups on Reporting Rates 45
 - 2.7. A Procedure for Estimating Unreported Noncontacts..... 47

- 3. Noncontact’s Contribution to Nonresponse Error 55**
 - 3.1. Introduction 55
 - 3.2. Approaches to Evaluating Postal Survey Nonresponse Bias 56
 - 3.3. An Empirical Analysis of Postal Survey Nonresponse Bias..... 65
 - 3.4. Response and Bias Trends across Multiple Postal Surveys 69
 - 3.5. Noncontact as a Contributor to Net Nonresponse Bias 74

- 4. Approaches to Reducing Noncontact Bias..... 83**
 - 4.1. Introduction 83
 - 4.2. Existing Approaches to Nonresponse Bias Reduction 83
 - 4.3. Potential Mechanisms for Targeting Postal Noncontact Bias 89
 - 4.4. Predicting Noncontact: Developing a Propensity Score 99
 - 4.5. A Proposed Noncontact Propensity Sampling (NPS) Procedure 107

5. Simulated Performance of the NPS Scheme	113
5.1. Introduction	113
5.2. Simulating NPS Procedure Performance	114
5.3. The Effect on Response Distributions	124
5.4. The Effect on Survey Estimates	126
5.5. Interaction with Three Common Post-Survey Procedures.....	130
5.6. A Promising Procedure	137
6. Summary, Applications and Future Directions	141
6.1. Introduction	141
6.2. Key Findings and Implications.....	142
6.3. Potential Applications	150
6.4. Limitations and Directions for Future Research	152
7. References	157
Appendix 1: Information on the Thesis Supplementary CD.....	167
A1.1. Workings for Total Noncontact Estimates.....	168
A1.2. Copies of ISSP Questionnaires from 2001 to 2006	168
A1.3. Copies of Census Forms from 2001 and 2006	169
A1.4. Walk-through of Calculation Steps in the Proposed NPS Scheme	169
A1.5. Modelling and Simulation SAS Code	170
A1.6. Detailed Result Tables for the Simulation Study.....	172
Appendix 2: Sources of Census Figures	173
A2.1. Notes on Census Data Sources and Calculations	174
Appendix 3: Logistic Regression Models.....	177
A3.1. Detailed Logistic Regression Model Specifications.....	178

List of Tables

Table 1: Cooperation rates for ISSP postal surveys in New Zealand	14
Table 2: Younger individuals were more likely to change details.....	35
Table 3: Some employment classes were more likely to change details	35
Table 4: People in younger households were more likely to change details.....	36
Table 5: People in multi-surname households were more likely to change	36
Table 6: Address type did not have a significant effect on address change	36
Table 7: Location type did not have a significant effect on address change	37
Table 8: Survey response by roll change classification	38
Table 9: Households with a higher average age returned at a higher rate	40
Table 10: Single and many surname households returned at a higher rate	40
Table 11: Split address households under-returned	40
Table 12: Households in metro areas were least likely to return	40
Table 13: Response by household average age and roll detail status.....	42
Table 14: Response by household surnames and roll detail status.....	43
Table 15: Response by address type and roll detail status	44
Table 16: Response by location type and roll detail status.....	44
Table 17: Multiple waves and an envelope message increased GNA returns	45
Table 18: The envelope message reduced the number of inactives.....	46
Table 19: Sample units with changed details responded in lower numbers	47
Table 20: Total noncontact by estimation method, treatment and wave.....	52
Table 21: Response to the six ISSP surveys.....	70
Table 22: Unweighted survey estimates compared to census figures	71
Table 23: Percentage difference between valids and the full sample on frame data	72
Table 24: Percentage change in estimated bias after multiple contacts	73
Table 25: Average values for frame variables by response disposition	75
Table 26: Relationship between frame and survey variables.....	77
Table 27: Wave extrapolated unweighted estimates compared to census	79
Table 28: Frame variables retained in the final logistic regression models.....	105
Table 29: Distribution of propensity scores in each modelled dataset	108
Table 30: Estimated adjustment rates for noncontact underreporting	119
Table 31: Proportion of the population in each census age/sex cell.....	123
Table 32: Effect of the NPS scheme on survey response [†]	125

Table 33: Proportion of the valid group in each propensity decile [†]	126
Table 34: Effect of the NPS scheme on estimates for frame variables [†]	127
Table 35: ‘Best scheme’ results for survey estimates compared to census.....	129
Table 36: Age/sex weights under each sampling scenario [†]	131
Table 37: ‘Best scheme’ results for survey estimates with frame weighting	133
Table 38: ‘Best scheme’ results for survey estimates with census weighting	134
Table 39: ‘Best scheme’ results for survey estimates with wave extrapolation.....	135
Table 40: Sources for individual census parameters	174

List of Figures

Figure 1: The relationship between total survey error, bias and sampling error	10
Figure 2: Cooperation rates for the ISSP survey appear in decline	15
Figure 3: Conceptual determinants of postal survey response	22
Figure 4: Timing of the frame snapshots and fieldwork for the study.....	32
Figure 5: Gains chart for model predictions on ‘build’ datasets	106
Figure 6: Gains chart for model predictions on ‘test’ datasets	107

List of Equations

Equation 1: General coverage error formula for a linear statistic.....	4
Equation 2: General nonresponse error formula for a linear statistic.....	7
Equation 3: A possible nonresponse rate formula	8
Equation 4: A common postal survey cooperation rate formula.....	9
Equation 5: Estimated total noncontact for the ‘changed’ group.....	49
Equation 6: Estimated total noncontact rate (overall)	49
Equation 7: Estimated total noncontact rate (Iceberg method).....	51
Equation 8: The oversampling rate for an NPS scheme band	110
Equation 9: Nonresponse error incorporating response propensity.....	138