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 ineligibiles, 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

List of Figures

List of Equations

1. Background and Objectives

"One of the important scientific challenges facing survey methodology at the beginning of this century is determining the circumstances under which nonresponse damages inference to the target population. A second challenge is the identification of methods to alter the estimation process in the face of nonresponse to improve the quality of sample statistics." (Groves, Dillman, Eltinge, & Little, 2002, p. xiii)

1.1. Introduction

Researchers typically conceptualise survey error as arising from four sources: sampling, coverage, measurement and nonresponse. All are worthy of consideration but nonresponse is of increasing concern, with longitudinal studies suggesting response rates are declining, or at best stable, in many countries across all modes (Curtin, Presser, & Singer, 2005; de Leeuw & de Heer, 2002). The consequences of this for valid inference from probability samples are now frequently discussed in the literature, as are potential methods for avoiding or mitigating nonresponse bias.

Components of nonresponse such as refusal, ineligibility and noncontact are not as easily separated in postal surveys as they are in interviewer-led modes. In particular, it is often not possible to distinguish between unreported noncontacts and passive refusers (i.e., between those who are not contacted at all and those who receive the survey but do not respond in any way). Current strategies for reducing mail nonresponse therefore revolve around questionnaire design, incentivisation and repeated general contacts, all of which only address those who receive the survey invitation. One consequence is that the postal nonresponse literature largely ignores the potentially differential contribution to survey error of noncontact.

Yet, there is reason to expect that a better understanding of postal survey noncontact may facilitate the development of methodological techniques effective at reducing any bias it introduces. Therefore, in addition to examining existing research relating to postal survey nonresponse and presenting a conceptual model of its key components, this chapter outlines the objectives of a series of studies aimed at investigating the nature and extent of noncontact in the postal mode.

1.2. Nonresponse as an Important Error Source

Sample surveys are a valuable tool for researchers in their efforts to aid decision making, whether it be guidance in the development of policy, gauging likely consumer response to business initiatives, or tracking changes in population state over time. Yet, one does not have to search long to find examples of survey applications that have contributed to erroneous decisions or predictions. For instance, in 1936 the now infamous Literary Digest poll, which since 1920 had enjoyed a perfect record of predicting US elections, forecast a landslide victory for Republican Alf Landon. The election that year was won by incumbent Franklin Roosevelt by a margin of 24%. Although it was impossible to conclusively identify the causes of this substantive error, Squire (1988) undertook an extensive analysis of information related to the poll from the time and concluded the error was likely due to non-random sampling along with substantial nonresponse bias.

More recently, at a *Research Industry Summit for Improving Respondent Cooperation* held in 2006 by executives from leading global research and consumer goods companies, *Procter & Gamble's* VP of consumer and market knowledge presented one example in which online and postal surveys on an instant-coffee concept came up with opposing results (Neff, 2006).

Such failures provide a compelling motive to better understand the causes of survey error and how to avoid them. Unsurprisingly, then, there is a large body of literature dedicated to bias identification and reduction across various populations, survey modes and resource constraints.

1.2.1. Error Sources and their Classification

Groves (1989) asserted that researchers in disciplines as diverse as psychology, econometrics, and anthropology have considered various aspects of the error problem, but that a cross-disciplinary understanding of its component sources was required. He put forward an error source taxonomy that has subsequently been used extensively by social survey methodological researchers. The taxonomy separates errors into four distinct types: coverage, sampling, measurement, and nonresponse.

Coverage Error

Coverage error relates to discrepancies between the set of people or other entities of interest (the 'target population') and the list or 'frame' used to select a sample from that set (the 'frame' population'). Discrepancies occur when:

- The target population and frame population do not correspond to one another at a conceptual level, such that even if the sampling frame were complete it would not represent the target population. Sometimes termed 'overcoverage' because some entries on the frame are linked to nonmembers of the population, this would occur, for instance, if a researcher used a general sample of households to survey people of a specific ethnicity.
- There is conceptual correspondence between the target population and frame population, but the sampling frame is not complete and therefore does not fully represent the target population. Termed 'undercoverage', an example of this would be the New Zealand electoral roll. The roll enjoyed an impressive coverage rate of more than 90% of eligible voters prior to the last two elections (New Zealand Electoral Enrolment Centre, 2005). Yet, because it 'undercovers' the population of voters, samples taken from it may be exposed to coverage error.
- Each member of the frame population does not have an equal chance of selection in the sampling frame because some members are disproportionately represented. That is, one or more members of the population are linked to more than one entry on the frame, thus giving them multiple chances to be chosen. This would occur, for instance, if a telephone directory was used to sample households. Sometimes, one household may have multiple listings in a telephone directory because more than one member has their name included.

The coverage rate for a given frame, target population and sampling procedure is a measure of the effective representation of target population members in the frame. Combinations of the discrepancies described can occur and lead to a reduction in the coverage rate for a specific study. However, a coverage rate of less than 100% does not necessarily lead to bias in estimates obtained from the sample. This is because, for linear statistics (e.g., means, proportions, counts) such as those examined in this thesis, coverage error is a function of both the coverage rate and the difference in values between the covered and noncovered elements of the target population. Equation 1, below, illustrates this.

Equation 1: General coverage error formula for a linear statistic

$$y_c = y + \frac{n_{nc}}{n}(y_c - y_{nc})$$

Where:

y_c = The value of the statistic for those covered by the frame;

y = The value of the statistic on the full target population;

n_{nc} = The number in the target population not covered by the frame;

n = The total number in the target population;

y_{nc} = The value of the statistic for those not covered by the frame.

(Source: Groves, 2004, p. 85)

The potential for coverage error exists in all survey modes and is typically dependent on the population under study and the frame employed. Take, for instance, a survey of the general population to be conducted over the internet with a sample based on email addresses published in the telephone directory. One would expect a very low coverage rate, and corresponding higher potential for coverage error, in such a study. This is because many members of the population do not have access to the internet and only very few of those that do actually publicly list their email addresses. Conversely, a high coverage rate would be expected in an intra-organisational postal survey of employees using the organisation's human resource data as a frame.

Random Sampling Error

Of the four sources of survey error, random sampling error is arguably the most well known and understood. It occurs because not all members of the frame population are surveyed in a randomly selected sample. As such, the data collected for those members surveyed cannot normally be expected to perfectly reflect those that would have been collected had the entire frame population been surveyed.

Unlike the other categories of error, random sampling error can be estimated in certain conditions via standard statistical techniques underpinned by the *Central Limit Theorem*. A detailed discussion of the theorem, methods of estimating random sampling error (i.e., standard errors), and the associated calculation of error ranges or ‘confidence intervals’ for a survey statistic can be found in foundation research texts such as Hair, Bush and Ortinau (2006) or Churchill and Iacobucci (2005).

In brief, the theorem states that for simple random samples of a reasonable size (typically 30 or greater), the sample means will be approximately normally distributed. The distribution of sample means has parameters related to the population sampled that can be employed to infer degrees of certainty with regard to any individual sample. Furthermore, the degree of certainty relating to a survey estimate is greater when the sample size is larger. A key requirement of the theorem is that a probability-based method of selection must be employed. In contrast to coverage error, then, sampling error is related to the sampling procedure employed rather than the frame from which the sample is selected.

Because of its relative ease of measurement, sampling error is often given most attention in the consideration of, and planning for, survey error. However, sampling error may be swamped by the other, *non-sampling*, error sources.

Measurement and Design Error

The third source of error in Groves’ (1989) typology relates to the implementation of the survey instrument itself rather than the frame or sampling procedure employed. Often termed error due to ‘measurement’ or ‘design’, and classified by Groves (2004) as ‘observational’ error in contrast to the ‘nonobservational’ nature of the other error sources, it comprises a vast array of factors including:

- Questionnaire layout design effects;
- Question ambiguity and item order effects;
- Interviewer error in question delivery and recording;
- Respondent error as a result of task misunderstanding;
- Reporting error related to the memory or level of knowledge of respondents;
- Errors in the design of the sampling process;
- Data entry error (including keypunching, coding, and programming errors).

It is not possible within the scope of this thesis to provide a detailed description of the factors related to survey measurement error. Hence, the reader is directed to Groves (2004), who breaks his extensive discussion of this source of error down into three sub-categories relating to the interviewer, the respondent, and the questionnaire. Similarly, Dillman (2000) provides a comprehensive overview of questionnaire-based measurement error as it applies to self-administered surveys (i.e., mail and internet), and many of the concepts he discusses translate to other modes.

Measurement error occurs when there is a difference between survey statistics and the “true value” in the population due to the above factors. Often, it is very difficult or impossible to determine the “true value” of a variable because the population value itself may fluctuate over successive measures (Groves, 2004; Kish, 1995). Indeed, evidence suggests that constructs such as attitudes and opinions can be influenced merely by the act of measuring them (e.g., see Morwitz, 2005).

There are, however, common techniques that can be employed to obtain information about the likely direction and extent of measurement error. For instance, Groves (2004) outlines several options including laboratory experiments resembling the survey interview, comparing against external measures, randomised assignment of measurement procedures to sample persons and repeated measures on individuals.

Nonresponse Error

Nonresponse error relates to those situations in which some selected frame elements fail to complete some or all of the survey questions. It is similar in nature to coverage error in that nonresponse error is not a necessary consequence of a response rate less than 100%. Rather, as demonstrated in Equation 2, nonresponse error for a linear statistic is a function of both the response rate and the difference in values between those who respond and those who do not respond to the survey instrument.

Thus, nonresponse error is not inevitable at any level of survey response. However, since researchers generally cannot quantify nonrespondent values, but often *do* know the nonresponse rate, much effort is taken in practice to minimise the nonresponse rate in an attempt to minimise the overall potential for nonresponse bias.

Equation 2: General nonresponse error formula for a linear statistic[1]

$$y_r = y_n + \left(\frac{nr}{n}\right)(y_r - y_{nr})$$

Where:

y_r	= The value of the statistic for those who responded;
y_n	= The value of the statistic for the entire sample;
nr	= The number of nonresponders;
n	= The total number in the sample;
y_{nr}	= The value of the statistic for those who did not respond.

(Source: Groves, 2004, p. 133)

In contrast to coverage error, researchers undertaking studies based on restricted invitation samples (e.g., probability, pseudo-probability, or quota) are often able to make a reasonable determination of the level and source of nonresponse.

Specifically, nonresponse (nr) may be decomposed as follows[2]:

val	= units presenting complete valid responses to the item;
part	= units presenting partial valid responses to the item;
inv	= units presenting invalid responses to the item;
ref	= units actively refusing to complete the item;
inact	= units from whom no form of response to the item is presented;
inel	= units identified as being ineligible to complete the item;
nc	= units that were not exposed to the item due to noncontact.

[1] This is a simplified calculation for nonresponse error because it assumes that, for a given survey design, all potential respondents have a response propensity of either zero or one (i.e., that nonresponse is deterministic). An alternative equation, incorporating the more realistic assumption that individual nonresponse is probabilistic, is presented and discussed in a later section (5.6, p. 137).

[2] *The American Association for Public Opinion Research* (2008) has developed a set of "standard definitions for final dispositions of case codes and outcome rates for surveys" that presents a much more detailed typology of survey outcomes for different survey modes. However, the categories outlined here represent the nonresponse outcomes at a level sufficient for the discussion developed in this chapter. Furthermore, for the purposes of this discussion, partial returns are classed as cooperative responses.

Hence, the nonresponse rate (P_{nr}) for an item could be expressed as follows:

Equation 3: A possible nonresponse rate formula

$$P_{nr} = \frac{nr}{n} = \frac{inv + ref + inact + inel + nc}{n} = 1 - \frac{val + part}{n}$$

All of the component terms in Equation 3 are present in one form or another in common calculations of response rates. Yet, a difference exists in the structure of different formulae employed by researchers such that the nonresponse rate is not always the inverse of the response rate. In fact, one of the problems facing those interested in nonresponse is that different researchers employ different formulae to determine response rates (Shaw, Bednall, & Hall, 2002; Wiseman & Billington, 1984). In response to this, at least two American industry organisations have worked to establish standards in this area (Frankel, 1982; The American Association for Public Opinion Research, 2008). Nevertheless, these standards are voluntary and will take time to diffuse, meaning differences in practice are likely to remain in the medium term.

In the postal mode, one common deviation from the structure outlined in Equation 3 involves excluding ineligible and noncontacted sample units from the calculation. For example, in a review of response rates to postal survey studies published in medical journals in 1991, Asch, Jedrziewski and Christakis found the following:

> "*Response rates reported in manuscripts often differed from the response rate calculated by dividing the number of surveys received by the number distributed. Many of these differences reflected adjustments to account for surveys considered unusable – either because they were returned by the post-office as undeliverable, or because the subjects failed to meet study criteria. However, there was also great inconsistency and confusion about how to make these adjustments. Some authors deleted unusable surveys from the numerator, effectively lowering their reported response rate. Others deleted unusable surveys from the denominator, raising their reported response rate.*" (Asch, Jedrziewski, & Christakis, 1997, p. 1131)

Although it is not possible to establish with certainty because, as Asch et al. (1997) note, many studies do not report their response rate formulae, it appears that the latter approach of subtracting ineligibles and noncontacts from the denominator is common. Indeed, Asch et al. (1997) do this when determining response rates for the supplementary postal survey undertaken in their study, and some International Social Survey Programme (ISSP) members also take this approach when reporting response to studies fielded in postal format (International Social Survey Programme, 2003, 2004)[3]. Equation 4 presents the structure of this formula visually.

Equation 4: A common postal survey cooperation rate formula

$$P_{cr} = \frac{val + part}{(val + part + inv + ref + inact)} = \frac{val + part}{n - (inel + nc)} \neq 1 - P_{nr} = \frac{val + part}{n}$$

Here, the equation is labelled a 'cooperation rate', as this is the term given to a response calculation that excludes ineligible and noncontact dispositions from the denominator in the standard formulae developed by *The American Association for Public Opinion Research* (2008). In studies undertaken for this thesis, the term 'cooperation rate' will be employed wherever such a calculation is performed.

Although subtle, the common exclusion of ineligibles and noncontacts in the denominator term is relevant because it can lead postal survey researchers to ignore the potential impact of these dispositions on nonresponse error. The implications of this are discussed further in section 1.3.2.

Total Survey Error

The four types of error described above occur to varying degrees in surveys of all modes (Dillman, 2000; Groves, 2004; Kish, 1995). Kish (1995 - first published 1965) describes how they contribute to total survey error – the difference between sample estimates and population parameters – by conceptualising their relationship on a right-angle triangle (e.g., see Figure 1).

[3] Only a subset of members field the surveys for this programme in postal format. Of those that did, Canada, Denmark, Finland, the Netherlands, and New Zealand all subtract noncontacts and/or ineligibles from the denominator.

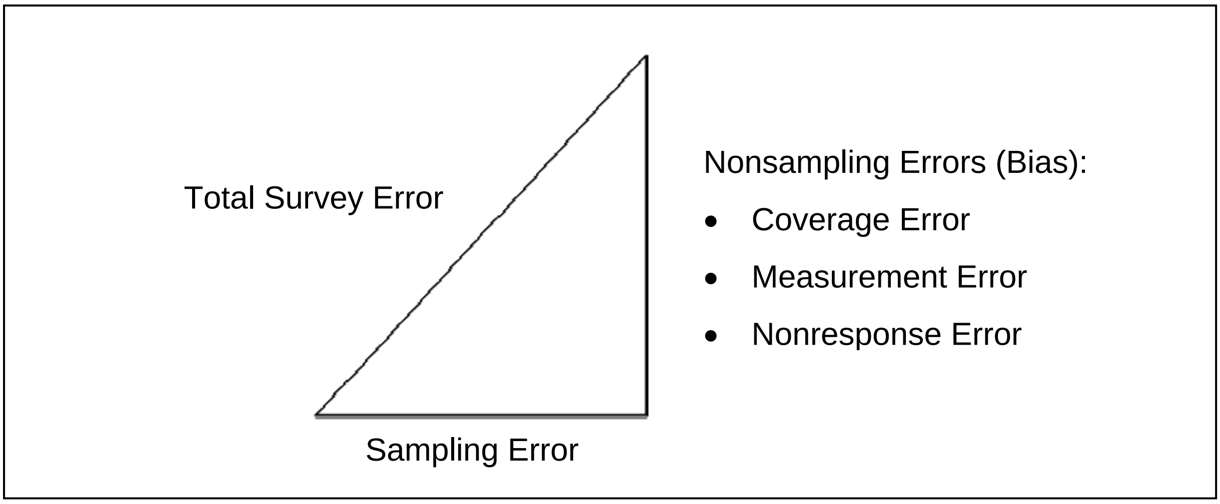

Figure 1: The relationship between total survey error, bias and sampling error[4]

Readers can find a comprehensive discussion of total survey error in Kish (1995), but one key point arising from his conceptualisation is that nonsampling errors may easily swamp those from sampling. Indeed, in the examples of survey failure cited at the beginning of this section, random sampling error is likely to have been trivial. With this in mind, research industry leaders such as Lavrakas (1996) have increasingly begun to advocate a total survey error approach to maximising the effective use of scarce research resources. This approach encourages a wider perspective on survey error than that taken by many researchers and managers, who Lavrakas (1996, p. 35) suggests show *"an obsequious devotion to 'the god of sample size'"* by focusing myopically on the one source (sampling error) for which theory-based estimates of error can be obtained:

> *"The TSE [total survey error] perspective presents a compelling argument that it is both foolish and wasteful to let sampling error drive decisions about survey design and resource allocation. Within the almost certain future climate of tight and 'balanced' … survey budgeting, sample sizes must be reduced and fixed resources redeployed to reduce and measure other sources of total survey error more successfully."* (Lavrakas, 1996, p. 35)

[4] An assumption inherent in this diagram is that sampling and nonsampling errors are not correlated. Where sampling and nonsampling errors are correlated, the triangle would not be right-angled.

1.2.2. The Importance of Nonresponse

Of the three nonsampling error sources, nonresponse is arguably the most visible because it is clearly evident in response rate calculations and its incidence is typically more easily quantified than that of coverage or measurement errors. Furthermore, evidence suggests that response rates to household sample surveys in the United States and Western Europe have decreased or, at best, remained static over time despite efforts aimed at increasing them (e.g., see Groves, Fowler et al., 2004).

Concern amongst US market research practitioners about declining response is not new: the industry formed an organisation in 1992, *the Council for Marketing and Opinion Research*, to help combat it. However, disquiet does appear to have gained momentum in recent years. For instance, Neff (2006, p. 1) reports that figures released during a September 2006 *Research Industry Summit for Improving Respondent Cooperation* suggest "*Some 59% of research companies are concerned about respondent cooperation, up from 49% in 2005. Moreover, 16% list it as their biggest concern.*"

Academics have also directed substantial attention to better understanding nonresponse over the past decade. This effort is exemplified by the recent publication of research compilations (Groves & Couper, 1998; Groves et al., 2002; Koch & Porst, 1998), the establishment of an annual international nonresponse conference (see www.nonresponse.org), and the appearance of special issues dedicated to the topic in three top-ranking methodological journals (de Leeuw, 1999; Lynn, 2006; Singer, 2006). This focus is unlikely to wane in the near future, as societal pressures continue to work against respondent cooperation and budgetary constraints force researchers to carefully manage total survey error. There is a clear need, then, for research in the area.

Although much has been done to better understand the factors influencing respondent cooperation, many aspects of nonresponse remain under-investigated. For instance, more work is necessary to support or refute the numerous theories of response choice in current existence (Gendall, Hoek, & Finn, 2005; Groves, Fowler et al., 2004) and examine the effect of ineligibility on survey estimates in different situations (Groves, Fowler et al., 2004). Furthermore, aspects of nonresponse that have been investigated in detail for some modes remain neglected for others. Noncontact nonresponse in postal surveys is one such aspect.

1.3. Postal Survey Nonresponse and its Components

Sample surveys can be undertaken in a number of modes, each relating to a general form of methodological implementation. Four commonly defined modes represent the vast majority of surveys conducted: face-to-face, telephone, internet and postal mail. Of these, the first two can be classified as ‘interviewer-led’, because a human intermediary presents the survey and records responses. Conversely, the latter two are typically classed as ‘self-administered’.

Data from the United States (Dillman, 2000) suggests that postal mail is not currently the dominant mode for large-scale general population studies. This is because, compared to interviewer-led modes, it can be slow to field (typically a minimum of six weeks for fieldwork), assumes high levels of literacy amongst respondents, provides limited opportunities for response probing or question clarification and does not readily allow for confirmation of respondent identity.

It can also be difficult to find adequate sampling frames of individuals for postal surveys in some settings and the sampling approaches employed in interviewer-led modes (e.g., Random Digit Dialing and Random Walks) do not readily translate to postal implementation. For instance, although it may be possible to select households using a map or phone list, the procedures for selecting individuals within these units typically rely heavily on the fluid exchange of information enabled by the presence of an interviewer. Attempts to implement them via an introductory letter or set of screening items on the questionnaire may confuse respondents and, ultimately, it would be hard to tell whether the selection rules had been applied at all.

Nevertheless, postal surveys *are* often used where the research is intra-organisational or is to be developed and implemented in-house (Dillman, 2000). This includes customer surveys such as those aimed at examining satisfaction with an organisation. Furthermore, health and epidemiological studies are often undertaken via postal survey (Edwards et al., 2002). Reasons for this mode's popularity in such circumstances include reduced costs compared with interviewer-led modes, the potential for good response rates with careful implementation, its appropriateness for widely dispersed populations and the option to easily present visual concepts. The self-administered nature of postal surveys also eliminates interviewer-related bias and allows respondents to reflect on their answers in their own time.

Indeed, there is good reason to believe postal mail will gain in importance as a mode in the future. Specifically, Dillman and Miller (1998) point out that advances in scanning technology now mean data entry can be automated, reducing the time and cost associated with this aspect of postal survey implementation. Furthermore, Dillman (2000) asserts that there is a societal trend toward self-administration and self-service both online and offline which works in the favour of mail and internet modes.

Additionally, there are a number of issues facing telephone as a survey mode, including increased mobile-only phone use[5], the advent of call blocking technologies, greater intolerance toward phone-based intrusions on time and the commonplace use of answer-phones. Faced with this situation, as well as significant problems in sourcing frames for internet surveys, some researchers are turning to postal mail as part of a mixed-mode strategy for reaching a representative sample of their target population (Best & Radcliff, 2005). Together, these factors point to a growth in surveys completed or initiated by post and, in turn, continued interest in postal survey nonresponse.

1.3.1. Trends in Postal Survey Response Rates

Common understanding is that response rates to all traditional modes of survey research have been declining over time. Indeed, this view is supported by recent research examining trends for a selection of longitudinal telephone surveys (e.g., Bednall & Shaw, 2003; Curtin et al., 2005) and face-to-face surveys (Groves, Fowler et al., 2004). However, comprehensive meta-analyses of long-term nonresponse trends in face-to-face or telephone surveys present only moderate evidence of a

[5] The journal *Public Opinion Quarterly* dedicated a special issue (vol. 71, issue 5) to this topic in 2007.

substantive decline in response over time. Specifically, de Leeuw and de Heer (2002) report the results of a long-term cross-national study of government surveys and claim that, although there was evidence for an international decline in response rates over time, the rate varied significantly by both country and survey.

In one of the few studies to investigate longitudinal response rates for postal surveys, Hox and de Leeuw (1994) found that levels appeared relatively stable across the period they examined (1947-1992) while at the same time rates for face-to-face and telephone surveys declined. Thus, it is possible that, although nonresponse to personal interview methods increased over those decades, it did not to any great degree for mail methods. However, the Hox and de Leeuw (1994) data are now 15 years old. Thus, little is known about mail survey response rates in recent times.

In order to examine the postal response situation relating to the context of this thesis, cooperation rates to an ongoing yearly survey undertaken as part of the *International Social Survey Programme* (ISSP) were examined. The topics for the ISSP survey are rotated every 7 years. Each survey is administered to the general population and, in New Zealand, has been undertaken by the same organisation since 1991. Table 1 presents the cooperation rates by survey topic over the last 15 years.

Table 1: Cooperation rates for ISSP postal surveys in New Zealand

Survey Topic	**Replication Years**	**Rate Yr. 1** (%)	**Rate Yr. 2** (%)	**Change** (%)
Religion	1991, 1998	66	65	-1
Social Inequality	1992, 1999	68	61	-7
Environment	1993, 2000	70	62	-8
Social Networks	2001	61		
Family/Gender Roles	1994, 2002	71	60	-11
National Identity	1996, 2003	67	57	-10
Citizenship	2004	62		
Work Orientation	1997, 2005	70	58	-12
Role of Government	1997, 2006	70	60	-10

All rates were calculated according to Equation 4, p. 9.

There were differences in some of the design details from one study replication to another, so conclusive statements about response trends cannot be made. However, it does appear that nonresponse to the ISSP surveys has increased over the last decade. A best fit line through the response rates in Table 1 suggests they have declined by just under a percentage point per year on average (see Figure 2).

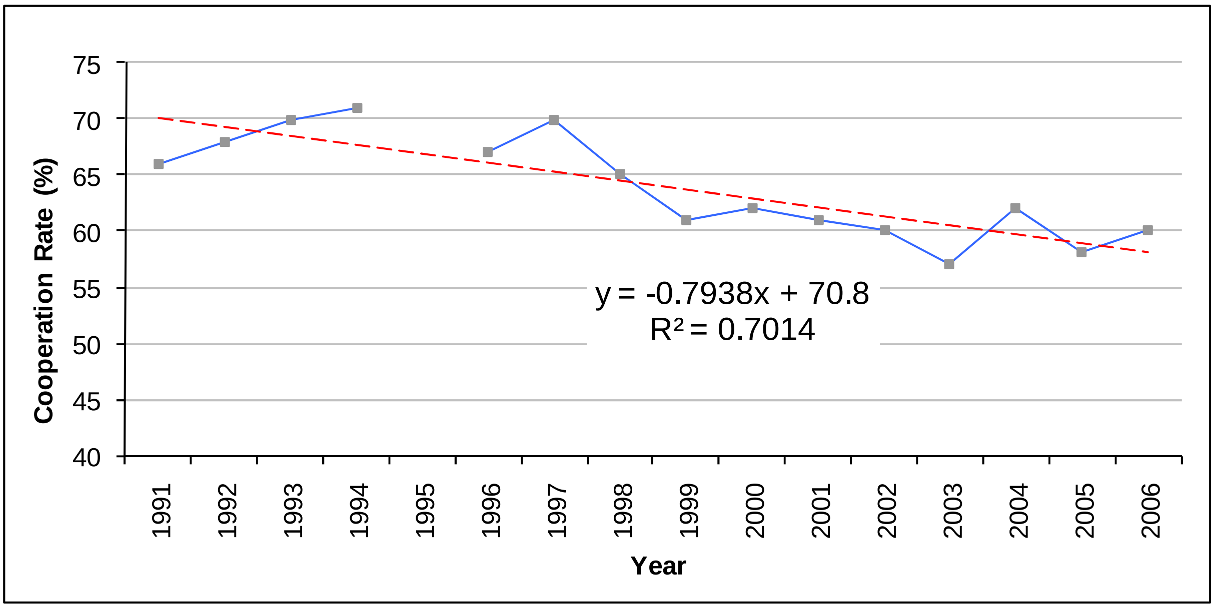

Figure 2: Cooperation rates for the ISSP survey appear in decline

There are a number of reasons why a decline in postal survey response might not be easily observable in the meta-analysis cited earlier. For instance, comparisons can be distorted by modifications to respondent selection, survey design, nonresponse measurement or calculation, fieldwork procedures and changes in the research organisation. Additionally, many of the studies included related to governmental or academic surveys, which may have experienced levels of decline different from commercial surveys. Finally, there has been a concerted effort on the part of researchers to improve postal response rates over recent decades via techniques such as personalisation, pre-notification, increased follow-up contacts and the use of participation incentives (Dillman, 2000). It is likely that these efforts have had a degree of success in the face of societal changes counting against response, such that the response rates examined in longitudinal studies may appear stable.

There are, however, many reasons why response rates might have been expected to decline over time in the absence of countermeasures. Specifically, rising work pressures mean individuals are likely to feel they have less time to undertake voluntary activities. In conjunction, people are faced with more commercial impositions on their time in the form of telemarketing and direct mail promotions. This material competes directly with survey requests for individual attention and has no doubt led to ‘promotional burnout’ amongst consumers, such that less overall attention is now willingly given to any unsolicited communications. Indeed, commercial organisations increasingly use customer “surveys” as a form of promotional tool, further blurring the distinction between research and sales pitches. Looking to the future, it is unlikely these pressures on survey response will abate.

In summary, it appears that postal survey response rates have, at best, been stable over time despite concerted efforts by researchers to improve them. Further, the societal factors that might be expected to contribute to nonresponse look set to continue and compound in the future. The implication of this is that work focused on examining and reducing postal survey nonresponse and, more importantly, its associated bias, will continue to be valuable to the research community.

1.3.2. The Components of Nonresponse in Postal Surveys

Nonresponse is an umbrella term employed to refer to a failure to collect data from a sample unit. Where there is a failure to collect *any* intended data from a sample unit, nonresponse is said to have occurred at the unit level. Conversely, where data are successfully collected from the sample unit but some pieces are incomplete, nonresponse is said to have occurred at the item level. Conceptually, then, it could be argued that unit level nonresponse is a special case of 100% item level nonresponse. However, some of the causes of nonresponse only lead to unit level nonresponse, while others can be the cause of either unit or item level nonresponse.

Furthermore, although all modes are susceptible to the various sources of survey error, the presence or absence of a human intermediary in the process means that they differ in the degree to which components of nonresponse can be separated out and targeted by researchers. The rest of this section therefore presents a breakdown of the key components of nonresponse as they relate to the postal mode. The terms introduced here are referred to throughout the thesis.

Ineligibility[6]

In the context of the postal surveys examined in this thesis, ineligibility is used to describe situations in which contacted sample units cannot provide information because they do not understand the language used, are mentally or physically incapable of adequately responding, are illiterate, or for some other reason are unable to comply with the survey request. A source of unit nonresponse, ineligibility is likely to affect surveys differentially depending on the population and survey topic under examination. For instance, a survey on mental health or one aimed at an immigrant or elderly population will probably encounter a substantial amount of ineligibility that could be expected to lead to a nontrivial level of bias.

For most general population and household surveys employing a robust frame and selection procedure, ineligibility is likely to be a small component of overall nonresponse. Certainly, in the six surveys analysed later in this thesis, reported levels never rose above 3% of the initial sample (see Table 21, p. 70).

Active Refusal

Active refusal occurs when a contacted sample unit declines to comply with either the entire survey request or with specific items in a postal questionnaire. Hence, it is a source of both unit and item level nonresponse. At both levels, active refusal is easily identifiable as an active negative response to the survey request. This could be via an indication on the returned questionnaire or a separate communication by the sample unit. In all modes of survey research other than postal mail, partial responses (breakoffs) are also a calculable form of active refusal. Although breakoffs no doubt occur for postal surveys, it is difficult to monitor their incidence because they are indistinguishable from an inaction response as discussed below.

---

[6] The term 'ineligibility' classifies cases that received the survey invitation but were known to be unable to reply. That is, they are clearly not refusals or noncontacts. Such cases are simply labelled 'other' in the terminology employed by the American Association of Opinion Researchers. However, 'ineligibility' is used here to provide a clear distinction between this group and the mixed 'inactive' group that makes up a large proportion of mail survey nonrespondents and which, arguably, could also be described as an 'other' group.

Like ineligibility, active refusal may affect surveys differentially depending on both population and survey characteristics. Certainly, studies into postal survey response correlates have found that factors such as salience of the survey topic (Dillman & Carley-Baxter, 2000) and source of the survey request (Fox, Crask, & Kim, 1988) influence people's propensity to respond.

Inaction

By far the largest component of nonresponse for most postal surveys, inaction represents those situations where no response is received from a sample unit to one or more requests. At the unit level, inaction can occur because the sample unit declines to participate but does not inform the researcher or simply does not 'get around' to completing or sending back the questionnaire (i.e., passive refusal). Additionally, it can occur where the researcher is not informed of a sample unit's ineligibility or noncontact. At the item level, inaction occurs where the sample unit does not answer one or more survey items because they refuse to answer but fail to indicate this or because they mistakenly skip a question.

This component of nonresponse is unique to self-administered survey modes. In interviewer-led surveys, nonresponders can be classified into clearly defined disposition codes – refusal, noncontact, partial response, or valid – based on the interviewer's determination after interaction with the respondent or their household. Furthermore, the presence of an interviewer to guide respondents through a survey reduces the chances questions will be inadvertently overlooked.

It is due to the existence of this 'mixed bag' component in postal survey research that highly targeted approaches to nonresponse bias reduction are not often pursued. Typically all nonresponders in this category are treated as though they are passive refusers by researchers, who attempt to reduce its size via techniques reliant on the sample unit receiving the survey invitation (e.g., incentives, multiple contacts, etc.).

Noncontact

A source of unit level nonresponse, noncontact arises in postal surveys whenever the survey request is not delivered to the intended sample unit. This can occur because the request is lost en-route, delivery is not accepted by the sample unit, or the intended recipient is no longer at the postal address used for the request. Of these causes, incorrect addressing is the largest source of noncontact for most surveys and is related to both the quality (with respect to age and data entry) of the sample frame and the movement over time of the population it relates to.

To the extent that more movement occurs amongst some subpopulations (e.g., the young, or Maori), noncontact is likely to affect surveys differentially depending on the population under examination. For instance, a survey of students may encounter more noncontact than one of retirees because the former may be more likely to change address in the time between the collection of original frame information and delivery of the survey invitation. One important distinction between noncontact and the other components of nonresponse is that, while noncontact arises because sample units never even receive the survey invitation, the other components arise because sample units receive the invitation but make a decision to not respond. Hence, propensity to be a noncontact can be considered conceptually independent of a sample unit's propensity to comply with a specific survey invitation they receive.

Very little research has been published regarding the incidence of noncontact across different postal survey situations, let alone its affect on survey estimates. Yet, noncontact is a non-trivial source of nonresponse in other survey modes (e.g., see Lynn & Clarke, 2002). Indeed, although no published studies of postal survey noncontact rates exist, a brief examination of historical records from recent ISSP studies conducted in New Zealand, Australia and Canada indicates that combined ineligibility and noncontact rates ranged from 4% to 14%, with 6 out of the 7 surveys reporting above 8% (International Social Survey Programme, 2003, 2004, 2005). As noted earlier, in the New Zealand studies, for which more detailed breakdowns are available, the ineligibility rate never rose above 3% while noncontact was never less than 8% (see Table 21, p. 70). This suggests that noncontact may be a non-trivial source of nonresponse for many postal surveys.

1.3.3. Opportunities for Reducing Error due to Noncontact

Declining response is not cause for concern in and of itself, since it is possible that nonresponders and responders do not differ with respect to the variables of interest in any particular study. However, it is often not practically possible to assess the degree of dissimilarity between groups and, so, an assumption is made that they *are* dissimilar and that reducing the size of the nonresponse group will reduce any bias due to nonresponse.

Although it is true that reducing nonresponse to zero would eliminate nonresponse bias in a survey, it is not necessarily true that partially reducing it will decrease bias. Hence, when addressing nonresponse it is important to take into consideration its different components, their sources and whether or not they contribute differentially to bias. It is possible, for instance, that in the right balance bias due to noncontact could cancel out bias due to active refusal. If this were the case, efforts to reduce nonresponse may alter that balance and, in turn, increase bias in the survey estimates. An understanding of the underlying nature of nonresponse should therefore be a critical input into sample and survey design.

As discussed earlier, the components of postal survey nonresponse arise for different reasons and are likely to relate to different survey-relevant population characteristics. Furthermore, evidence from research undertaken on other survey modes suggests that two key components of nonresponse, noncontact and refusal, are not only increasing for different reasons, but may also lead to different biases in survey estimates. For instance, de Leeuw and de Heer (2002) specifically separated noncontact and refusal nonresponse in their cross-national longitudinal study and concluded the following:

> *"…analyses showed that there are differences between countries in noncontact rate and that the noncontact rates are increasing over time, but that there are no differences between the countries in the rate in which the noncontacts are increasing. The difference in nonresponse trends over the countries is caused by differences between countries in the rate at which the refusals are increasing. For some countries, the increase in refusal rates is much steeper than for other countries." (p. 52)*
>
> *"Both contribute to overall nonresponse, but different factors influence each source" (p. 45)*

Similarly, in a study directed at understanding the bias contributed by nonresponse components as well as their level of incidence, Lynn and Clarke (2002) examined data from three national face-to-face household surveys in the UK. They found bias was indeed introduced by those who are difficult to contact and that it was different to that introduced by refusers.

Both de Leeuw and de Heer (2002) and Lynn and Clarke (2002) examined nonresponse in interviewer-led modes of research (telephone and face-to-face). Thus, their findings cannot be extrapolated directly to the postal survey context because of the different nature of noncontact in self-administered surveys. Specifically, noncontact in postal surveys occurs because the sample unit is not at the address specified, whereas noncontact in interviewer-led modes tends to occur because the sample unit is not available at the time of call. Nevertheless, the results from other modes suggest the possibility that nonresponse components differ in both incidence and influence in postal surveys. Certainly, the one published study to attempt to address the 'influence' aspect of this question (Mayer & Pratt, 1966) concluded that there was a difference in the nature of noncontact and refusal bias.

One thing apparent from the literature with respect to postal surveys is that, apart from the exploratory Mayer and Pratt (1966) study, no systematic examination of bias due to refusal, ineligibility or noncontact has been undertaken and published. As stated earlier, this may be because it is very difficult to separate out refusers from noncontacts (Moore & Tarnai, 2002; Sosdian & Sharp, 1980). Whatever the reason, it is important that the various components of postal survey nonresponse are better understood if researchers are to develop and employ design mechanisms for reducing their incidence and bias (Groves & Couper, 1995). Indeed, given the considerable achievements made during the 1970s and 1980s when focus was placed on improving response from those who actually receive an invitation to participate (Best & Radcliff, 2005), it is possible that similar focus on noncontact nonresponse may lead to as yet unrealised improvements in estimates for postal surveys.

Figure 3 outlines a wide range of factors that might be expected to contribute to the survey response outcome for a given survey invitation to an individual.

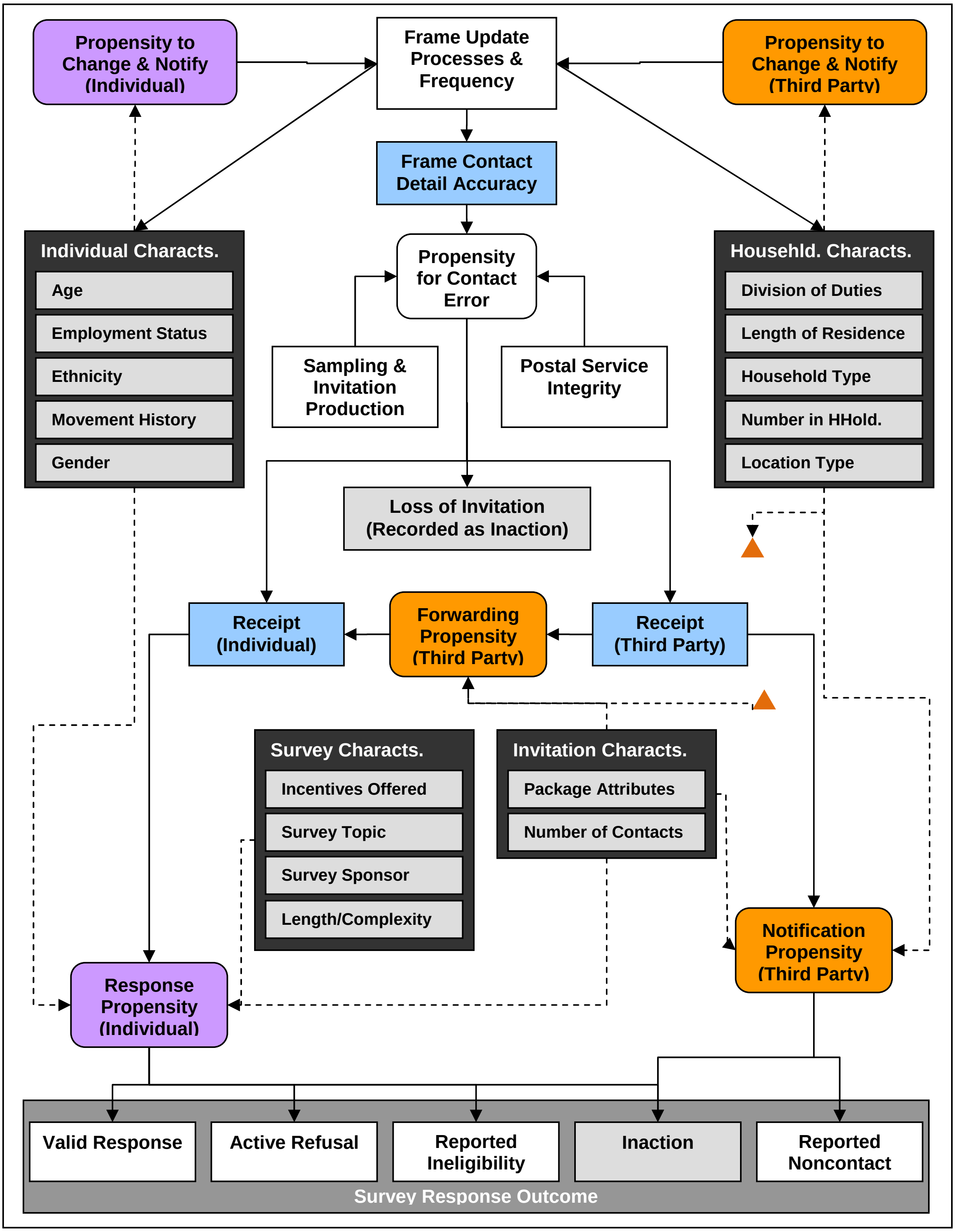

Figure 3: Conceptual determinants of postal survey response

Note: Lists of characteristics are intended to be indicative only, but many are discussed in compilations of nonresponse and survey methodology research such as Dillman (2007), Groves et al. (2002) and Groves, Fowler et al. (2004) or in published population mobility studies (e.g., Statistics New Zealand, 2007j).

As mentioned above, those factors influencing active or passive refusal and valid responses have been extensively studied, while those leading to noncontact have not. With that in mind, the following key points about noncontact may be deduced from the diagram:

- Frame accuracy is dependent on update processes and frequency, which in turn depend on third parties (typically households) or individuals to notify the frame keeper of changes. Different individuals and households may have different propensities both for changing (e.g., moving) and notifying the frame keeper of that change. To the extent that those propensities are related to individual and household characteristics, frame inaccuracies may be skewed. Furthermore, frames that are updated less frequently or actively are likely to be less accurate.
- Both sample selection and invitation production procedures may contribute to the potential for postal error. Errors may occur, for example, if a data sorting error led to names being mismatched with addresses. Similarly, the integrity of the postal service (e.g., how often it loses mail) may be a factor. For experienced researchers working with professional production and postal firms, these issues should be insignificant. Furthermore, for surveys involving multiple contact attempts, the chance that all attempted contacts would be lost should be very low.
- Where a contact error leads to third party receipt of an invitation, the third party may forward the invitation, notify the sender, or do nothing. The propensities for forwarding or notifying are likely to be related to the characteristics of both the third party (assumed to be a household in the diagram) and the invitation. Just as a survey can be considered a stimulus that an individual may or may not respond to, an addressed envelope can be considered a stimulus to which a household may or may not respond. If a household chooses not to respond, the researcher will record an inaction outcome.
- Although not included in the diagram, it is possible that some individuals who correctly receive a survey invitation send a noncontact response as a form of refusal. Also, a third party may complete and return the survey even though it was not meant for them. Given that both of these possibilities would involve deceit and effort, they are expected to make up only a small number of noncontact and valid responses.

These conceptual relationships provide useful direction for efforts aiming to investigate and address postal survey noncontact bias and informed development of the project objectives detailed below.

1.4. Project Structure and Objectives

Previous sections established that survey researchers are increasingly concerned about survey nonresponse and its associated error. Postal surveys have benefitted from this concern in that a number of techniques aimed at reducing nonresponse due to active or passive refusal have met with success. Yet, these techniques can only go so far. Many postal survey sample units are never contacted because of address inaccuracies and it is possible this introduces a nontrivial level of bias. Certainly, evidence suggests noncontact results in significant and distinct error in other modes.

This thesis therefore aimed to investigate the under-explored phenomenon of postal survey noncontact, with the ultimate goal of providing insight into how any bias it introduces may be identified and reduced. To achieve this, an empirical investigation was undertaken of nonresponse to a series of New Zealand general population postal surveys fielded between 2001 and 2006. All base surveys, of named individuals, were selected by simple random sampling or stratified random sampling from the New Zealand electoral roll and had design effects close to one. Specific high-level objectives of the project were to:

1. *Empirically estimate the levels of total noncontact present in the surveys examined and identify key correlates of both noncontact incidence (e.g., sample unit movement) and reporting (e.g., by households*).

   As outlined in Figure 3, address inaccuracies were expected to be related to individual and household characteristics, while reporting of any resulting noncontact was expected to depend on third party (household) and survey invitation characteristics. Hence, it was necessary to understand both the profile of noncontacts and the proportion of noncontact that goes unreported before an examination of noncontact bias could occur. The details and results of the study addressing these issues are presented in chapter 2.

2. *Identify the direction and magnitude of postal survey bias introduced by noncontact and compare it to error introduced by other survey nonresponse components.*

   Evidence from other modes suggests there is a difference in the nature of error from different nonresponse sources such that they contribute differentially to total survey error. If this is also true for postal surveys, then a clear opportunity exists for the development of methods aimed at targeting this source of bias. The details and results of a study examining the nature of noncontact bias are presented in chapter 3.

3. *Investigate targeted in-field mechanisms for reducing postal survey bias introduced by noncontact. In particular, examine the utility of a noncontact propensity sampling scheme for this purpose.*

   Just as in-field mechanisms such as incentives, multiple contacts, and personalisation have had success in reducing refusal nonresponse bias, it was expected that in-field mechanisms for reducing noncontact bias may bear fruit. Chapter 4 presents the results of an examination of a number of such potential noncontact-targeted design interventions. Noncontact propensity sampling was identified as the most promising of the alternatives. Hence, chapter 5 details the results of an empirical examination of that method's ability to reduce noncontact bias and, in conjunction with common post-survey weighting methods, total survey error.

Readers interested in a high-level summary of the methodologies and results of the studies mentioned above are directed to chapter 6. That chapter also outlines the limitations of the research, presents directions for future research in the area, and discusses a number of practical activities to which the key findings may be applied.

2. The Nature and Extent of Postal Survey Noncontact

2.1. Introduction

Many postal surveys source samples from population or membership registers such as an electoral roll. As detailed in chapter 1, the occurrence of address inaccuracies in these frames will likely depend on frame update processes along with individual and household movement. Furthermore, the reporting of any resulting noncontact is expected to depend on third party (household) and survey invitation characteristics. Together, these factors may lead to noncontact being both systematically underreported and unevenly distributed amongst the target population.

Although little is known about the nature and extent of postal survey noncontact, there is good reason to suspect this is true. Recent *Statistics New Zealand* research into population movement found it is related to key demographic variables such as age, ethnicity, living arrangements, employment status, and region of residence. Furthermore, prior studies examining reporting of misaddressed mail established that a significant portion goes undetected. Unfortunately, the *Statistics New Zealand* research does not indicate the degree to which movement might translate into noncontact for a given frame. Similarly, there are clear limitations to existing noncontact reporting studies, which failed to test reporting of misaddressing as it would occur in a typical survey.

The study presented here therefore examined noncontact in a general population survey of 2,400 people. It did so by exploiting a unique frame update situation to identify addresses that were likely to be inaccurate at the time the survey was fielded and comparing these with ‘gone, no address’ (GNA) returns to the survey invitation.

2.2. An Underexamined and Underreported Phenomenon

Common postal survey frames inevitably contain various inaccuracies, but one kind in particular, out-of-date address information, causes recurring misaddressing issues for survey researchers. The noncontact nonresponse that results from such inaccuracies has the potential to generate more than just a financial cost. Specifically, it may lead to erroneous cooperation rates and bias in survey estimates.

Cooperation rate inaccuracies may occur when only a portion of total noncontact due to misaddressing is reported to researchers in the form of ‘gone, no address’ (GNA) returns. This is because the remainder, unreported noncontact, is typically combined with passive refusal into an inaction category when reporting survey response, leading to underestimates of cooperation rates (Gendall et al., 2005; Sosdian & Sharp, 1980). Such a practice occurs because there is typically no mechanism for distinguishing between the two main components of inaction nonresponse. In an environment of increasing concern about declines in response across survey modes, this is of interest for two reasons.

First, it masks the proportion of postal survey nonresponse attributable to noncontact rather than noncooperation, thereby confounding analyses of the causes of declines and the efficacy of efforts to address them. Second, it hinders investigations into noncontact’s contribution to any overall bias in survey estimates. As outlined earlier (see Figure 3, p. 22), noncontact is likely to be linked to individual characteristics, while the reporting of it is expected to depend on third party (household) and survey invitation characteristics. These factors could lead to noncontact being both systematically underreported and unevenly distributed amongst the target population. To the extent that those more likely to be noncontactable differ from others in the population in their behaviour and attitudes, noncontact may be an important contributor to postal survey nonresponse error.

Indeed, there is good reason to suspect this is the case. For example, the March 2007 quarter *Survey of Dynamics and Motivation for Migration in New Zealand* (Statistics New Zealand, 2007j) reports that movement is related to age, ethnicity, marital status, living arrangements, income, employment status, occupation, and current region of residence. These key demographic variables are likely to correlate with a range of behaviours and attitudes. Furthermore, studies examining reporting of misaddressed mail have established that a significant portion goes undetected. For instance, (Hutt, 1982 cited in Esslemont and Lambourne, 1992) posted 300 deliberately misaddressed envelopes to households in the UK, of which only 68%

were returned. Similarly, Esslemont and Lambourne (1992) sent 200 misaddressed questionnaires within New Zealand of which 70% were returned. More recent research suggests underreporting continues to be an issue and that it may be worse than it has been in the past. For example, Healey and Gendall (2005) sent 1,400 misaddressed envelopes in New Zealand and received only 53% back to their 'normal' treatment (the rate rose to 67% when a 'please return' message was included on the envelope). Similarly, Braunsberger et al. (2005) found that only 41% of the 1,000 deliberately misaddressed questionnaires they mailed in the United States were returned unopened.

Two of these studies (Braunsberger et al., 2005; Healey & Gendall, 2005) examined household characteristics and their relationship to reporting of misaddressed mail. However, Braunsberger et. al. only examined gender, and did so by relying upon an assumption about the gender of the receiver that was likely to have been wrong in a number of cases (see Healey & Gendall, 2005). The Healey and Gendall (2005) study looked at a range of frame-based variables including address type, average age of electors in the household, number of elector surnames in the household, and geographic location. It found clear differences in noncontact reporting at different levels of those variables and tentatively concluded that "*identified non-contacts (i.e., 'gone no address' returns) to a single-shot mailing without an envelope message should be doubled"* to estimate total noncontact (p. 44).

Unfortunately, the population movement research from *Statistics New Zealand* does not indicate the degree to which movement translates into noncontact for a given frame. Furthermore, as Healey and Gendall (2005) noted, the misaddressing studies were limited because they involved a single mailing when, in practice, multiple follow-up mailings may be employed; a factor which may improve reporting rates. Moreover, because they sent deliberately misaddressed mail to random population samples, the studies failed to examine a realistic distribution of noncontact amongst households. Specifically, it is unlikely that misaddressing and its associated noncontact incidence occurs at random because some people (e.g., those more likely to move) may be more likely to have inaccurate address information against them in standard sampling frames. Indeed, it is also possible that a form of 'double jeopardy' exists with respect to noncontact; those more likely to be noncontactable in a sample may also be more likely to have resided in households that won't notify researchers about it. If this occurs, prior studies may have overestimated the reporting rates that can be expected in typical misaddressing situations.

The study presented in this chapter attempted to address these limitations and develop a robust understanding of postal survey noncontact incidence and reporting. This was necessary before a comprehensive examination of noncontact bias could be undertaken. Furthermore, the study sought to extend Healey and Gendall's (2005) work aimed at increasing the proportion of noncontact reported, and to explore potential mechanisms for estimating underreporting via the decomposition of the inactive disposition category.

The vehicle for the study was a general population survey of 2,400 New Zealanders undertaken in 2005. Noncontact incidence was examined by exploiting a unique frame update situation to identify addresses that were likely to be inaccurate at the time the survey was fielded. These were compared with 'gone, no address' returns to the survey invitation. Independent frame information was also used to develop profiles of individuals more likely to change addresses and third parties (e.g., households) more likely to report noncontact. Finally, the study tested a 'please return' message on the invitation envelope aimed at increasing reporting rates.

2.3. Exploring Noncontact Reporting using Frame Change Data

2.3.1. Procedural Overview

In June 2005 an age-stratified random selection of 2,400 individuals was taken from the New Zealand electoral roll for the purpose of undertaking an International Social Survey Programme (ISSP) survey on work orientation. Equal strata (of 800) were selected of those aged 18 to 34, 35 to 55, and 56 or over. The roll information had been extracted on the 30th of April 2005 and was received in early May. This sample was sent a series of postal mail invitations to participate in the ISSP survey by completing and returning a paper questionnaire. Three waves of mail were sent: an initial invitation and two reminders. Standard 'A4' envelopes were used for the initial contact and second reminder, which contained a replacement questionnaire. A standard 'Banker' envelope was used for the first reminder letter.

The survey invitations and reminders were sent between the 1st of August and the 8th of September 2005, and each sampled address was randomly allocated to a number of survey design treatments unrelated to this thesis. Respondents would only have been exposed to these if they received and opened the invitation envelope. Additionally, a split envelope message test was run. According to the procedure first tested by Healey and Gendall (2005), each sampled address was randomly allocated to one of two treatments, either an envelope with a 'please return to sender' message, or an envelope with no message. The message was centred on the bottom front of the envelope and consisted of the following statement:

> **IMPORTANT:** If this mail has not reached the intended person and cannot be forwarded, please mark the envelope "**Return to Sender**" and place it in a NZPost box.

To enable analysis of response at the household level, a number of electoral roll variables relating to individuals registered at the same sampled households were retained, including age (within 5 year band), surname and title.

A special frame update situation

The field period of the 2005 ISSP survey coincided with an enrolment update campaign undertaken in advance of the New Zealand general election to be held in mid September. As part of the data cleaning exercise, mail was sent to every eligible elector in New Zealand at their address on the roll by the *Electoral Enrolment Centre* (EEC). The mail contained a prominent *"If this isn't for you, pass me on or post me back"* message on its outer and an information update form inside. Electors who received the mail were asked to check their details. Where no changes were required, no response was to be made. Where changes were necessary (e.g., to the address, surname, or occupation), the elector was asked to send the amended form back. Extensive radio, television and print advertising also encouraged those no longer at their old address to contact the EEC via a freephone number or dedicated website. Furthermore, where a 'gone no address' (GNA) response was received from the house to which the update form was sent, the elector's details were removed from the roll, as indicated by this statement from the EEC website:

"It is easy to keep your enrolment up to date as your details change, particularly if you make a redirection order with NZ Post. We also run enrolment update campaigns from time to time. If any of our letters to you are "returned to sender" then you will have to enrol again." (Elections New Zealand, 2005)

On August 17th 2005, a fresh version of the electoral roll was published for use in the election, taking into account the changes, additions, and deletions uncovered in the enrolment update campaign. The enrolment update campaign is only undertaken prior to general and local-body elections, which occur once every three years, so the electoral roll is rarely as accurate as it was at that date.

Figure 4 presents the timing of roll updates and the survey field period visually.

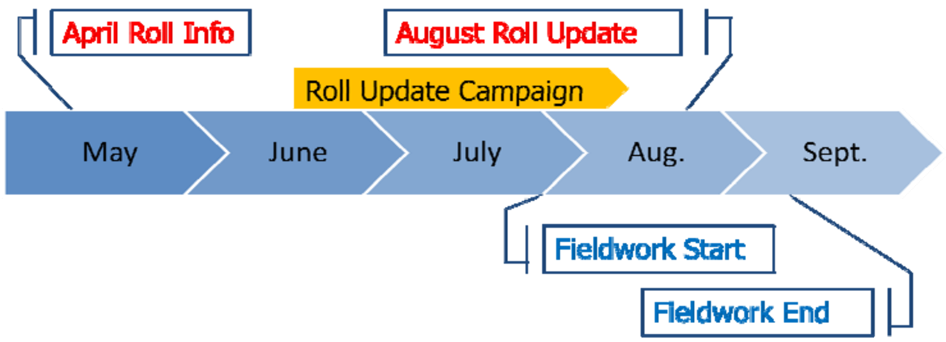

Figure 4: Timing of the frame snapshots and fieldwork for the study

Frame detail comparison

The enrolment update campaign provided a unique opportunity to compare changes in roll details for the 2,400 people selected in the ISSP sample and to undertake an analysis of the corresponding survey response profile of those individuals.

Specifically, the sample roll details obtained at the end of April 2005, and used to send the ISSP survey invitations, were compared with the details published in the pre-election roll of August 17th, around the time the ISSP surveys were in the field. Comparisons were made at an electorate level (there are 69 electorates in New Zealand). For instance, details for a person sampled from the Palmerston North electorate were compared with the updated electorate details for Palmerston North.

Comparisons were limited to this level because it was not practical to look in all 69 electorate rolls for each of the 2,400 individuals originally sampled within a reasonable time period. The effect is that those who had moved from one electorate to another within New Zealand would be classified as 'Other' rather than 'Moved' under the scheme outlined directly below. From an analysis point of view, this is unlikely to be of consequence, since any source of address change presents an opportunity for noncontact to occur. Nevertheless, these two categories are analysed both separately and in combination in the initial analysis of results presented in section 2.4.

Roll entry differences were noted where they occurred and sampled units were allocated to one of three categories:

- **Same**: Where sampled address and name information was listed the same in both the 30th April and 17th August rolls for the same electorate. This categorisation would apply to people who had not moved during the period, or who had moved but the household failed to notify EEC that the person was now 'gone no address'.
- **Moved**: Where sampled name information was the same, but address details were different between the 30th April and 17th August rolls for the same electorate. This categorisation would apply to people who moved within their electorate during the period and notified EEC of this, either through a NZPost redirection or via return of the roll update form.
- **Other**: Where the sampled name information could not be found in the 17th August rolls for the same electorate. This categorisation would apply to people who moved outside of their electorate during the period and notified the EEC of this, changed names due to marriage or deed poll, or were removed from the electoral roll due to death, incarceration, or lost contact (i.e., a GNA return).

2.3.2. Hypotheses

Based on the findings of prior studies and given the conceptual determinants of noncontact incidence and reporting outlined in Figure 3 (p. 22), the following effects were expected to be found:

1. Address detail changes and, therefore, noncontact incidence, would be correlated with movement-related demographic variables such as age, household composition, and address type.
2. Not all noncontact would be reported and reporting rates would be lower from households more likely to contain individuals who had changed address details.
3. An envelope message would improve reporting rates, as would follow-up contacts (i.e., reminder postings) to sampled individuals.
4. Despite the use of an envelope message and follow-up contacts, some unreported noncontact would remain.

2.4. Characteristics of Sample Units that Changed Details

Prior to examining the response profile of those with changed or unchanged address details, it was important to determine whether there were clear differences between the groups. Table 2 through to Table 7 present the results of this investigation based on available frame variables.

To facilitate comparisons, an ‘Any Change’ column, which simply combines the figures from the ‘Moved’ and ‘Other’ groups, is included in each table. Furthermore, to ascertain whether address type and location had an influence on reporting rates, two variables were constructed from available frame data according to the methods used by Healey and Gendall (2005).

For address type, an address relating to a rest home, hall of residence, or other group accommodation was classed as a Multi Residence. Addresses containing a Rural Delivery code or Post Office reference (e.g., PO Box), were classified as a Delivery Centre. Simple residential addresses (e.g., 10 Smith Street) were classed as ‘Residential – Whole’ addresses, while more complex addresses (e.g., 10-A Jones Street) were classed as ‘Residential – Split’ addresses, to differentiate those more likely to be family homes from those more likely to be flats. Conversely, the location variable classifications were based on the town or city of the address. Although imperfect, these classifications enabled a basic examination of differences in roll change patterns.

Readers should also note that two sets of age-related information are presented in the analysis for this section. Specifically, roll change classifications are compared both by age of individual sample units (‘Individual Age’) and the average age of electors at the address of each sampled unit (‘Average Age of Electors in Household’). The first provides insight into links between an individual’s age and their likelihood of changing address, while the second examines address change in relation to one aspect of household composition. Later sections analyse reporting of noncontact by a number of household composition variables. Hence, the household-level address change data presented here provides a foundation for assessing the patterns identified in follow-on analyses.

Table 2: Younger individuals were more likely to change details

		Roll Change Classification			
Individual Age	n	Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
18-29	518	79	8	14	*21*
30-39	471	87	7	6	*13*
40-49	421	91	4	5	*9*
50-59	368	92	5	2	*8*
60-69	315	96	2	2	*4*
70+	307	94	2	5	*6*

Note: χ^2 (5, n=2,400)=82.3, p<0.01, for ‘same’ vs. ‘any change’ by age group

Table 3: Some employment classes were more likely to change details

		Roll Change Classification			
Individual Employment Status	n	Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
Student	251	80	6	14	*20*
Not Stated	118	85	9	6	*15*
On Benefit	50	86	8	6	*14*
Employed	1,384	89	5	6	*11*
Unemployed	56	89	2	9	*11*
Homemaker	280	92	4	4	*8*
Retired	261	95	2	3	*5*

Note: χ^2 (6, n=2,400)=34.7, p<0.01, for ‘same’ vs. ‘any change’ by employment status

Table 4: People in younger households were more likely to change details

Average Age of Electors in Household	n	Roll Change Classification Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
18-29	224	74	13	13	*26*
30-39	705	87	5	8	*13*
40-49	629	88	5	7	*12*
50-59	348	94	3	3	*6*
60-69	240	95	3	2	*5*
70+	254	95	2	3	*5*

Note: χ^2 (5, n=2,400)=80.0, p<0.01, for 'same' vs. 'any change' by household age group

Table 5: People in multi-surname households were more likely to change

Surnames in Household	n	Roll Change Classification Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
One	1,498	92	4	4	*8*
Two	599	87	5	8	*13*
Three	183	75	8	17	*25*
Four	56	80	9	11	*20*
Five or more	64	73	9	17	*27*

Note: χ^2 (4, n=2,400)=72.0, p<0.01, for 'same' vs. 'any change' by household surname group

Table 6: Address type did not have a significant effect on address change

Household Address Type	n	Roll Change Classification Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
Multi Residence	27	78	4	19	*22*
Delivery Centre	164	87	6	7	*13*
Resident. - Split	473	88	6	6	*12*
Resident. - Whole	1,492	89	4	6	*11*
Rural Delivery	244	90	5	5	*10*

Note: χ^2 (4, n=2,400)=4.3, p=0.36, for 'same' vs. 'any change' by address type

Table 7: Location type did not have a significant effect on address change

		Roll Change Classification			
Household Location Type	n	Same (% row)	Moved (% row)	Other (% row)	*Any Change (% row)*
Metropolitan	1,339	89	4	7	*11*
Provincial	569	87	8	5	*13*
Rural	492	91	4	5	*9*

Note: χ^2 *(2, n=2,400)=4.3, p=0.37, for 'same' vs. 'any change' by location type*

There are clear signals in the above tables that support the hypothesis that address changes are correlated with individual and household characteristics. For instance, there is a strong linear trend in the age group data (Table 2), with younger individuals much more likely to be associated with changed address details. There is also a small increase in 'other' changes for those over 70 years of age, which is likely to relate to removal from the electoral roll due to death. Turning to employment status (Table 3), it appears students are more likely than the employed to change address details and that the employed are in turn more likely to change than retirees.

With respect to household characteristics, those who live in younger households or multiple-surname households were more likely to change address details (see Table 4 and Table 5). This makes intuitive sense, as such households are more likely to contain people with a higher propensity to move (younger individuals and renters). However, substantial differences in address change rates were not found for address type or location type. There were indications that those in multi-residence households (e.g., rest homes or university dormitories) or provincial locations were more likely to change. Even so, the number of multi-residence sample units and the difference between the address types was too small to be of practical use (see Table 6 and Table 7).

Similarly, no significant differences were found between the rates of address change by gender (10% for females, 12% for males, p>0.10) or Maori descent (14% for those indicating yes, 11% for those indicating no, p>0.10). However, the direction of difference for these variables is consistent with independent research relating to population mobility in New Zealand (Statistics New Zealand, 2007j).

In order to examine the relationship between address change and noncontact reporting, it is necessary to look at how the different groups responded to the survey request. Table 8 therefore presents the correspondence between survey response and roll detail change. Looking first at the '% of Row' breakdowns, significantly fewer 'gone, no address' (GNA) returns came from the 'Same' category than was the case for the other response classes (55% vs. at least 86% for the others).

Table 8: Survey response by roll change classification

		% of Row			**% of Column**		
Response	*n*	Same	Moved	Other	Same	Moved	Other
Valid	*1,307*	94	3	3	58	33	25
Inactive	*751*	86	6	7	30	39	37
GNA	*182*	55	16	29	5	25	35
Refused	*117*	98	2	0	5	2	0
Ineligible	*43*	86	5	9	2	2	3
Total (n)					*2,131*	*118*	*151*

Note: χ^2 *(4, n=2,400)=256.5, p<0.01, for 'same' vs. 'any change' (moved plus other) by response*

Furthermore, looking to the '% of Column' breakdowns, those identified as 'Moved' or 'Other', returned GNA responses in much greater proportions than the 'Same' group (25% and 35% vs. 5%). Thus, as expected, there is a link between address change and reported noncontact. There also appears to be a relationship between address change and inaction, with those in the 'Moved' and 'Other' categories neglecting to respond at higher rates compared to the 'Same' group (39% and 37% vs. 30%).

Although the relationship between address change and noncontact reporting is clear, it is not perfect. Half (55%) of all GNA responses come from those who did not change address details. A likely explanation for this, as suggested by the conceptual model presented in Figure 3 (p. 22), is that a number of movers fail to update their roll details themselves and the households that some of them lived in also failed to notify the EEC during the enrolment update campaign. Additionally, some people will have moved in the time between the completion of the update campaign and the publication of the updated roll. Thus, a small proportion of those who appear to have kept the same address on the roll may have actually moved.

The fact that a non-trivial portion of those whose roll details changed returned valid responses (33% of ‘Movers’ and 25% of ‘Others’ in Table 8) is also not surprising. First, some movers will have had their mail redirected to them by NZPost. Second, some will have had their mail redirected to them via alternative means, the most likely being forwarding by the current occupants of their old household. Finally, a very small proportion of people complete surveys not addressed to them (5.0% in Esslemont and Lambourne (1992) and 0.5% in Braunsberger et al. (2005)).

From a practical perspective the findings above suggest that address change can be employed as a key indicator of noncontact incidence. Furthermore, it is apparent that a relationship exists between sample unit movement (as indicated by roll modification), noncontact and GNA reporting. However, the higher inaction rates for the ‘Moved’ and ‘Other’ groups suggest that this relationship is moderated by the propensity of households receiving misaddressed mail to return it to researchers. In order to investigate the nature of this propensity for reporting, an investigation of survey response by address change and household characteristics was undertaken.

2.5. Characteristics of Third Parties Reporting Noncontact

2.5.1. Reporting in a Comparative Deliberate Misaddressing Study

Prior to examining the results from the current study, it is worth revisiting the findings of a prior study based on deliberate misaddressing undertaken in 2004 on the same general population. Healey and Gendall (2005) sent 1,400 misaddressed envelopes to a random sample of households from the New Zealand electoral roll. As tables 9 through 12 show, using data available for all respondents from the roll they found that noncontact return rates were strongly related to household composition.

Specifically, households comprising younger people (e.g., where the average age of electors in the household was 18-29) were more than 2.5 times less likely to return than households comprising older people (e.g., where the average age of electors was 70+). Furthermore, households in which the inhabitants shared the same surname were significantly more likely to return than mixed households (i.e., those with two, three, or four surnames). Turning to location-related variables, Split Address and Metropolitan dwellings were least likely to return the mail.

Note: The four tables below are reproduced from Healey and Gendall (2005).

Table 9: Households with a higher average age returned at a higher rate

Avg. Age (HH)	n	% Returned
18-29	138	30
30-39	419	49
40-49	376	63
50-59	200	69
60-69	128	77
70+	139	84
Overall	*1,400*	*60*

Note: Returns from all of the age group pairings except 40-49 and 50-59, and 60-69 and 70+ were significantly different at the 90% level.

Table 10: Single and many surname households returned at a higher rate

Surnames	n	% Returned
One	855	66
Two	320	50
Three	121	38
Four	35	49
Five or more	69	72
Overall	*1,400*	*60*

Note: Single surname households returned at a significantly higher rate than Two, Three, and Four surname households at the 90% level.

Table 11: Split address households under-returned

Address Type	n	% Returned
Multi Residence	23	83
Rural Delivery	172	76
Post Centre	86	71
Resid. - Whole	861	57
Resid. - Split	258	50
Overall	*1,400*	*60*

Note: Whole and Split Address households returned at significantly different rates to all others at the 90% level.

Table 12: Households in metro areas were least likely to return

Location	n	% Returned
Metropolitan	796	53
Provincial	410	67
Rural	194	72
Overall	*1,400*	*60*

Note: While the Provincial and Rural addresses had similar return rates, the 'Metropolitan' households returned at a statistically different rate from the others at the 90% level.

The results of Healey and Gendall (2005) provide good support for the hypothesis that households more likely to contain movers are also less likely to report noncontact. However, as noted earlier, the study involved only one wave of mailing and assumed a random incidence of noncontact across the population. These factors were likely to mean the level of overall noncontact reporting achieved was not the same as would occur in a typical survey situation.

2.5.2. Reporting in the Roll Address Change Study

An assessment of the rate of return of noncontact mail in the *Roll Address Change* study could not be undertaken in as clear-cut a fashion as that in Healey and Gendall (2005) because incidence of noncontact was not known with certainty. Nevertheless, given the results in section 2.4, address change was employed as a proxy for misaddressing so that the tendency of households to return mail across change classification could be examined.

Given the various factors contributing to survey nonresponse (active refusal, passive refusal, noncontact, and ineligibility), the correlates of address detail change identified earlier, and the findings of noncontact reporting propensity from Healey and Gendall (2005), it is worth considering what patterns might be expected in a household level analysis of return rates in the present study. First, if roll detail change is a good proxy for misaddressing, the proportion of households returning GNAs for sample units with address detail changes should be much higher than for the unchanged group. Second, if the results from Healey and Gendall (2005) generalise, there should be evidence of the patterns they found in noncontact returns (e.g., households comprising younger people being less likely to report noncontact).

It is important to note, however, that patterns are likely to be confounded by factors not present in the Healey and Gendall (2005) study. Specifically, some of the households that receive misaddressed mail will be able to forward it on to the intended recipient and, as such, it may be returned as a valid, refusal or ineligible response. Indeed, those households motivated enough to return GNAs may also be expected to forward mail if they can. Rather than focusing solely on the patterns in GNA returns, then, analysis should also examine patterns in non-return. Households associated with low rates of reporting in Healey and Gendall (2005) should show relatively high rates of non-return (as signified by the inactive category) in this study.

Another pattern of interest that could not occur in Healey and Gendall's study relates to the profile of households from which no response is received (i.e., inaction) when address details did *not* change. Some of this nonresponse will undoubtedly relate to unreported noncontact. However, the vast majority should relate to passive refusal. To the extent that the two behaviours share a similar root cause (e.g., a lower propensity for altruistic behaviour), it is likely that passive refusal patterns will be similar to those for noncontact nonreturn. That is, households associated with movers may also be more likely to contain individuals who, even if they correctly receive a survey request addressed to them, are more likely to ignore it.

An examination of tables 13 through 16 shows that the patterns hypothesised do appear in the data. For the sake of presentation clarity, the 'Responded' category in these tables relates to a grouping of valid, refusal and ineligible responses. Furthermore, given the similarity of the 'Mover' and 'Other' groups in prior analyses, these two categories have been grouped in the tables below.

Table 13: Response by household average age and roll detail status

	Address Changed				Address Unchanged			
Average Age of Electors in Household	*n*	Responded (% row)	GNA (% row)	Inactive (% row)	*n*	Responded (%row)	GNA (%row)	Inactive (%row)
18-29	*59*	27	29	44	*165*	45	8	47
30-39	*91*	26	30	44	*614*	60	4	36
40-49	*73*	40	22	38	*556*	61	4	34
50-59	*21*	33	43	24	*327*	69	4	28
60-69	*12*	42	50	8	*228*	83	3	14
70+	*13*	31	54	15	*241*	77	7	16
Overall	***269***				***2,131***			

Note: χ^2 *(10, n=269)=17.2, p=0.07, for the 'Address Changed' cells.*

χ^2 *(10, n=2,131)=101.7, p<0.01, for the 'Address Unchanged' cells.*

Consistent with expectations, households comprising younger people in the 'Address Changed' group had the highest levels of inaction and lower levels of reported GNAs.

Table 14: Response by household surnames and roll detail status

	Address Changed				Address Unchanged			
Surnames	*n*	Responded (% row)	GNA (% row)	Inactive (% row)	*n*	Responded (% row)	GNA (% row)	Inactive (% row)
One	*117*	36	26	38	*1,381*	70	4	26
Two	*79*	28	35	37	*520*	58	5	36
Three	*45*	29	29	42	*138*	51	9	41
Four	*11*	27	9	64	*45*	38	9	53
Five or more	*17*	29	53	18	*47*	45	13	43
Overall	***269***				***2,131***			

Note: χ^2 *(8, n=269)=11.1, p=0.20, for the 'Address Changed' cells.*

χ^2 *(8, n=2,131)=65.3, p<0.01, for the 'Address Unchanged' cells.*

Furthermore, although the base numbers are too small to determine clear trends, it does appear that the households associated earlier with movers (two to four surnames) generate more inaction. Also, as found by Healey and Gendall (2005), the 'Five or more' group counters this trend by having the highest GNA reporting rates of all the groups. This is likely to be due to the fact that these are often rest homes or shared residences such as student hostels, which have different processes for dealing with mail to those used by typical households.

Again, although the base numbers are small, the residential address groups (see Table 15) appear to have lower noncontact reporting rates than the other dwelling types, as evidenced by the proportion of inactive responses for the 'Address Changed' group. This is consistent with the findings of Healey and Gendall (2005). Also of interest is that the Rural Delivery and Delivery Centre address types generated higher 'Responded' rates amongst the 'Address Changers' (although not significantly so), suggesting that surveys sent to such addresses are more likely to be forwarded if misaddressing occurs.

Table 15: Response by address type and roll detail status

Address Type	Address Changed *n*	Responded (% row)	GNA (% row)	Inactive (% row)	Address Unchanged *n*	Responded (% row)	GNA (% row)	Inactive (% row)
Multi Residence	*6*	33	67	0	*21*	48	19	33
Rural Delivery	*25*	48	24	28	*219*	73	2	25
Delivery Centre	*22*	45	23	32	*142*	66	4	30
Residential - Whole	*158*	30	29	41	*1334*	65	4	31
Residential - Split	*58*	22	36	41	*415*	62	6	32
Overall	***269***				***2,131***			

Note: χ^2 *(8, n=269)=13.0, p=0.11, for the 'Address Changed' cells.*

χ^2 *(8, n=2,131)=22.5, p<0.01, for the 'Address Unchanged' cells.*

Turning to location type, metropolitan households appeared to have lower reporting rates for the 'Address Changed' group (i.e., they had the highest inaction rate), which is again consistent with the findings of Healey and Gendall (2005). However, the difference was not significant (see Table 16, below).

Table 16: Response by location type and roll detail status

Location Type	Address Changed *n*	Responded (% row)	GNA (% row)	Inactive (% row)	Address Unchanged *n*	Responded (% row)	GNA (% row)	Inactive (% row)
Metropolitan	*150*	27	32	41	*1,189*	62	5	33
Provincial	*74*	41	26	34	*495*	69	5	26
Rural	*45*	31	33	36	*447*	69	3	28
Overall	***269***				***2,131***			

Note: χ^2 *(4, n=269)=4.5, p=0.35, for the 'Address Changed' cells.*

χ^2 *(4, n=2,131)=13.7, p<0.01, for the 'Address Unchanged' cells.*

Given the findings presented thus far, the overall conclusion to be drawn is that there is a clear relationship between demographics and household characteristics, likelihood of address change, and noncontact. Furthermore, households containing people who are more likely to change address tend to report noncontact at lower rates. This ‘double jeopardy’ effect means the overall noncontact reporting rates established in prior *deliberate* misaddressing studies probably underestimate the level of underreporting that would occur in a typical postal survey of the general population.

2.6. Effect of Envelope Messages and Follow-Ups on Reporting Rates

Prior to attempting to estimate the total level of noncontact in the Address Change study, an examination of the effect of multiple waves or envelope messages on returns was undertaken. Table 17 presents response to the survey by wave of contact. Additional waves of contact substantially improved returns across all categories of response, including GNAs. Additionally, the incorporation of a ‘please return if misaddressed’ envelope message improved reporting of GNAs by over 70% (5.6% vs. 9.6%).

Table 17: Multiple waves and an envelope message increased GNA returns

	Unmessaged		**Messaged**		**Overall**	
Response	Wave 1 (% col.)	Final (% col.)	Wave 1 (% col.)	Final (% col.)	Wave 1 (% col.)	Final (% col.)
Valid	29.8	54.6	27.8	54.3	28.8	54.5
Inactive	63.8	32.7	64.0	29.9	63.9	31.3
GNA	2.8	5.6	5.2	*9.6	4.0	7.6
Refused	2.8	5.2	2.4	4.6	2.6	4.9
Ineligible	0.8	2.0	0.7	1.6	0.8	1.8
Total	100.0	100.0	100.0	100.0	100.0	100.0
Group Size (N)	*1,200*	*1,200*	*1,200*	*1,200*	*2,400*	*2,400*

* The 4.0% difference between the unmessaged and messaged treatments in final GNA returns is significant at the 95% level. No other differences in final returns between the unmessaged and messaged treatments were significant.

These results corroborate those from Healey and Gendall (2005) which also found a significant improvement in GNA return rates when the envelope message was incorporated into the study design[7]. Of note is that the message was able to elevate levels of GNA reporting beyond that achieved via the implementation of follow-up contacts. Thus, the reporting gains from the envelope message are incremental to those from multiple contacts and the two design components can be deployed together to maximise noncontact reporting. Indeed, the 70% improvement in GNA return rate suggests that the efficacy of the envelope message is much higher in typical postal surveys of the population than the 26% improvement found in the Healey and Gendall (2005) deliberate misaddressing study.

The difference in response make-up across the two envelope treatments suggests that, as would be expected, the message draws most of the additional GNAs from the inactive category (see Table 18).

Table 18: The envelope message reduced the number of inactives

	Unmessaged		**Messaged**		**Difference**
Response	*n*	% Column	*n*	% Column	n
Valid	*655*	54.6	*652*	54.3	-3
Inactive	*392*	32.7	*359*	29.9	-33
GNA	*67*	5.6	*115*	9.6	48
Refused	*62*	5.2	*55*	4.6	-7
Ineligible	*24*	2.0	*19*	1.6	-5
Total	*1,200*	*100.0*	*1,200*	*100.0*	

Although it cannot be said that the message leads to significantly fewer inactive responses, since the only significant difference is in the number of GNAs reported, the pattern does at least suggest two things.

- Even after three waves of contact there are a good number of households that do not notify the researcher of a GNA unless there is a message on the envelope

[7] Interestingly, envelope 'teasers' encouraging those to whom the survey invitation was addressed to open the envelope have also been found to substantially improve response rates to postal surveys (Dommeyer, Elganayan, & Umans, 1991).

prompting them to. Indeed, one could speculate that there will also be a good number that do not notify the researcher despite the presence of the message;

- If the message encourages anyone to return a GNA in place of a refusal or ineligible response, at most it has this effect on a handful of people.

A comparison of the proportion of valid returns across the treatments also suggests that the message does not appear to stimulate additional forwarding of mail; the valid return figures are essentially the same. Thus, it seems households who do forward on mail take such action, where possible, independently of a prompt.

2.7. A Procedure for Estimating Unreported Noncontacts

One question left unanswered is how much noncontact remains unreported despite the improvements from an envelope message and multiple waves of contact. To provide a foundation for assessing this, Table 19 presents a breakdown of final response by treatment and contact detail classification. Again, some categories have been aggregated for the sake of clarity. Specifically, valid, refusal and ineligible responses are grouped because they are all responses from individuals to whom the mailing was sent. Together, they reflect the proportion of people who received the stimulus and acted upon it. Similarly, the 'Moved' and 'Other' roll change groups described in the methodology section are collapsed here because they both represent cases with a high likelihood of misaddressing and, therefore, noncontact.

Table 19: Sample units with changed details responded in lower numbers

	Unmessaged		**Messaged**		**Overall**	
Response	Same (% col.)	Changed (% col.)	Same (% col.)	Changed (% col.)	Same (% col.)	Changed (% col.)
Valid/Ref/Inel	65.5	35.8	64.2	26.3	64.9	31.6
Inactive	31.2	43.0	29.8	31.4	30.5	37.9
GNA	3.3	21.2	*6.0	*42.4	4.7	30.5
All	100.0	100.0	100.0	100.0	100.0	100.0
Group Size (N)	*1,049*	*151*	*1,082*	*118*	*2,131*	*269*

** Messaged treatment value is significantly higher than the corresponding unmessaged value at the 95% level.*

As expected, the response profile of those with changed details was dramatically different in all treatments, with much higher reported noncontact and fewer survey responses coming from them compared to the group that had not changed details. Furthermore, the envelope message generated significant increases in the proportion of cases returned as GNAs whether or not the roll details of sample units had changed. Significant differences between message treatments did not exist for any of the other response classifications at the 95% level.

These results provide a foundation for estimating total noncontact rates because they enable decomposition of inaction into unreported noncontact and passive refusal (i.e., those who received the invitation but did not respond). By way of example, total noncontact in the unmessaged treatment (i.e., the first two columns in Table 19) is predicted to be 12%, based on the following *cross-group comparison* procedure.

First, the response rate of those who were likely to have received the invitation because their details did not change (65.5%) can be used to estimate the number of those with changed details who also received their invitation (e.g., via forwarding). Since 54 (35.8% of 151) people with changed addresses responded, and they are likely to represent approximately 65.5% of the people in that group who actually received the invitation, the total number of receivers in that group can be estimated at 82 (54 divided by 65.5%).

Second, using this to decompose the inactives for the address change group, we can predict that 28 (82 minus 54) got the invitation and chose not to respond, while the remaining 37 (43% of 151, less 28) were noncontacts. Third, adding these 'inactive noncontacts' to the reported noncontacts (21.2% of 151 equals 32, plus 37 gives 69) enables us to calculate a noncontact notification rate of 46% (the reciprocal of 37 divided by 69). In step four, the notification rate can be applied to the group of people whose address details did not change in order to estimate total noncontact amongst them at 76 people (3.3% of 1,049, divided by 46%). Finally, the estimated total noncontacts from both groups can be added (69 plus 76 gives 145) and divided by the total sample size for the treatment to find a final estimated noncontact rate of 12% (145 divided by 1,200).

Algebraically, the calculations can be represented as follows:

Equation 5: Estimated total noncontact for the 'changed' group

$$TOTNC_{Chg}=\left(INACT_{Chg}-\left(\frac{VRI_{Chg}}{RR_{Same}}-VRI_{Chg}\right)\right)+GNA_{Chg}$$

And

Equation 6: Estimated total noncontact rate (overall)

$$NCR_{All}=\left(\frac{GNA_{Same}}{GNA_{Chg}/TOTNC_{Chg}}+TOTNC_{Chg}\right)\Big/N$$

Where:

- GNA_{Chg} = Number of reported GNAs in the 'Changed' group;
- $INACT_{Chg}$ = Number of reported Inactives in the 'Changed' group;
- VRI_{Chg} = Number of Valids, Refusals and Ineligibles returned for the 'Changed' group;
- $TOTNC_{Chg}$ = The estimated total number of noncontacts in the 'Changed' group;
- GNA_{Same} = Number of reported GNAs in the 'Same' group;
- RR_{Same} = The 'responded' rate for the 'Same' group (the proportion of Valids, Refusals and Ineligibles out of the total sample size for that group);
- N = The original overall sample size (across both the 'Same' and 'Changed' groups);
- NCR_{All} = The estimated total noncontact rate across both groups, expressed as a proportion.

The same method was employed to estimate total noncontact in the messaged treatment (13%) and overall sample (13%). It was also used to generate estimates on cumulative data from only the first and then second waves of contact for the overall sample (12% in both cases). Readers are directed to Appendix section A1.1, p. 168, for information about a spreadsheet containing full workings for these figures on the thesis supplementary CD.

That the procedure yields very similar estimates across these varied design scenarios suggests it holds promise as a decomposition mechanism. Furthermore, it can be applied in a range of circumstances, provided a sub-sample is sent survey invitations using old address data so that response comparisons can be made. Many organisations retain customer address change information that would enable this on a survey-by-survey basis. Alternatively, post-hoc analyses could be undertaken after a general frame update, as was done here, to establish a notification rate to be applied to future studies.

Significant implications for survey practice arise from these findings. First, it appears that noncontact is underestimated in typical postal surveys using frames such as the electoral roll. Figures from Table 17 (p. 45) suggest estimated total noncontact is as much as 400% higher than the reported level in a single-contact unmessaged study (2.8% vs. the estimated 12% established above). Indeed, even in a study with three contacts and an envelope message, total noncontact is likely to be more than 30% higher than reported (9.6% vs. an estimated 13%). The cooperation rates for many postal surveys are therefore likely to be understated.

Second, the results suggest noncontact is a much larger component of total nonresponse than generally acknowledged. Given widespread concern about declining survey response, this is important to know. Efforts aimed at understanding the reasons for declines, identifying any associated bias, or developing tools to combat the problem, all require knowledge of the size and nature of nonresponse components. The demographic comparisons, envelope message technique and notification rate estimation procedure outlined here work to generate that knowledge.

The *cross-group comparison* procedure developed above relies on two key assumptions:

1. That the total proportion of noncontact in the ‘Same’ group (those who have not changed address) is small enough to have minimal impact on the ‘responded’ rate calculated for that group, and
2. That the response rates amongst those who receive a request, and noncontact notification rates for households that receive a misaddressed envelope, remain constant across the ‘Same’ and ‘Changed’ groups.

The first assumption appears reasonable given the results in Table 19. Unfortunately, the second assumption is untestable; all that can be said given the results presented in this chapter is that the response and notification rates in both groups are greater than zero. However, there is no obvious reason to suspect a substantial difference in either rate between groups.

2.7.1. An Alternative Estimation Procedure

There are some situations where the first assumption above cannot be expected to hold. For example, if an attempt was made to estimate 'responded' rates for the 'Same' group on some subpopulations from the sample, the rate may become very sensitive to the number of GNAs in the group. This may happen because cell sizes become too small. It also could occur if the subpopulation itself is defined on a variable highly correlated with noncontact (such as age). In such cases, even the 'Same' group may contain a relatively high proportion of noncontact and, as such, a reliable base 'responded' rate will not be calculable.

In these situations, it may be prudent for researchers to use an alternative, although less robust, estimation procedure. One approach would be to calculate unreported noncontacts at an aggregate level by splitting the inactives according to a simple ratio of responders (i.e., valids, refusals and ineligibles) to reported noncontact. For example, if 1,000 people were surveyed and 450 gave some form of response while 500 gave no response at all (inactives) and 50 were returned GNA, then the total noncontact rate would be estimated at ≈10% according to the formula below.

Equation 7: Estimated total noncontact rate (Iceberg method)[8]

$$NCR_{All} = \left(GNA_{All} + \left(\frac{GNA_{All}}{VRI_{All} + GNA_{All}} * INACT_{All}\right)\right) \Big/ N$$

[8] The estimate calculation can be simplified to $GNA_{All}/(VRI_{All}+GNA_{All})$. However, the full equation is presented above to more clearly express the logic of the procedure.

Where:

GNA_{All}	= Number of reported GNAs overall;
VRI_{All}	= Number of Valids, Refusals and Ineligibles returned overall;
$INACT_{All}$	= Number of reported Inactives overall;
N	= The original overall sample size;
NCR_{All}	= The estimated total noncontact rate overall, expressed as a proportion.

The assumption here is that, much in the same way an iceberg tip indicates the size of the underlying structure, the size of the 'responded' group indicates the proportion of people that actually received the survey invitation while the size of the reported GNAs indicates the proportion of the total sample that were noncontacts. One benefit of this approach is that, because it does not rely on any particular group having a trivial level of noncontact, it would not break down in subpopulation analyses.

Table 20 presents the results of a comparison of total noncontact estimates for the current study made under both the *Cross-Group Comparison* and *Iceberg* procedures.

Table 20: Total noncontact by estimation method, treatment and wave

		Survey Treatment		
Method	**Wave**	Unmessaged (%)	Messaged (%)	Overall (%)
Cross-Group	1	10.9	13.0	11.9
	1,2	10.8	13.8	12.3
	1,2,3	12.0	13.4	12.8
Iceberg	1	7.6	14.3	11.0
	1,2	8.5	12.8	10.7
	1,2,3	8.3	13.7	11.0

As might be expected, the *Cross-Group* method appears to be more robust in the current situation; its estimates are more consistent across treatments. Certainly, the *Iceberg* method appears to underestimate total noncontact rates in the unmessaged treatment. A potential contrary issue with the *Iceberg* method is that it would be susceptible to poor survey design. Specifically, if a survey were to have a very low response rate, the correspondingly low ‘responded’ rate in the estimation calculation could lead to noncontact being *overestimated*. Researchers should therefore apply the procedure with these issues in mind.

3. Noncontact's Contribution to Nonresponse Error

3.1. Introduction

The Address Change study in chapter 2 found that noncontact is a larger component of postal survey nonresponse than typically recognised, that it appears related to population movement, and that it occurs disproportionately amongst a subset of the population. Nevertheless, it is not necessarily true that noncontact contributes error to survey estimates. Indeed, even if it does, it is possible that any error is either the same as, or entirely offset by, that contributed by other nonresponse components (see Groves, 2006, for a formal discussion of the possible interaction of component biases). It was therefore important to develop an understanding of the error introduced by the various components of postal survey nonresponse prior to directing effort toward targeting noncontact as a specific source.

Unfortunately, nonresponse bias is a notoriously difficult phenomenon to examine because it arises from missing data. A range of techniques are therefore employed by methodologists interested in it, depending on the external data available to them or the auxiliary information able to be collected as part of fieldwork. Because most involve assumptions that are untestable, researchers often rely on internal consistency arguments to support insights into nonresponse bias. General consensus is that multiple techniques and replication studies should therefore be used wherever possible to examine the phenomenon from different perspectives and provide a solid foundation for any conclusions drawn.

Certainly, prior studies exploring postal survey nonresponse bias have employed a variety of techniques to assess the error-reducing effects of field efforts aimed at improving response (e.g., incentives or multiple contacts). Most have examined changes in estimates over waves of contact or compared survey results against known frame information and population parameters (e.g., from census data). In addition to highlighting the weaknesses of individual methods of examining nonresponse bias, these studies have demonstrated that improving response does not necessarily reduce error. Furthermore, the only study to explore postal nonresponse bias at the component level found marked differences in the contribution made by noncontact and refusal (Mayer & Pratt, 1966). As a result, the authors urged researchers to further consider the interplay between the sources of nonresponse and survey error.

The study presented here therefore sought to identify the direction and magnitude of postal survey noncontact bias and compare it to error introduced by other nonresponse components. It established estimates of bias due to noncontacts, active refusals, ineligibles and inaction for a selection of general population surveys fielded between 2001 and 2006. Multiple techniques for estimating error were employed, including benchmarking against population parameters, comparisons on individual-level frame data and analysis of valid responses over time.

3.2. Approaches to Evaluating Postal Survey Nonresponse Bias

The potential for examining nonresponse bias is moderated by a number of factors including external data availability, financial resources, and survey mode. Hence, a range of approaches to bias estimation have been developed. It is generally acknowledged that no one approach is able to give a full picture of the potential error and so, where possible, multiple techniques should be employed to enable a comparative analysis.

3.2.1. General Nonresponse Bias Assessment Methods

Groves and Brick (2006) outline four general categories of nonresponse bias assessment methods, as outlined below. Each approach uses different tools in an attempt to measure bias, but all ultimately aim to measure the degree of covariance between propensity to respond and the value of key survey variables. Not all can be applied in postal surveys. However, the full range of methods is briefly discussed here to provide a context to the review of postal survey specific studies to follow.

Benchmarking

Under this approach, results from a study with nonresponse are compared with those of recent, independent studies that had very high response rates. Examples include figures from a census or a well-resourced government survey. Such comparisons are often easy and inexpensive to undertake. However, benchmarking data may suffer from errors of measurement, coverage, and nonresponse that must be considered when employing them as a ‘gold standard’. Furthermore, the variables in common between the studies may be limited and unrelated to the key items of interest, thereby reducing the utility of making the comparison.

Measurement against external individual-level data

Where high-quality data are available, robust estimates of nonresponse bias with respect to a select set of variables can be achieved with these methods:

- *Information available on, or able to be matched to, the sampling frame*
  Many frames contain age, gender, and location details for each individual. Some organisational lists also hold information on items such as length of membership and products purchased. Furthermore, it may be possible to match data from other sources to individuals on a frame. As this information is known whether or not a subject responds, analysis can be undertaken to assess the degree to which respondent values differ to nonresponders on those variables. An example of this approach can be seen in Lin and Schaeffer (1995).

- *Observational data collected during fieldwork*
  A field force can collect information about subjects or households approached that can be utilised in a nonresponse study. For instance, in a face-to-face survey, the gender of refusers could be captured, as could the dwelling type and other overt characteristics of all households (e.g., see Lynn, 2003).

- *Response from an ‘add-on’ sub-sample for which external data are available*
  Because external information is available for all members of the sub-sample, response differences amongst that group can be extrapolated to the full sample for which external information is not known. An example of this method can be seen in Groves et. al. (2004).

These approaches allow accurate estimates of bias due to nonresponse to be made, at least with respect to the external variables available. Furthermore, any relationships found between the external variables and response propensity can be extended to respondents' answers to key survey variables in an effort to assess whether bias was likely to have been introduced. However, the techniques are limited by the type and level of external variables available. In some cases the external variables may not relate well to either response propensity or the key variables under examination when, ideally, they would relate to both. Furthermore, if the external variables themselves are subject to missing data or a high level of measurement error, their efficacy for nonresponse estimation may be compromised.

Examining internal variation within the data collected

It is often not possible to obtain individual-level external data. Hence, many studies attempt to extrapolate nonresponse error from differences observed within responses received over time or across sub-groups. The most common techniques are:

- *Comparison of response rates by sub-group*

  A simple mechanism for considering whether nonresponse bias exists is to group sample units according to common demographic variables (e.g., age, gender) and compare their levels of response. Where each group responds at roughly the same rate, it is assumed that bias has not occurred. Where one or more groups under-respond, a comparative examination of their response distributions for key survey variables is then undertaken. If the response distributions differ, bias is said to exist and post-survey adjustments are made.

- *Use of screening or prior-wave data from multi-stage studies*

  This approach is similar to the 'observational data' technique described earlier. Effort is made to maximise response to a limited set of screening questions. Variables from the screening data are then used to examine the characteristics of nonrespondents in the second stage of the study. Longitudinal studies may employ this technique by matching data to both responders and nonresponders from a prior round of measurements in which members of both groups responded.

- *Nonresponse follow-up studies or nonresponse experimentation*

  Here, extended fieldwork efforts (e.g., additional callbacks, incentives, alternative contact modes) are made to get information from nonrespondents. These can be undertaken during the original survey or as part of a 'follow-up' study of nonrespondents. Values obtained during extended efforts are assumed to be representative of nonrespondents and are therefore used to assess total bias. A slightly different approach involves randomised experiments during fieldwork which vary design elements thought to affect response. The different treatments achieve different response rates and the survey estimates achieved under each treatment are then compared to estimate the likely effect of nonresponse in the lower response treatments (e.g., see Groves et al., 2005).

- *Examination of variation by level of recruitment effort or wave of response*
  This type of analysis involves a standard field operation and is commonly employed as a post-hoc bias assessment technique. Once fieldwork is completed, nonresponse is estimated using variation in respondent values by effort required to generate a response (e.g., number of callbacks required). The assumption made is that 'easy to get' respondents differ from those that are 'hard to get', and that the 'hard to get' respondents are similar to those who do not respond at all. Bias is estimated by extrapolating any trends identified along this 'continuum of resistance'. Lynn and Clarke (2002) and Craighill and Dimock (2005) represent examples of the application of this technique. Furthermore, Colombo (2000), Filion (1975) and Armstrong and Overton (1977) outline different methodological processes for this class of nonresponse analysis.

Because they do not require access to an independent data source, these methods can be applied in a wide range of studies. Furthermore, many are low cost and therefore meet the restrictions of a variety of funding situations. However, the 'Achilles heel' of these approaches is that they depend heavily on one underlying assumption: that all nonresponders are well represented by respondents who were 'difficult to get'. Most of the time, it is impossible to test this assumption. However, where it has been tested it does not always hold (e.g., see Lin & Schaeffer, 1995).

Contrasting alternative post-survey adjusted estimates

The fourth methodological category discussed by Groves and Brick (2006) involves post-survey adjustment for nonresponse under different weighting models. Methods can range from simple post-stratification on one or more variables, through to propensity modelling employing combinations of characteristics and imputed results.

The aim is to conduct bias sensitivity analyses under different assumptions about the differences between respondent and nonrespondent values, given the response rates of the study. Differences between various estimates of bias and the unweighted survey estimates are used to indicate the likely extent of error in the sample statistics. This approach employs the same type of data as the methods in the previous section. It also ultimately rests on the same assumption that variation in values amongst respondents of different classes can be extrapolated to those for whom no responses were collected.

3.2.2. Methods and Results of Published Postal Survey Bias Studies

The bulk of the literature dealing with nonresponse in postal surveys focuses on improving response rates. A number of meta analyses, literature reviews and texts document the variety of such studies undertaken (e.g., see Dillman, 2000; Kanuk & Berenson, 1975; Mangione, 1995; Yammarino, Skinner, & Childers, 1991) and there is general agreement that follow-up contacts, incentives, and type of postage used are effective in this regard. Ultimately, efforts to improve response aim to reduce nonresponse bias. It is surprising, then, that few studies have sought to empirically estimate postal survey nonresponse bias, let alone its relationship with component sources or response rates. Those that have typically use wave analysis and external individual-level data to estimate bias. A small number employ benchmarking or simulation studies, and some early studies used double sampling (nonrespondent follow-ups) to generate comparative data.

Reid (1942) employed wave analysis and double sampling techniques in one of the first studies to attempt an exhaustive examination of postal survey nonresponse bias. Specifically, in his study of radio use in schools he achieved a 67% response rate with one follow-up mailing. Alternate contact methods were then used to achieve a 95% response from a sub-sample of nonrespondents to the initial study. He found statistical differences between respondent answers to questions across the three response groups and suggested that researchers should not assume that multiple-contacts reduce bias to a trivial level.

Reid also cites six other mail studies (Reid, 1941; Rollins, 1940; Shuttleworth, 1940; Stanton, 1939; Suchman & McCandless, 1940; Toops, 1926) reporting the results of nonrespondent follow-ups. A number of those found significant differences between estimates from the initial returns and those from follow-up efforts, leading Reid to conclude that "*replies from respondents cannot be considered representative of non-respondents*" (1942, p. 90).

In an effort to generate estimates of nonrespondent values in the inevitable absence of data for nonresponders, researchers have examined the assertion that trends in responses over successive waves predict at least the direction, if not magnitude, of nonresponse bias (Filion, 1975; Pace, 1939). In addition to wave analysis techniques, these studies employ external data or the inclusion of specific questions to provide another perspective on error due to nonresponse. The results do not lead consistent support to the utility of wave analysis as a bias estimation procedure.

For example, Clausen and Ford (1947) anticipated a problem in employing wave extrapolation because their veteran population was undergoing significant changes in employment status at the time of their study. By incorporating an anchoring question in the survey they were able to show that, had the problem not been anticipated, wave analysis would have suggested substantial differences in response by status across waves when no such difference existed.

Furthermore, both Mayer and Pratt (1966) and Lankford et al. (1995) used external data to show that although wave analysis would have suggested no difference between respondents and nonrespondents in their study, significant differences did in fact exist. Finally, Ellis et al. (1970) report that estimates across respondent groups as measured by external data did not follow a linear pattern, as would be expected under a 'continuum of resistance' model. They conclude that "*late respondents do not provide a suitable basis for estimating the characteristics of nonrespondents*" (p. 108). Interestingly, although Reid (1942) did not attempt wave extrapolation in his study, there is also some evidence of nonlinearity across respondent groups in the results he presents (see p.92).

Armstrong and Overton (1977) attempted to overcome the practical limitations of extrapolation across waves by introducing a judgemental procedure to inform the process. Using data from 16 prior studies with multiple waves of contact, they examined judges' ability to predict the direction of difference between first and second wave responses to a range of items. When combined with an extrapolation method to predict the direction of bias in a third wave, they found that the use of judgemental input helped "reduce major errors", but noted that this was at the expense of "an increase in the percentage of items overlooked" (p. 399). That is, the procedure helped identify situations where a non-linear bias relationship was likely to exist across waves, and for which linear extrapolation should therefore not be performed. However, it also generated a number of false positives.

Turning to estimation of bias *only for those items judged to be likely to exhibit linear changes* across waves, Armstrong and Overton examined a range of extrapolation techniques. They found, unsurprisingly, that any extrapolation from the first two waves generally lead to better estimates of final sample means (i.e., incorporating third wave response) than no extrapolation at all. The study is open to criticism because it used incomplete data as its 'gold standard' for testing whether bias was adequately mitigated by extrapolation. However, even if the findings are valid, it appears wave extrapolation as a bias estimation technique remains fraught with subjectivity and limitations in its application.

Another set of postal survey nonresponse bias studies have approached the problem as part of an examination of the effect of response inducements, or sub-group response rates, on estimates. These studies typically employ external data from the frame or a prior survey along with experimental manipulation of survey design. For example, Jones and Lang (1980) looked at how sponsorship, cover letter message, notification method and questionnaire format influenced response rates and survey estimates. Comparing estimates to external data for each individual in their study, they found that improving response rates through design manipulations can differentially draw in respondents that actually contribute to an *increase* in nonresponse bias.

Using frame data, Moore and Tarnai (2002) also found that incentives exacerbate composition differences amongst various respondent groups. However, the differences were not large, and the low sample size of their non-incentive treatment limits the certainty of the finding. In contrast, Shettle and Mooney (1999) used prior survey data available for all respondents to measure bias and found that it was lower for the incentivised group in the mail component of their study, but not significantly so. Furthermore, they note that incentives did not appear to differentially attract certain subpopulations and concluded that it is "*reasonable to assume that increasing response rates through the use of incentives will lead to a decrease in nonresponse bias*" (p. 242). Taken together, the evidence regarding the bias mitigating effects of efforts to increase response rates in postal surveys via incentives is, at best, inconclusive.

Conversely, results relating to multiple contacts suggest that extra efforts in that area can reduce overall bias. Specifically, studies reporting results by wave of response and making comparisons to independent data (Clausen & Ford, 1947; Ellis et al., 1970; Filion, 1975; Mayer & Pratt, 1966; Reid, 1942; Shettle & Mooney, 1999) typically show improvements in cumulative estimates over successive waves. In addition, the incorporation of multiple contacts provides a foundation for wave extrapolation; a technique which can *sometimes* be effective in estimating and adjusting for bias.

One conclusion that can clearly be drawn from prior postal survey studies is that efforts to improve response can only go so far in reducing nonresponse bias. For example, the studies cited directly above overwhelmingly find that nonresponse bias remains in postal surveys even when multiple contacts are incorporated. Furthermore, it is not clear that implementation of incentives, one of the few design features found to consistently improve response rates over and above repeated contacts, can reduce bias. Hence, if progress is to be made in nonresponse bias reduction in postal surveys, efforts must focus on a) the development of more robust extrapolation procedures and b) careful attention to differential design manipulations targeted at managing response at a group, rather than survey level. Yet, for these efforts to proceed, a clearer picture is required of the contribution made to overall bias by different components of postal survey nonresponse.

3.2.3. How Component-Focused Studies may Contribute

It is fair to say that nonresponse research as it relates to the mail mode has focused almost exclusively on bias due to active or passive refusal. Where design manipulations were examined in studies cited here, they related to efforts that could only work if the respondent was contacted. Furthermore, noncontact did not receive consideration either because the population studied was not prone to it (Ellis et al., 1970; Reid, 1942), it was incorporated into an overall nonresponse group (Armstrong & Overton, 1977; Jones & Lang, 1980; Moore & Tarnai, 2002), or it was noted in response figures but its contribution to bias not examined (Clausen & Ford, 1947; Filion, 1975; Shettle & Mooney, 1999).

Filion (1975) did take time in his study of water fowl harvest to note that "the trend observed over successive cumulative waves revealed a tendency for surveys with a low response rate to underestimate the number of deceased persons and unclaimed letters" (p. 490). However, only one study (Mayer & Pratt, 1966) undertook a detailed examination of the effects of noncontact and refusal on estimates. They concluded that:

> *"the biases introduced are not similar. In fact, for 3 out of the 7 characteristics considered, the biases are* offsetting*. For the others, there are marked differences between the two nonresponse groups. Accordingly, in evaluating the potential seriousness of nonresponse bias, as well as in prescribing a weighting scheme, we feel that an independent examination of both size and character of the major nonresponse segments provides the analyst with a far more meaningful approach than does the conventional reliance on over-all nonresponse rates alone. For example, the present findings demonstrate that the practice of excluding undeliverable questionnaires from the sample frame could lead to ignoring a serious bias source."* (p. 644)

It appears likely that the same disparity between refusal and noncontact nonresponse found in face-to-face and telephone modes (e.g., see Lynn & Clarke, 2002; Stinchcombe, Jones, & Sheatsley, 1981) is also present in the postal mode. If it is, then, as Mayer and Pratt (1966) noted, the opportunities for improving estimates afforded by component-based studies of nonresponse may be significant:

> *"Inasmuch as the biases tend to be offsetting for certain characteristics, the researcher who has carefully segmented nonresponse by source could minimize total nonresponse bias by (1) controlling the relative sizes of offsetting nonresponse segments, or by (2) applying differential weights based on the relative sizes of these segments."* (p. 644)

Additionally:

> *"If the nature of an individual's involvement in the subject matter of the survey underlies his motivation to respond, motivation, in turn, provides a useful approach to explaining (or predicting) the distribution characteristics of those who refuse. Biases introduced by nonrecipients of a questionnaire tend to coincide with characteristics of the mobile portion of the population being studied. As long as the relative sizes of the nonresponse groups are known, and as long as the directions of bias can be evaluated through knowledge about the motivations of the "refusers" and the characteristics of the "mobiles," meaningful techniques can be developed to adjust for possible nonresponse bias."* (p. 645-646)

3.3. An Empirical Analysis of Postal Survey Nonresponse Bias

Given the potential for a component-focused approach to generate improvements in postal survey accuracy, and the substantial contribution made by noncontact to total survey nonresponse established in chapter 2, a decision was made to further examine this error source. The following study therefore examined the direction and level of postal noncontact bias in comparison to that introduced by the other nonresponse components, across a number of completed surveys. Multiple techniques for estimating error were employed.

3.3.1. Procedural Overview

The Surveys Analysed

This study used data collected from the following general population surveys of named individuals undertaken by the *Department of Marketing* at *Massey University* between 2001 and 2006. All sourced their samples from the New Zealand electoral roll and were fielded as part of the *International Social Survey Programme* (more information at www.issp.org). Although some involved stratified random sampling, the estimated design effects for all surveys were close to one (see section 4.4.2, p. 103 for details). Sample questionnaires from each study have been included on the thesis supplementary CD. See section A1.2, p. 168, for more information.

- *"Social Networks in New Zealand" (2001)*
  Covering a range of questions related to group membership, friendships, support networks and socio-demographics, this survey was sent to a sample of 2,200 people. The sample was selected at random, without stratification, from a copy of the electoral roll extracted during 1999. Fielded from August to October 2001, the survey had one invitation and three follow-up postings.

- *"The Roles of Men and Women in Society" (2002)*
  Covering a range of questions related to attitudes to women working, sharing of home responsibilities, financial arrangements, work-life balance and socio-demographics, this survey was sent to an initial sample of 2,075 people. The sample was selected at random, without stratification, from a copy of the electoral roll extracted during 2000. Fielded from August to September 2002, the survey had one invitation and three follow-up postings.

- *"Aspects of National Identity" (2003)*
  Covering a range of questions related to personal identity, group affiliation, nationalism, social or political views and socio-demographics, this survey was sent to an initial sample of 2,200 people. The sample was selected at random, with 100% over-sampling of those on the Maori roll (15% instead of 7.25%), from a copy of the electoral roll extracted during 2002. Fielded from September to November 2003, the survey had one invitation and three follow-up postings.

- *“New Zealanders’ Attitudes to Citizenship” (2004)*

  Covering a range of questions related to democratic process, rights, political activity, government corruption and socio-demographics, this survey was sent to an initial sample of 2,500 people. The sample was selected at random, without stratification, from a copy of the electoral roll extracted during 2004. Fielded from June to August 2004, the survey had one invitation and two follow-up postings.

- *“New Zealanders’ Attitudes to Work” (2005)*

  Covering a range of questions related to work history, job satisfaction, work-life balance, job security and socio-demographics, this survey was sent to an initial sample of 2,400 people. The sample was selected at random within three age-bands (18-34, 35-55, 56+) of 800 people, from those less than 90 years of age with a New Zealand address in a copy of the electoral roll extracted during 2005. Fielded from August to October 2005, the survey had one invitation and two follow-up postings.

- *“The Role of Government” (2006)*

  Covering a range of questions related to democracy, government responsibility, politics and socio-demographics, this survey was sent to an initial sample of 2,250 people. The sample was selected at random within six age/sex bands (male/female by 18-34/35-55/56+) of 375 people each, from those with a New Zealand address in a copy of the electoral roll extracted during 2006. Fielded from August to October 2006, the survey had one invitation and two follow-up postings.

For each survey, the following information was available for analysis:

- Demographic variables from the frame including age, gender, location, occupation, and the same information for other registered voters at the same address.
- Response disposition information: including type of response (GNA, valid response, refusal, ineligible or inaction), date of response, and whether a reminder had been issued.
- Survey item information for those sample units that returned a valid response to the survey request.

Bias Estimation Methods Employed

Because survey fieldwork had long finished, double-sampling was not suitable as a bias estimation method in this study. However, the other methods commonly employed in postal survey bias studies to date were utilised. They were:

- *Benchmarking*

  Parameters from the national census from 2001 and 2006 relating to variables included in the survey were compared with survey estimates to indicate direction and magnitude of bias.

- *Comparisons against individual-level external data*

  Values from the frame were compared across all response groups to provide estimates of direction and magnitude of nonresponse bias on those items. Correlations between these variables and answers to survey questions were also examined to indicate bias on items not contained in the frame.

- *Extrapolation of wave results*

  Extrapolation of cumulative survey estimates by wave was employed on a number of items in an attempt to generate indications of the direction and magnitude of bias. The results of the extrapolation were compared against census variables to assess the efficacy of the technique.

3.3.2. Key Hypotheses

The study sought to estimate and characterise the bias introduced by different components of postal survey nonresponse. It also aimed to examine the contribution each component made to net nonresponse bias. Given the findings reported by Mayer and Pratt (1966) and the results presented in chapter 2, it was hypothesised that the following would be the case:

1. Despite multiple contacts, net nonresponse bias would exist in the final estimates for the surveys and variables examined.

2. Successive waves of contact would reduce net nonresponse bias in the surveys and variables examined.

3. Each component source would contribute a different profile of nonresponse bias to the studies and variables examined. Specifically, the sources would affect different variables, or the same variables in different ways. Noncontact was expected to relate to mobility, while refusal was expected to relate to motivation and survey topic (Brennan & Hoek, 1992; Mayer & Pratt, 1966).

4. Where component sources affect the same variable, a net bias would still arise. That is, in most cases the components would not cancel each other out completely, even if there was some limited offsetting effect.

In addition, the following speculative hypothesis was proposed:

5. Extrapolation of wave trends would improve estimates but not adequately account for net bias in the survey estimates. This is because, in a postal survey situation, 'hard to contact' people do not respond in any wave and, so, no waves of contact will contain information on these sample units that can subsequently be extrapolated.

3.4. Response and Bias Trends across Multiple Postal Surveys

Table 21 presents a breakdown of response to the six studies under examination. Although ineligibility is stable across the studies, the other disposition categories vary. There was an apparent decline in GNAs as a proportion of total response over the years. However, in 2001 and 2002, the rolls used to select the samples were at least one year old. Furthermore, although the 2003 sample was also taken from an older roll (i.e., a copy from 2002), that roll was from an election year and, as such, would have been more accurate than those from prior years. Finally, the 2005 and 2006 surveys employed stratifications on age that increased the proportion of younger people in the sample relative to the population, which was likely to have led to a slight increase in the level of underreporting of noncontact.

Table 21: Response to the six ISSP surveys

Response (%)	Issp01	Issp02	Issp03	Issp04	Issp05	Issp06
Valid	52	49	47	54	54	56
Inactive	27	31	35	34	31	34
GNA	12	14	10	8	8	5
Refused	6	3	4	2	5	3
Ineligible	3	3	3	2	2	3
Total	*100*	*100*	*100*	*100*	*100*	*100*

Note: All figures represent percentages of the column.

In order to examine whether the estimates from valid responses to the ISSP surveys contained errors, a selection were compared to figures from the 2001 and 2006 census (see Table 22). Overall, the unweighted results from the surveys appear to consistently underestimate the proportion of males, overestimate the number of older people, and underestimate the proportion of people of Maori ethnicity in the population (see footnote 9, next page).

Furthermore, estimates of marital status, qualifications, income, and household size appear to contain consistent error when compared to census. However, the figures do not suggest whether this bias could be due to noncontact, refusal, or some other error source. For example, where questions in the surveys and census were not presented in exactly the same form, any differences found may be due to measurement error. Furthermore, although the electoral roll enjoys a high enlistment rate (approximately 95% of the eligible population, according to F. Thompson, 2007) because of a legal requirement to enrol, it is possible that some of the error presented in Table 22 is due to incomplete coverage. Indeed, a combination of error sources is likely to be the cause of underestimates for variables such as marital status. That variable had different wording in the ISSP from the census and may also have been subject to both noncontact and passive refusal nonresponse bias as those who are not married may spend less time at home and be more likely to move.

Table 22: Unweighted survey estimates compared to census figures

	ISSP Survey Estimate						Census Result[9]	
Variable	2001	2002	2003	2004	2005	2006	***2001***	***2006***
% Male	43	43	45	44	46	48	***48***	***48***
% 20-34 Years old	18	19	19	20	*25	*23	***29***	***28***
% 65+ Years old	21	20	24	23	18	22	***17***	***17***
% Maori Ethnicity	9	9	^17	11	10	11	***11***	***11***
% Marital: Single	18	17	19	19	19	21	***31***	***31***
% Bach/PG Qual	14	18	16	21	19	21	***10***	***14***
% Income <$20k	41	34	40	39	34	34	***49***	***39***
% Income > $50k	16	21	20	22	24	27	***13***	***20***
% Not Religious	26	29	26	29	34	33	***28***	***33***
% Employed Fulltime	44	46	47	46	49	47	***46***	***48***
% 1 Person HH	13	12	13	12	11	12	***23***	***23***
% 5+ Person HH	14	12	14	13	14	12	***12***	***12***

* The 2005 and 2006 ISSPs contained stratification by age.

^ The 2003 ISSP oversampled from the Maori roll

Therefore, in an attempt to isolate error due to nonresponse, data from the frame were used to compare values for those who returned a valid response against the values for the entire sample. The use of this independent data excludes measurement, coverage or sampling error as potential causes of any differences found. Table 23 presents the results of the comparison in the form of percentage deviation values for each survey and item.

[9] Readers are directed to Appendix section A2.1, p. 174, for a background to the census figures presented in the table above. As discussed in the Appendix, although the survey estimates for '% Maori Ethnicity' and '% Not Religious' appear to track the census figures fairly closely, differences in base populations between the survey sample and underlying census data, along with measurement differences, are likely to mean that these variables are in fact consistently underestimated by ISSP survey returns.

Table 23: Percentage difference between valids and the full sample on frame data

Frame Variable	Issp01	Issp02	Issp03	Issp04	Issp05	Issp06
Average Age	4	5	7	6	4	7
% Male	-10	-6	-5	-8	-4	-4
% Maori Descent	-23	-18	-19	-21	-18	-16
% Employed	3	3	2	3	8	2
% Student	-23	-29	-21	-19	-28	-17
% Retired	6	21	20	19	-2	22
Avg. Age (HH)	3	4	6	4	2	5
Avg. # Electors (HH)	-16	-14	-14	-7	-8	-5
Avg. # Surnames (HH)	-27	-24	-21	-12	-16	-9

Note: Figures represent the values for valid respondents minus those for the overall sample, divided by the overall sample value. Thus, a negative figure indicates that the valid group under-represented the sample on the variable of interest by x%.

Consistent with the trends presented in Table 22, the proportion of males and people of Maori descent[10] in the sample are underestimated by the valid respondent group. Similarly, age is overestimated. Furthermore, there are compositional differences between the respondent group and the total sample on occupation and household measures.

The fact that the direction of many results in Table 22 and Table 23 are uniform across both estimation methods and survey instances gives strong support to hypothesis one. Specifically, it appears that bias in some survey estimates (i.e., at least those relating to sample demographics) does exist even after multiple contacts have been made and that nonresponse is a material contributor to that bias. However, although the bias is in the direction one might expect if noncontact was the main cause, it is not clear which component sources are responsible for the error.

[10] Descent signifies ancestry, whereas ethnicity is considered to relate more to cultural identity (Statistics New Zealand, 2007c). Hence, it cannot be said that responses to an ethnicity question (e.g., as reported in Table 22) are directly comparable to responses to a question of descent (the source of the frame variable in Table 23). Nevertheless, it is reasonable to assume that the two are related; people of Maori descent are more likely to signal that they are also Maori ethnicity. Hence, the consistent underestimation of these variables across the survey and frame data indicates a lower response propensity for those who identify themselves as Maori in one or the other form.

The second hypothesis put forward was that, even though they would not eliminate nonresponse bias, multiple contacts would reduce the net error ultimately incurred. To test this hypothesis, frame variable analysis was performed by wave of response for each of the ISSP surveys. In most instances, there was improvement in estimates over waves. Table 24 presents a summary of the analysis. Each cell reflects the percentage change in estimated nonresponse bias from wave one to the final result. A negative percentage reflects an improvement, such that estimated bias reduced from the first wave to the final result. For example, the figure '-36%' for ISSP01 with respect to 'Average Age' indicates that the estimated bias in this variable reduced by 36% (from 7% to 4%) between the results from the first contact for that survey and the final result after four contacts.

Table 24: Percentage change in estimated bias after multiple contacts

Frame Variable	Issp01	Issp02	Issp03	Issp04	Issp05	Issp06
Average Age	-36	-41	-39	-42	-52	-40
% Male	-38	-12	-38	**68**	-50	-31
% Maori Descent	-25	-47	-33	-42	-29	-18
% Employed	**90**	**105**	-59	**34**	**3**	**193**
% Student	-51	-50	-17	-51	-36	-41
% Retired	-72	-57	-43	-42	-94	-62
Avg. Age (HH)	-40	-48	-41	-37	-56	-38
Avg. # Electors (HH)	-16	-40	-10	-69	-27	-15
Avg. # Surnames (HH)	-11	-26	-3	-60	-27	-8

In all but six cases, estimated bias on the frame variables was reduced by follow-up contacts, and five of the counter-cases occurred for one variable (% Employed). Hence, hypothesis two gains moderate support; increasing response via follow up contacts does improve estimates in many cases, but does not completely remove bias. Researchers implementing only one contact can therefore expect that their results will suffer more bias due to nonresponse, at least with respect to the majority of the variables examined here, than would be the case had they attempted follow-ups.

The anomalous result, and swings in estimated bias, for the '% Employed' is related to the fact that the bias for that variable is the lowest of all of the variables (see Table 23, p.72). Specifically, after first wave returns, the valid group differed from the entire sample by an average of only 1.2%. After follow-ups, this increased to an average of 3.6%; still the smallest of the frame variables examined. Thus, even though multiple contacts increased bias, they did not do so dramatically; the large percentage changes presented in Table 24 for the '% Employed' variable are due to the small base upon which the changes occurred.

Nevertheless, the fact that a number of estimates were actually degraded by follow-up contacts reinforces the idea that, in the absence of independent data, caution must be taken when undertaking wave-analysis to estimate nonresponse bias. Where repeated contacts bring even a small increase in bias (e.g., by improving response from those who already respond at adequate rates, such as the employed) wave extrapolation will exacerbate the error, rather than alleviate it.

3.5. Noncontact as a Contributor to Net Nonresponse Bias

In order to clarify the contribution component sources make to the net nonresponse bias found in the prior section, an attempt was made to estimate error by response disposition type. First, differences on frame variables were examined. Then, for a number of variables for which no frame data existed, correlation analysis and wave extrapolations were employed in an attempt to generate indications of nonresponse bias direction and magnitude.

3.5.1. Component Bias as Measured Against Frame Variables

The existence of frame data for a number of key demographic and household variables presented an opportunity to compare differences across the response groups. For each survey, the average value for sample units in each response group was calculated for the same variables in the frame-based analyses undertaken in the previous section. Table 25 presents a summary of the comparisons. It is important to note that the figures presented are averages across the six ISSP studies. That is, they are 'averages of averages'. For example, the value '50' in the 'Valid' column for 'Average Age' indicates that, across the six studies, the valid response groups had an average age of 50. This average comprises the average ages for the valid groups in ISSP01 through ISSP06 of 49, 49, 52, 52, 48 and 49.

It is acknowledged that the presentation of 'averages of averages' can obscure substantial differences in individual results. However, the key trends apply in all but a few cases across each of the six individual studies.

Table 25: Average values for frame variables by response disposition

			Response Disposition			
Frame Variable	All	***Valid***	Inactive	GNA	Refusal	Ineligible
Proportion of Sample (%)	*100*	***52***	*32*	*9*	*4*	*3*
Average Age	47	***50***	42	43	59	60
% Male	48	***45***	52	51	41	45
% Maori Descent	14	***11***	19	17	9	7
% Employed	59	***61***	59	58	41	34
% Student	8	***6***	11	10	2	7
% Retired	11	***13***	6	8	25	36
Avg. Age (HH)	47	***49***	43	44	58	61
Avg. # Electors (HH)	3.2	***2.8***	3.0	3.8	3.0	10.0
Avg. # Surnames (HH)	2.1	***1.7***	2.0	2.9	2.1	8.4

Note, this includes samples with stratification in 2003 (Maori), 2005 (Age) and 2006 (Age/Gender). The patterns still apply despite this.

The different sources of nonresponse error do appear to exhibit different profiles. Specifically, the refusal and ineligible groups are similar to one another on many variables but differ on gender, occupation and household composition. Furthermore, the GNA and inactive groups are strikingly similar on all variables except household electors and surnames. More importantly, the refusal and ineligible groups tend to contribute opposing bias compared to the GNA and inactive groups. Indeed, for every variable except "% Employed", the net nonresponse bias across the surveys, as indicated by the difference between the all and valid columns, is in the same direction as that for the refusal and ineligible groups. Said another way, incorporating more refusers or ineligibles in the valid group would typically make estimates *worse*. Thus, the net nonresponse bias in the studies examined is attributable to nonresponse from the GNA and inactive groups.

These results lend support to hypotheses three and four; that the component sources would contribute a different profile of nonresponse bias to the studies and variables examined and that, where component sources affect the same variable, a net bias would still arise despite limited offsetting effects.

Moreover, they indicate an opportunity to reduce noncontact bias via methods targeted at noncontact. Given the findings of chapter 2, it is reasonable to assume that the contribution of noncontact to the net nonresponse bias presented in Table 25 is underestimated because not all noncontacts were reported. Indeed, in the Address Change study, the total noncontact rate was estimated to be double the amount reported in GNA figures for an unmessaged treatment after multiple contact waves (12% total vs. 6% reported GNAs, see Table 17, p. 45). Since unreported noncontacts reside in the inactive group, the GNA and inactive groups are so similar in nature, and net nonresponse bias is attributable to an underrepresentation of people from them, the proportion of net bias attributable to noncontact can be estimated at around 40%. Specifically, if reported GNAs represent half of total noncontact then, on average, noncontact accounts for 18% (9%*2) of response to the surveys covered by Table 25. This reduces the proportion attributable to inactives to 23% (32%-9%). Thus, noncontact represents 44% (18%/(18%+23%)) of the total noncontact bias they contribute.

In addition, there are two good reasons to expect that efforts targeted at noncontact nonresponse may be more effective than those aimed at further reducing inactives via common appeals to motivation or interest. First, some studies cited in section 3.2.2 found that general inducements and other design modifications can exacerbate respondent composition disparities. Second, the results in Table 25 reflect response after several follow-up contacts and for surveys following good design practice. Hence, in the absence of reliable and effective targeted inducements, it could be said that those people who remain in the inactive category are unlikely to be swayed by further appeals.

3.5.2. Nonresponse Bias in Non-Frame Variables

One of the criticisms levelled against bias estimation via frame variable analysis is that there are often only a select number of such variables available, and those are not necessarily related to items of interest in the survey.

Table 26: Relationship between frame and survey variables

Frame variable	Correlates with these survey variables:
Age	Religiosity, Views on importance of religion, Time at residence, Current employment status, Activity re finding work, Attitude toward men's/women's roles in relationships, Attitudes toward women working, Likelihood of having disability, Type of disability, Marital status, Household size, Type of housing, Children in household, Highest level of education, Level of income, Attendance at political meetings, Interest in politics, Attitudes toward marriage/cohabitation, Attitudes toward maternity leave, Feelings of stress, Mother worked while respondent was young, Views on Treaty of Waitangi, Views on strike activity, Views on publicity for extremist ideas, Number of people in daily contact with, Views on alcohol use reduction measures, Views on overweight people.
Gender	Household duties undertaken, Level of income, Likelihood of being a homemaker, Work history, Attitudes toward women working, Hours of work, Worked while children less than school age, Gender of closest friend, Work in dangerous conditions, Reason ended last job.
Maori Descent	Chances of voting for Maori party, Views of importance of NZ ancestry on citizenship, Views on offshore ownership of land, Views on land claim limits, Views on Treaty of Waitangi, Views on indigenous governance, Views on Maori language, Views on naming of NZ, Views on foreshore and seabed legislation.
Occupation: Employed	Level of income, Marital status, Feelings of stress, Work history, Level of education, Main source of economic support, Seriousness of disability, Number of people in daily contact with.
Occupation: Student	Marital status, Job prospects, Want for job, Ever had paid job, Age, Type of housing, Live with siblings/parents, Level of education, Main source of economic support.
Occupation: Retired	Marital status, Level of income, Number of children under 18 in household, Age, Type of housing, Live with siblings/parents, Length of residence in current town, Attitudes toward women working, Feelings of stress, Reason last job ended, Main source of economic support, Number of people in daily contact with.
Household Age	Most of those identified for the individual age variable, Attitude towards demands from family, Caring for a dependant.
Household Surnames	None.

In light of this, a correlation analysis was performed on survey variables from all six surveys to identify those related to items on the frame. Table 26 (above) presents the findings, reflecting variables that had correlation coefficients of magnitude +/-0.25 or higher with the various frame items. As might be expected, age was related to the widest range of survey variables. Yet, the other frame variables were also related to a number of survey items, many of which were not also highly correlated with age. This indicates that the net nonresponse bias found in the frame variables, attributable to the under-representation of noncontacts and inactives, is likely to have also occurred to varying degrees in the range of survey variables identified above.

It is also worth noting at this point that, although there are correlations between the frame variables examined, those relationships vary in strength and direction. Hence, the under-representation problem is a multivariate one. Indeed, evidence from Table 22 (p. 71) suggests that over-sampling on ancestry (ISSP03 – Maori Descent) affects estimates for ethnicity but does not make much difference to bias in age estimates. Similarly over-sampling on age (ISSP05, ISSP06) does not make much difference to estimates of ethnicity. It therefore appears that efforts aimed at mitigating nonresponse should take a multivariate approach rather than relying on improving representativeness of the achieved sample on only one dimension.

Another approach to investigating nonresponse bias in variables beyond those for which frame information exists is to analyse and extrapolate point estimates across waves of response. As noted earlier, there are many potential pitfalls with this technique due to the frailty of the 'continuum of resistance' assumptions it relies upon. Nevertheless, it does represent another lens through which to view nonresponse error in ISSP survey items.

Table 27 presents wave-extrapolated point estimates for the surveys and survey variables examined previously, and for which census figures were available. The "projected respondent" extrapolation procedure was used to generate item estimates for a 100% response rate. Filion (1976) describes the procedure as follows:

"Changes observed in an estimated parameter value as the response rate is increased using follow-ups of nonrespondents may be used to predict what the parameter value should be for a 100% response rate. Thus, a linear regression line may be fitted to data depicting an observed variable as a function of the cumulative response rate after each wave of replies. That is to say, fit

$$Y = a + bx$$

where Y = observed value of a variable per unit of response based on the respondents up to a given wave of replies

x = cumulative response rate up to a given wave

a,b = regression parameters (intercept and slope)" (p. 403)

Table 27: Wave extrapolated unweighted estimates compared to census

	Extrapolated Survey Estimate						*Census Result*[11]	
Variable	2001	2002	2003	2004	2005	2006	*2001*	*2006*
% Male	52	42	45	41	50	52	*48*	*48*
% 20-34 Years old	26	26	27	27	*31	*27	*29*	*28*
% 65+ Years old	14	10	15	16	13	13	*17*	*17*
% Maori Ethnicity	11	11	^22	14	12	13	*11*	*11*
% Marital: Single	24	22	27	24	21	25	*31*	*31*
% Bach/PG Qual	14	19	19	24	19	25	*10*	*14*
% Income <$20k	38	27	38	38	31	31	*49*	*39*
% Income >$50k	14	21	17	17	25	27	*13*	*20*
% Not Religious	24	34	27	29	32	36	*28*	*33*
% Employed Fulltime	50	51	53	50	52	52	*46*	*48*
% 1 Person HH	13	10	10	9	9	10	*23*	*23*
% 5+ Person HH	18	18	22	17	18	16	*12*	*12*

Note: The 2001-2003 surveys had four waves of contact while the 2004-2006 surveys had three.

* The 2005 ISSP was stratified by age group. The 2006 ISSP was stratified by age/gender group.

^ The 2003 ISSP oversampled from the Maori roll.

[11] Readers are directed to Appendix section A2.1, p. 174, for a background to the census figures in the table above.

Although it is not immediately evident from the table of extrapolated estimates, a comparison with the pre-extrapolation estimates presented earlier in Table 22 (p. 71) reveals the following:

- Compared to census, the extrapolated estimates generally still underestimate the proportion of 20-34 year olds, ‘Singles’, and 1 person households in the sample. Furthermore they overestimate the proportion of people with a bachelor’s degree, high incomes, and 5+ person households.
- Although the extrapolation moved some estimates closer to the census parameters[12], this occurred consistently for only two variables: ‘%20-34 year olds’ and ‘% singles’. Overall, only 27 (38%) of the 72 cells in Table 27 represent an improvement over pre-extrapolation estimates.
- In 17 cases, the extrapolated estimates led to deviations compared to census figures that were in a *different direction* to those from pre-extrapolated estimates. That is, the extrapolated estimates overshot the census figures. This occurred in every instance for the ‘%65+ year olds’ variable. In some cases, this still led to estimates closer to the census figures. However, in 9 cases it did not.
- Thirty six cells contained estimates that were in the same direction of deviation from census results, but greater than those for the pre-extrapolated estimates. The vast majority of these (30 out of 36) occurred for the survey variables in the bottom section of the table (i.e., from ‘% Singles’ down). The implication is that, for a number of survey-only variables (qualifications, low income, and large household size in particular), later waves attracted respondents that led to *more* bias in estimates. This underscores the point made earlier, that while multiple contacts often do improve estimates, this does not occur in every case and, so, care must be taken when employing wave extrapolation to estimate survey bias.

Consistent with the findings of prior research (e.g., Ellis et al., 1970), a number of cells (18 out of 72) did not exhibit linear changes in cumulative estimates across waves, contrary to ‘continuum of resistance’ assumptions. This rose to 32 out of 72 when wave estimates were examined individually, rather than cumulatively.

[12] Census figures for 2002-2005 were interpolated from the 2001 and 2006 figures.

Overall, then, it is not clear that wave extrapolation is a reliable mechanism for estimating the direction or magnitude of nonresponse bias in the studies and variables examined here, let alone a good tool for enabling the isolation of noncontact bias. The fact that the trends in underestimation and overestimation for certain variables identified earlier remained despite extrapolation for nonresponse is at least consistent with this technique's systematic exclusion of noncontact. Nevertheless, it is not possible to draw any solid conclusions about whether this a key cause of the method's performance, because there are too many potential sources of error involved.

For instance, measurement and coverage error are likely to account for some of the difference between survey estimates and census figures, which confounds any analysis of bias magnitude. Moreover, because wave extrapolation treats all forms of nonresponse together and is dependent on response rates, it is not clear what portion of the variation it its performance above can be attributed to small sample sizes for later waves, the absence of information relating to noncontacts, or residual differences between later responders and remaining passive refusers. Thus, hypothesis 5 is not supported.

4. Approaches to Reducing Noncontact Bias

4.1. Introduction

Prior chapters established that noncontact is an underreported and nontrivial contributor to postal survey nonresponse, and indicators suggest it leads to bias in a range of estimates that is not completely counteracted by other nonresponse components. Thus, efforts to reduce noncontact or adjust for its biasing effects may hold promise for improving postal survey estimates.

The existing nonresponse literature posits two general approaches to bias reduction; post-survey adjustment via techniques such as weighting or imputation, and in-field design interventions aimed at limiting it ‘up front’. Both are commonly employed, but the success of post-survey approaches ultimately rests on the amount of data gathered during the field period and the validity of assumptions about the relationship between responders and nonresponders. Hence, many methodologists advocate a ‘responsive design’ approach to fieldwork that allocates resources to in-field interventions targeted at low-response groups.

Based on this, the project’s focus moved to exploring potential in-field mechanisms targeted at noncontact nonresponders. After examining the literature for techniques that could be modified for such a purpose, four candidates were identified: finding and subsampling noncontacts, sampling movers from an independent source, substitution from within mover households and sampling based on propensity to be noncontact. Of these, the first three were found to have significant limitations in the postal mode. However, the fourth option, noncontact propensity sampling (NPS), appears to have both a compelling theoretical foundation and potential for wide practical applicability.

4.2. Existing Approaches to Nonresponse Bias Reduction

Researchers typically manage postal survey nonresponse bias via efforts to motivate passive refusers or the application of post-survey weighting procedures. These activities take place at different points in the survey process and are often used in combination, since no one technique is likely to completely eliminate bias.

A review of the techniques most commonly employed to reduce or adjust for bias is presented below. Readers will already be somewhat familiar with many of them, since they were also introduced in the bias estimation literature review undertaken in section 3.2.1. Rather than reiterate that discussion, the following section focuses specifically on the technical aspects of existing procedures that serve to inform the justification and development of noncontact targeted approaches. In general, the procedures can be categorised according to the point in the survey process at which they are employed.

4.2.1. Post-Survey Adjustment

Adjustment for unit nonresponse using response or auxiliary data can be made by weighting respondent values or, in some circumstances, imputation of nonrespondent values. With respect to the first approach, a range of weighting techniques may be applied at the point of survey analysis if there is reason to suspect nonresponse has occurred differentially across important subgroups in the sample. All involve splitting the respondents into mutually exclusive cells based on a selected set of characteristics (e.g., age, sex) expected to correlate with both response propensity and response to key survey items. Weights are then defined for each cell and these are applied to all cell members such that the weighted response rate is equalised across the cells and any associated survey estimates are adjusted along with the response distribution.

One of the simplest ways to develop weights is to take the reciprocal of the response rate within each cell, where cells are defined using one or more pre-existing frame variables. More complex methods may incorporate survey paradata (such as number of contacts prior to response) into cell definitions and judgement into whether every cell is then weighted (i.e., conditional weighting), or use multivariate modelling of response to develop classes and weights. Called response propensity weighting, the latter approach is described by Lynn (1996) as follows:

> *"weights are defined by the estimated coefficients of a multiple regression model (where survey response is the dependent variable). With this strategy, the weights are reciprocals of estimated (by the model) response rates for*

classes, where the classes are defined as all possible combinations (represented in the sample) or categories of the predictor variables. (Note that an alternative use of a regression model is simply to define the classes, to which simple inverse response rate weights can then be applied).” (p. 210).

Examples of applications of this technique can be seen in Woodburn (1991), Goksel et al (1992), Czajka et al. (1992), Fisher (1996) and Lee (2006).

In some cases, auxiliary data (e.g., from the frame) are either unavailable for both respondents and nonrespondents or are available but not expected to correlate with response or survey answers. An alternative mechanism for determining weights in these situations is to define weighting cells using external population data such as those from a census. Specifically, population figures are matched to questions on the survey and weights are calculated according to the ratio of proportions in the population to those achieved in the survey returns.

Although it is very common, there are two practical issues associated with this approach. First, differences between the survey data and population figures may in part be caused by measurement and coverage error, leading to weights that do not adequately equalise cells on nonresponse (see Lynn, 1996 for a discussion of this issue). Second, population data are not always available at the required level of detail, especially if the researcher wishes to develop cells based on the intersection of multiple variables (e.g., age group, sex, income range). Where this occurs, Raking (aka Iterative Proportional Fitting or Sample Balancing) may be used to adjust the weights of cells iteratively until the weighted sample marginal totals converge as closely as possible to the known aggregate population totals (see Battaglia, Izrael, Hoaglin, & Frankel, 2004; Little, 1993; Oh & Scheuren, 1983). However, doing so adds another level of potential error into the adjustment process.

Whereas weighting adjusts estimates for all survey variables by changing the contribution of values from individual responders, another approach to post-survey adjustment for nonresponse is to impute nonrespondent values via modelling. Imputation is commonly employed for item nonresponse, where values from a range of completed variables and across many individuals may be used to predict missing items. However, a form of imputation can be undertaken at the unit level using aggregate data from multiple field waves to predict the values of nonrespondents for each variable. As noted in the previous chapter (see the discussion of wave extrapolation in sections 3.2.1 and 3.2.2), there are many pitfalls associated with this approach, some of which relate to its assumptions about the similarity between late respondents and nonresponders. Fuller (1974) expresses this as follows:

> *"An assumption underlying these techniques is that those infrequently at home are similar to those never contacted during the survey, or those responding late to a mail survey are similar to those who do not respond at all. However, there is no a priori reason to believe that the nonrespondents in either instance would have answered the survey in the same way as those who were infrequently at home or those whose returns were mailed late."* (p. 242)

Fuller (1974) also outlines his concerns about post-survey weighting in general and advocates gathering more data in preference to relying on these techniques to resolve nonresponse bias:

> *"When the returns of respondents who are judged to be similar to the nonrespondents are multiplied to adjust for nonresponse, population estimates will be biased to the extent that the weighted returns differ from those that would have been obtained from the nonrespondents. The best procedure for avoiding such bias in a probability sample appears to be that of obtaining 100 percent returns from a random sample of the nonrespondents."* (p. 246)

Other authors have expressed additional reservations about post survey weighting. For instance, it is possible for weighting to inflate the variances of survey estimates (Kalton, 1983; Little & Rubin, 1987), although it is not a necessary consequence of these procedures (Little & Vartivarian, 2005). Furthermore, as alluded to earlier, weighting procedures commonly treat all forms of nonresponse as one, when there is often good reason to expect that differential treatment may lead to more robust adjusted estimators (Groves & Couper, 1995). Finally, there is the issue of choice of adjustment cells and variables. Ideally, there will be a relationship between the auxiliary variables and the inference variables such that an improvement in representativeness on one leads to an improvement in representativeness on the other. Furthermore, adjustment cells should contain reasonable sample sizes and be internally homogenous with respect to the adjustment variable (Little & Rubin, 1987). In practice, judgement is required to make these decisions and their validity is often not able to be empirically verified.

The dependence of post-survey weighting procedures on sometimes unverifiable assumptions and limited information mean that they should generally be considered a secondary defence against bias. As Holt and Elliot (1991, p. 334) note, *"[it] is better to collect the intended data than to rely on subsequent methods of adjustment at the survey analysis stage*."

4.2.2. In-Field Design Considerations

Given the known issues with post-survey weighting, researchers often focus on preventing missing data by design (McKnight, McKnight, Sidani, & Fiqueredo, 2007). As discussed earlier, a number of successes have been achieved in this regard. For instance, pre-paid incentives, advanced contact, callbacks, messages incorporating personalisation and persuasive techniques, and aspects of questionnaire design (e.g., length) have all been found to affect unit response in a variety of circumstances (Dillman, 2000, 2007; Groves et al., 2002; Groves, Fowler et al., 2004; Yammarino et al., 1991).

These techniques do not have to be administered from the start of the survey process or to all respondents. Indeed, it is often difficult to judge how the various response improvement options available might affect response rates, survey statistics, or costs at the outset of a survey project. Rather, they may be used as part of a responsive design methodology, in which progress toward goals are monitored throughout the process and emphasis is placed on targeted in-field interventions aimed at reducing total survey error (Groves & Heeringa, 2006; Lavrakas, 1996; S. K. Thompson & Seber, 1996).

Groves and Heeringa (2006) introduce the approach as follows:

> *"The development of computer-assisted methods for data collection has provided survey researchers with tools to capture a variety of process data ('paradata') that can be used to inform cost-quality trade-off decisions in realtime. The ability to monitor continually the streams of process data and survey data creates the opportunity to alter the design during the course of data collection to improve survey cost efficiency and to achieve more precise, less biased estimates. We label such surveys as 'responsive designs'. The paper defines responsive design and uses examples to illustrate the responsive use of paradata to guide mid-survey decisions affecting the non-response, measurement and sampling variance properties of resulting statistics." (p. 439)*

Readers are directed to Groves and Heeringa (2006) for details of the responsive design approach. However, in essence it advocates moving away from traditional 'set and forget' designs that attempt to treat nonresponse in an aggregate and pre-specified way, toward designs incorporating active monitoring via paradata (Couper, 2000), adaptive sampling and targeted interventions at different survey phases. Thus, responsive designs may include sub-sampling of nonresponders, alternative methods of follow-up contact, randomised experiments to test design elements in early phases and stratification targeted at nonresponse.

Response monitoring and bias estimation are key components of responsive design. Thus, although the approach represents a step away from the naïve focus on increasing response rates that may actually increase nonresponse error (Kessler, Little, & Groves, 1995), it remains susceptible to the known nonresponse measurement issues discussed earlier (see section 3.2.1) and requires the careful application of researcher judgement.

Furthermore, responsive design is currently harder to implement in postal surveys than other modes. For instance, the difficulty in separating out nonresponse components means monitoring is hindered and the development of targeted interventions, especially for noncontact, has been limited. Postal surveys are also often associated with research projects aimed at geographically diverse populations, constrained by modest budgets and reliant on a frame with limited individual data. As such there are fewer opportunities for procedures employed in other modes like alternate contact follow-ups or field-force collection of nonrespondent paradata.

Findings reported earlier regarding the effectiveness of an envelope ‘please return’ message (section 2.6) and procedures for estimating unreported noncontacts (section 2.7) suggest improvements can be made in monitoring nonresponse components throughout the postal survey process. However, they do not provide researchers with any interventions that might assist them to deal specifically with bias introduced by noncontact. With this in mind, the following sections detail some potential mechanisms for managing noncontact bias in the postal mode.

4.3. Potential Mechanisms for Targeting Postal Noncontact Bias

One obvious solution to the problem of noncontact is to avoid its occurrence by using a frame without misaddressing. Although it may be possible to move closer to this ideal by using a fresh snapshot of a frequently updated list, in many situations the number of out-of-date addresses in a list is beyond the researcher’s control. For instance, delays caused by bureaucracy associated with the survey process may mean the list ‘ages’ before a sample from it is fielded, or there may not be budget available to procure a fresh snapshot.

Furthermore, as suggested in the conceptual model of nonresponse sources on page 22, even a fresh snapshot of a frame will contain misaddressing because some movers will not notify the frame owner of their change of address, the frame owner’s update processes may be inefficient, or substantial frame cleaning activities are infrequent.

Given the various factors influencing frame address quality, it is likely to be more practical to attempt to compensate for noncontact than to eliminate it. To that end, a number of in-field design approaches could be considered.

4.3.1. Finding and Sub-sampling Noncontacts

Just as some studies attempt to estimate nonrespondent characteristics by following-up sub-samples of passive refusers, extra effort could be taken to find and then survey a sub-sample of noncontacts. The usefulness of this approach relies on two things: ability to find alternate contact details for the mover and a high response rate to the subsequent survey attempt. Furthermore, sufficient budget must be available to undertake the additional activity.

Unfortunately, the original choice of survey mode (postal) is often associated with tight budget constraints or a paucity of alternative contact information. For example, in one postal survey known to this author, an attempt was made to follow up GNA returns to the first wave of contact. The frame did not contain any telephone numbers or email details and budget restrictions meant physical visits to the geographically diverse GNA addresses were not feasible. In order to try and find alternative contact details for the household, the surnames and initials of other electors at the same address were taken from the frame. These were then cross-referenced with the public online telephone directory in an attempt to find an alternative address. Where a match was found, the survey was resent to the new address. Of the 47 GNAs this procedure was employed on, updated addresses could only be found for 8 (Matthews, 2006). Of those, one was returned GNA, and no reply was obtained from the others.

Although it represents only one attempt at follow-up of GNAs, this result does suggest that it is an approach fraught with difficulty in the postal context. This is not surprising, given that similar studies that have attempted to follow up postal survey nonrespondents in person or by telephone (e.g., see Gendall et al., 2005; Sosdian & Sharp, 1980) have also reported significant difficulties in finding and then gaining cooperation from nonrespondents.

4.3.2. Sub-sampling Recent Movers from an Independent Source

In some situations it may be possible to obtain a list of people who have recently moved, from which a sample can be taken to estimate GNA nonrespondent values. The feasibility of this approach will depend on the availability of such a list, the cost of sampling from it, and the degree to which those on the list are expected to represent movers in the population of interest. Hence, it may not be a viable option in many instances.

By way of example, the dominant postal provider in New Zealand offers a redirection service to customers who are moving house. In the past, the service was free for two months for residential redirections within New Zealand, with redirection for longer periods or to overseas addresses incurring a charge. However, in 2006 a $20 charge for the minimum two month redirection period was introduced. As part of the service setup, customers are able to opt-in to receive promotional material from New Zealand post partners. Those who do opt-in become part of the ‘New Movers Mailing List’, which New Zealand Post sells access to for a ‘setup plus per-record’ fee and which would therefore be available to survey researchers with the necessary budget.

Given that such a list is available in New Zealand, the question arises whether it could be used to generate a representative sample of movers. According to the New Zealand Post website, approximately 65% of the people that sign up for the redirection service each day opt-in to the list, adding around 178,000 records in a year (New Zealand Post Limited, 2008). Yet, results from the 2006 census place the number of people aged 20 or older who had been at their current residence for less than 12 months at approximately 615,000 (Statistics New Zealand, 2007k). Immigration may account for just under 60,000 of that number (Statistics New Zealand, 2007b). At best, then, the movers database is estimated to cover around 30% of those over 19 years old who move within New Zealand in a given year. This lack of coverage, in addition to the recently introduced fee for the redirection service, casts substantial doubt over the utility of the New Movers Mailing List to generate a representative sample of movers for research purposes.

An additional limitation of the list exists for those using the electoral roll to source samples (just as the surveys analysed in this thesis do). That is, all redirections requested via New Zealand post are automatically used to update the electoral roll frame. Customers are unable to opt-out of this particular update procedure because enrolment is a legal requirement and the electoral roll is not available for general commercial use. Thus, unless a significant period has passed between the frame snapshot and the survey field period, those that are returned GNA to the original mailing will either not be represented on the New Movers List because they did not opt in to it, or may be represented but with the same (incorrect) address.

4.3.3. Sub-sampling from within ‘Mover’ Households

Kish and Hess (1959) describe a novel procedure aimed at reducing noncontact nonresponse bias in face-to-face household surveys that was used to inform development of similar procedure for targeting noncontact in postal surveys. Specifically:

> *“…the plan consists in including with the survey addresses some nonresponse addresses from earlier surveys in which the sampling procedures were similar; interviews from the former nonresponse addresses become ‘replacements’ for survey addresses which result in non-responses. The plan is particularly well adapted to organizations which frequently conduct surveys with similar sampling procedures.”* (p. 17)

Two key ideas introduced by Kish and Hess’ proposal are that ‘difficulty of contact’ is survey independent and that, given enough attempts at contact, all households will eventually be interviewed (that is, each household has a propensity for contact that is greater than zero). These assumptions imply that the views or behaviours of nonrespondent households with similar contact propensities should be directly substitutable for a given survey and that contact attempts across surveys of similar methodology can be treated as contiguous for the purposes of substitution.

For example, consider two household surveys employing the same sampling procedures and providing for a maximum of five contact attempts to each household. At the end of the first survey, a number of households will remain uncontacted despite the five calls. This may lead to bias in the survey results because, as outlined in earlier sections of this thesis, the uncontacted households (i.e., those with a contact propensity less than 1 in 5) cannot be assumed to have the same views or behaviours as those that were contacted.

However, Kish and Hess suggest that the uncontacted addresses may be included in the second survey and treated as though the calls made as part of that field period are additional to those made to the households in the first field period. As such, any responses from those households can be used to estimate the respondent values for the second survey for households with contact propensities ranging from 1 in 6 through to 1 in 10.

Unfortunately, such a procedure would not function in the same way in a postal context as it would for a telephone or face-to-face survey. This is because many postal surveys are addressed to an individual, rather than a household, and the method of contact is asynchronous. As a result, noncontact in postal surveys does not occur because a person isn't home at a particular point in time, but rather because the person is no longer living at the address. Including postal noncontact sample units in later surveys would serve no purpose, because no number of follow-up contacts will lead to survey receipt and, so, no information will be obtained from those units.

One potential solution to this issue may be to move from the individual level to the household level when substituting for postal survey noncontact. Specifically, instead of resampling noncontact units in later surveys, another individual could be sampled at random from those households returning a GNA notification within the same field period. That is, noncontacts could be substituted with a member of the noncontact household.

The assumption behind such an approach is that, on average, those who live at addresses relating to a noncontactable individual have similar views and behaviours to the noncontactable individual. This might be expected to be the case when, say, a group of student renters move out of a house to be replaced by another group of student renters. However, it is also not difficult to imagine situations in which the substitutability assumption may break down for an individual household. For instance, a student boarder may move from a household leaving the host family in residence, or an owner-occupier may move from a house which then becomes occupied by a group of renters. In the absence of reliable data relating to typical household composition changes over time, it is not possible to determine whether the various combinations of changes ‘even out’ over a population such that a random sample of individuals from households about to change composition is equivalent to a random sample of individuals from households that have just undergone a change. Indeed, even if it is the case that compositional changes even out, household-level substitution of respondents makes determining individual selection probabilities difficult, thereby potentially introducing error via the sampling process.

Despite these potential assumptional and sampling difficulties, a small empirical test of this procedure was performed as part of a postal survey on attitudes to advertising undertaken by a colleague (Hoek, 2006). Addresses from which a GNA response was received after the first wave of contact were sent an invitation to ‘The Householder’. The invitation asked that the enclosed survey be given to the person aged 18 or older, mimicking the ‘next birthday’ method commonly employed in face-to-face surveys to select a pseudo-random member of a household. In addition, an insert was added to the survey asking about household composition and tenure of residents.

The results suggest a number of practical difficulties with this technique. For instance, of the 72 households returning a GNA report in response to the first wave of contact, one could not be sent a ‘Householder’ invitation because the address was a postal delivery centre. Furthermore, eight ‘Householder’ invitations were themselves returned GNA because the address no longer contained any occupants (e.g., as signalled by the postal delivery worker) or the address was not a typical household (e.g., a rest home or student hostel).

Seventeen questionnaires were returned, but they tended to have been filled in by people who had been at the household for a relatively long period (11 years on average – only five had been at the residence for less than a year) and who were older than both the respondents they replaced and the GNA group in general (54 years compared to 42 for the people being replaced, and 39 for the GNA group according to frame data). These initial results raise serious questions about the ability of the technique to generate an adequate substitute sample for noncontacts.

Future studies may be able to resolve some of the problems identified in the test. For instance, the ‘last birthday’ method could be replaced with a request to pass the survey on to the person who most recently joined the household, to attempt to better match substitute respondents with the original noncontacts. Alternatively, variations on methods such as the Kish grid (Kish, 1949) or a range of other mechanisms put forward by researchers for within-household selection in interviewer-led surveys (e.g., see Kennedy, 1993) may provide a better mix of respondents. However, as noted in an earlier chapter, it may be very difficult to administer these in a postal setting.

Notwithstanding potential improvements in sample selection, it is likely that many substantive issues will remain with this approach. In particular, it does not account for unreported GNAs and cannot generate substitutes where households become unoccupied or are non-standard. Furthermore, attempts to extend it beyond the first wave of GNA returns or to include reminders for the ‘Householder’ invitation would lengthen the survey field period and may border on harassment. Specifically, under such a scenario it would be feasible that a household would receive multiple contacts to a noncontactable individual before returning a GNA report, and then receive multiple contacts asking ‘The Householder’ to comply with a substitution request.

4.3.4. Sampling Based on Individual Propensity to be Noncontactable

Another potential noncontact-targeted procedure that draws on the ideas in Kish and Hess (1959) involves a form of substitution carried out at the sampling phase of a study. This approach is founded on the assumption that whether a postal sample unit is a noncontact or not is a function of both time and the individual’s demographic and household characteristics. Certainly, findings presented in earlier sections of this thesis (e.g., see 2.4, p. 34) lend support to the idea that noncontact has clear links to the latter.

Furthermore, as suggested by Kish and Hess, contact propensity is considered to be survey independent. Hence, although it cannot be known in advance whether a sample unit will be a noncontact to any particular survey request, it may be possible to assign a propensity score to individuals in a frame based on demographic and response disposition relationships modelled using data from prior surveys. Those sample units with similar propensity scores should be directly substitutable. That is, one could assume they would be missing (noncontactable) at random (Little & Rubin, 1987; McKnight et al., 2007) with respect to frame and survey variables of interest.

This assumption makes intuitive sense because, although some people are more likely to move than others (e.g., younger people in multi-surname households), the point in time at which they move can be considered a stochastic process unrelated to the survey request. Thus, any survey request addressed to a group of people with a particular propensity for movement (and, therefore, noncontact) will reach some who have, and some who have not, moved since their frame details were last updated. Within similar propensity groups, the key thing that separates movers from those who have not moved is time. This difference may be important in surveys related to topics such as time-in-residence. However, for most topics of interest to social science and market researchers (e.g., health, product usage, purchase behaviours, attitudes toward policy) it is likely to be ignorable.

Based on these premises, the proposed procedure would be applied at the sampling stage of a research project to modify the selection probabilities of individuals on the frame. Specifically, those with a higher propensity to be a noncontact would have a greater chance of selection, effectively adjusting the sample for an expected level and distribution of noncontact amongst the selected units. Given an unbiased predictor of noncontact propensity and an accurate estimate of noncontact rates for a period of time since the last frame update, it would theoretically be possible for the procedure described to eliminate bias due to noncontact in postal surveys.

The theoretical foundation for eliminating bias via propensity weighting is formally outlined in Rosenbaum and Rubin (1983) and further developed in Rosenbaum and Rubin (1984) and Rosenbaum (1987). Originally, the technique was created to enable unbiased estimates of treatment effects to be generated from observation studies with nonrandom assignment. However, it is applicable beyond that context. For example, Czajka et al. (1992) applied propensity weighting on early tax submission data to estimate values for final returns for the IRS. They achieved improvements in estimate accuracy on a range of variables over the existing (poststratification) method and noted that *"[t]he results demonstrate the value of propensity modeling, a general-purpose methodology that can be applied to a wide range of problems, including adjustment for unit nonresponse and frame undercoverage as well as statistical matching*" (p. 117). Others have used propensity weighting to adjust for nonresponse (e.g., Goksel et al., 1992), frame undercoverage (e.g., Duncan & Stasny, 2001), and nonprobability sampling (e.g., Lee, 2006).

Interested readers are directed to Lee and Valliant (2007) for a detailed and accessible account of the development, application, and mechanics of propensity score adjustment. They outline the following foundational assumptions of the approach:

> 1. *Strong ignorability of treatment*[13] *assignment given the value of a propensity score*
> 2. *No contamination among study units*
> 3. *Nonzero probability of treatment or nontreatment*
> 4. *Observed covariates represent unobserved covariates*
> 5. *Treatment assignment does not affect the covariates*
>
> (p. 176)

With respect to the application proposed here, there is good reason to expect that each of these assumptions is valid. The first (ignorability) has already been discussed in the first part of this section. Regarding assumption two, there is no reason to suspect that, in a random sample of individuals from a frame such as the electoral roll, the factors influencing the contact propensity of one individual will influence the contact propensity of another. Admittedly, this could occur if two individuals were from the same household, but the chances of this are very small. Turning to assumption three, it is reasonable to expect that, because they have an entry on the frame, each person will have a contact propensity greater than zero. Similarly, because there are no structural restrictions to movement, all individuals will

[13] Consistent with the original 'observational study' context in which propensity adjustment was applied (Rosenbaum & Rubin, 1983), Lee and Valiant (2007) use the term 'treatment' to refer to the variable that is the focus of the propensity adjustment or weighting (i.e., the dependant variable in the propensity model). Thus, in the context of this thesis, 'treatment' relates to the contact status of a given sample unit (i.e., whether they are a noncontact or not).

have a contact propensity less than one (Note: most frames, including the electoral roll, do not cover people in prison. Thus, whether or not these people are included is a coverage, rather than contact, issue).

The fourth assumption is ultimately untestable. However, the results presented in chapter 2, along with independent data from *Statistics New Zealand* (Statistics New Zealand, 2007j), suggest that the key correlates of noncontact are common demographic and household composition variables. Hence, to the extent that any noncontact propensity sampling (NPS) scheme is based on a propensity model that includes such variables, assumption four should be met. Finally, assignment to a treatment (i.e., whether a person is a noncontact or not at the point of survey invitation) cannot affect covariates of noncontact propensity such as age, occupation and household size (at the point of frame data collection).

There is a range of ways in which an NPS scheme might be operationalised, as discussed in a later section. However, at a conceptual level there are some clear advantages to this approach compared to the other targeted mechanisms explored earlier. Specifically, an NPS scheme would:

- Be more cost-effective than procedures that require follow-up of noncontacts. In particular, organisations that undertake multiple surveys from the same frame could expend effort building a noncontact propensity model which they could then apply across multiple surveys;
- Be founded on unambiguous and defensible assumptions;
- Allow use of a single frame for sourcing all sample units, thereby eliminating the potential for coverage error to be compounded across sub-samples;
- Maintain a probability-based sampling procedure that can be specified and documented, and potentially used in combination with other probability procedures.

As such, of the methods discussed in this section, the NPS scheme shows the greatest potential for achieving the ideal of an in-field design intervention that is "*practical, cheap, effective, and statistically efficient*" (Kish & Hess, 1959, p. 17).

It is important to reiterate that the NPS scheme would target noncontact in postal surveys. As such, although it may reduce this component of bias due to nonresponse, it would not work to reduce any bias associated with refusal or ineligibility. Furthermore, as an in-field procedure, it would aim to improve the representativeness of the raw data actually received, just as incentives and multiple contacts attempt to. Indeed, for this reason, the application of propensity adjustment proposed has the potential to reduce both bias and variance in postal survey estimates.

4.4. Predicting Noncontact: Developing a Propensity Score

A good model of noncontact propensity is a necessary precursor to a successful NPS scheme. However built, the model would need to be based on independent variables known to be available for all members of a frame and would involve a development dataset containing disposition (i.e., contact or noncontact) outcomes from prior surveys. To be successful, propensities generated by the model would have to discriminate potential sample units such that those with higher predicted propensity for noncontact were significantly more likely than chance to result in a response disposition relating to noncontact (i.e., GNA).

As discussed earlier, there is good reason to expect that noncontact can be predicted using common demographic variables. For instance, the results presented in section 2.4 suggest that both movement and GNA reports are related to age, employment status, address attributes, and household composition. Recent research from *Statistics New Zealand* corroborates these findings. Interestingly, that research also presented evidence that a high proportion of those who move do so relatively frequently, providing further support to the underlying assumptions of the proposed NPS scheme.

> *"The majority of those who moved within New Zealand during the previous two years had lived at their previous homes for less than a year (30.4 percent); 22.7 percent had lived between one and two years at their previous home"* (Statistics New Zealand, 2007j, p. 7)

A question that arises, then, is whether frames exist that would enable modelling of noncontact on the variables outlined above. At least in New Zealand, the answer is yes. For example, Social Science and Medical researchers have legal access to an electronic copy of the electoral roll for the purpose of selecting samples. The Roll contains address, occupation, and age band information for each individual. Limited household composition, ethnicity and gender information can also be calculated from those data. Furthermore, many corporate databases are likely to contain the age, gender, address and employment status of customers. Some may also contain ethnicity, household composition and marital status. Finally, even databases containing only name and address information may provide a potential basis for building a noncontact propensity model, as publicly available small area census data including average ages, household size, incomes and so on is available for free from *Statistics New Zealand*. This could be appended to augment limited existing data.

In order to examine the potential for predicting noncontact in the context of this thesis, and to provide a foundation for testing any proposed NPS scheme, an attempt was made to build models of noncontact propensity for the ISSP datasets employed in earlier sections.

4.4.1. Predicting Noncontact using Available Datasets

The following six ISSP survey datasets described in section 3.3.1 (p. 66) were analysed in an attempt to develop predictive models of noncontact.

- *"Social Networks in New Zealand" (2001)*
- *"The Roles of Men and Women in Society" (2002)*
- *"Aspects of National Identity" (2003)*
- *"New Zealanders' Attitudes to Citizenship" (2004)*
- *"New Zealanders' Attitudes to Work" (2005)*
- *"The Role of Government" (2006)*

All of the datasets contained frame information for all individuals from the New Zealand electoral roll, along with survey response disposition data. Logistic regression was chosen as the propensity modelling technique, as it is a good fit for the problem. Not only does it predict occurrence of a binary outcome (noncontact or contact) from multiple metric or categorical independent variables, it also generates propensities that can be interpreted as chances of the event occurring and these can be aggregated across subsets of individuals if necessary. It is also very commonly employed in practice as a precursor to post-survey response propensity weighting (Lee & Valliant, 2007) and direct marketing campaign target selection (i.e., identifying which customers are most likely to respond to a product offering). Hence, it is a technique for which skill exists in the marketplace and that is put to use for very similar purposes to those intended here. These are factors that would be important to the practical adoption of any targeted noncontact bias reduction procedure developed.

As part of the model building process, elementary data analysis was conducted on each dataset to identify variables likely to be good discriminators for noncontact. For instance, chi squared tests were applied to crosstabulations of response disposition code (GNA, refusal, valid, ineligible, inactive) and individual categorical frame variables to identify those that showed significant interrelationships. Finally, simple logistic regression models, incorporating weighting for stratified design where necessary, were built for each frame variable identified as a good candidate for predicting a GNA response outcome. In each case, over 2,000 individual records were used to build each model, with the number of GNA responses ranging from just over 100 cases (ISSP 2006) to just under 300 cases (ISSP 2002). Out of this process, the following variables were found to generate significant model coefficients, according to a Wald/Chi Square test ($p<=0.10$), in four or more of the six datasets examined:

- *Age*: The age of the individual. For some datasets, age is known within a five year range. For others, it is known within a one year range.
- *Dwelling Address different to Postal Address*: A flag generated when the dwelling address differs from the postal address in the database. This occurs, for example, when a person has a post office box for mail delivery or has specified a 'care of' address.
- *Dwelling 'Split'*: A flag generated when the dwelling address contains a dwelling identifier that signals a multi-dwelling situation. This occurs, for example, when an address is "12a Smith Street" or "1/34 Smith Street".

- *Employment Status*: A categorical variable indicating whether the person is employed, on some form of benefit (e.g., for sickness or widows), unemployed, retired, in study, a homemaker, or did not state an occupation. This is derived from the occupation field in the electoral roll.
- *Maori Descent Flag*: A flag reflecting whether the person signalled they are of Maori descent. As noted earlier (see section 3.4) Maori descent is not the same as Maori ethnicity, although the two concepts are related.
- *Postal Address Type*: A categorical variable indicating postal addresses that differ from standard residential addresses. The field is mostly blank, but may contain flags for Private Bags, Counter Delivery, Free Text, Overseas, P O Box and Rural Delivery addresses.
- *Electoral Roll Type*: A flag reflecting whether the individual was registered on the General or Maori roll.
- *Average Age of Electors in Household (Grouped)*: The average of individual ages for the electors at the same address as the individual, grouped into bands.
- *Number of Electors in Household (Grouped)*: The number of electors at the same address as the individual, grouped into bands.
- *Proportion of Electors in Household on the General Roll (Grouped)*: The proportion of electors at the same address as the individual who signalled they were registered on the General Roll, grouped into bands.
- *Proportion of Electors in Household that are Male (Grouped)*: The proportion of electors at the same address as the individual who were Male, grouped into bands.
- *Proportion of Electors in Household that are of Maori Descent (Grouped)*: The proportion of electors at the same address as the individual who signalled they were of Maori descent, grouped into bands.
- *Number of Different Elector Surnames in Household (Grouped)*: The proportion of different elector surnames at the same address as the individual, grouped into bands.

These variables are generally consistent with prior research regarding correlates of movement and noncontact. Furthermore, the fact that they generated significant regression coefficients across multiple studies covering different topics and slight variations in methodology suggests they meet the criteria of capturing the survey-independent causes of nonresponse expected when examining noncontact.

Based on these findings, the decision to continue investigation into a propensity oversampling procedure was made.

4.4.2. Modelling Process and Results

A number of decisions regarding approach were required to move forward with the propensity modelling phase. In particular, any models developed had to provide for a cross-study analysis of the follow-on NPS scheme. Furthermore, the modelling had to take into account the fact that different combinations of prior datasets could be used to predict noncontacts for a given survey and that differences in study methodology (e.g., number of follow-up contacts made), frame age at the time of a study, and changes in frame structure over years had the potential to affect predictions achieved. With these things in mind, rather than building one model of noncontact propensity using all prior datasets (e.g., 2001-2005) to predict for the latest available dataset (2006), it was decided to build a series of models across a range of years. Specifically, using data from the preceding two years, models were built to predict noncontact for the ISSP surveys in 2003, 2004, 2005 and 2006. For example, the model for 2003 was built using historical data from 2001 and 2002. Similarly, the model for 2004 was built using data from 2002 and 2003. So, each model was built on data independent of the set to which it was to be applied.

The aim in choosing such a combination of prior sets and predicted sets was to aggregate enough data to develop reasonably robust models of noncontact (each contributing dataset contained *at least* 4,000 total records and 380 reported GNAs) while retaining the flexibility to examine how the outcomes of the modelling effort might change over a range of time periods and studies. A ‘build using last two studies’ approach is also likely to be close to the situation faced by researchers wishing to implement an NPS scheme. Specifically, a limited range of prior studies with similar methodologies and frames are likely to be available and the most recent studies will be selected for modelling. The results would then be applied to an upcoming survey.

Exploratory techniques were employed as part of the building phase for each multivariate logistic regression model. For example, available variables were entered into the standard 'Stepwise' and 'Backward' automated selection procedures available in SAS, and the output examined for consistently included or excluded variables. Ultimately, however, the final models were selected by judgement, the aim being to build models that contained variables consistent with prior knowledge of noncontact determinants and with sensible coefficients.

Because combinations of datasets were used in model development and some of those (i.e., 2003, 2005 and 2006) were selected using age or ethnicity stratification (see section 3.3.1, p. 66, for details), all models were developed incorporating weights according to the inverse of the original selection probability for each sample unit. These weights were normalised before use. Selection probabilities were determined from the full original frame data. For example, if the sample unit was in a dataset selected using a simple random sample, then the selection probability was calculated as the reciprocal of the total number of units in the frame. Where the sample unit was part of a stratified group, the selection probability was calculated as the reciprocal of the total stratum size.

SAS's PROC SURVEYLOGISTIC procedure (see Anthony, 2002; Baisden & Hu, 2006) was used to develop the final models. This procedure gives more accurate estimates of variances and test statistics than the standard "PROC LOGISTIC" procedure, which does not adequately account for complex sample designs such as those incorporating stratification. The results of the model development process are summarised below. Detailed final model specifications are provided in Appendix section A3.1, p178. Of note is that the standard errors produced under the SURVEYLOGISTIC procedure were very similar to those from the LOGISTIC procedure. Thus, study design does not appear to have had a substantive effect on estimate uncertainty. This is unsurprising, since a test of design effects for a number of survey variable means[14] across the ISSP studies employing stratified designs (i.e., 2003, 2005 and 2006) found none greater than 1.13. Indeed, the vast majority of

[14] Tests were performed using the PROC DESCRIPT procedure available in the SUDAAN statistical software package. Variables examined included Age, Gender, Ethnicity: NZ European, Ethnicity: Maori, Religion: None, Religion: Christian, Marital Status: Single, Work Status: Employed, and Household Population.

estimated design effects were between 0.90 and 1.10, with larger positive effects (0.50 - 0.73) being limited to those variables upon which stratification had been directly applied (ethnicity, age or gender).

Table 28: Frame variables retained in the final logistic regression models

	Modelled Datasets			
Frame Variable	2001-02	2002-03	2003-04	2004-05
Individual age	✓	✓	✓	✓
Dwelling address split flag (e.g., 1/A Tea St)	✓	✓	✓	✓
Number of different surnames in household	✓	✓	✓	✓
Number of different electors in household	✓	✓	✓	✓
Proportion in household who are male	✓	✓	✓	✓
Postal address type	✓	✓	✓	✓
Dwelling address different to postal address			✓	✓
Individual enrolment type (Maori or General)			✓	✓
Proportion in household of Maori descent	✓			
Individual employment status				✓
Likelihood Ratio P Value	*<.0001*	*<.0001*	*<.0001*	*<.0001*
C statistic	*0.67*	*0.69*	*0.72*	*0.71*
Max Rescaled Rsquare	*0.09*	*0.10*	*0.13*	*0.11*

The data in Table 28 suggest that a number of variables consistently predict GNA returns, but that there is some variability in the total selected set of variables across different dataset combinations. Furthermore, although the models are significant, their fit is not particularly good; the c-statistics are at the lower end of the 0.7-0.8 range expected for "acceptable discrimination" (Hosmer & Lemeshow, 2004, p. 162). As alluded to earlier, this is likely to be due to differences in methodology, frame structure, and frame age across the studies. Furthermore, the fact that not all noncontact is reported means that the models are necessarily built on partial noncontact data, an issue discussed in more detail later.

Looking further at model performance, figures 5 and 6 present the 'cumulative gains' achieved by each model on its build dataset and intended prediction dataset, respectively. Cumulative gains charts are commonly used in assessing logistic model performance, with a larger gap between the control (or base) line and model line representing a better model. The charts below were produced by ordering the dataset on the propensities calculated by the model, splitting the set into deciles, and then graphing the cumulative number of *actual* GNAs occurring across the ordered deciles (with the first decile containing those sample units with the highest predicted propensities for noncontact).

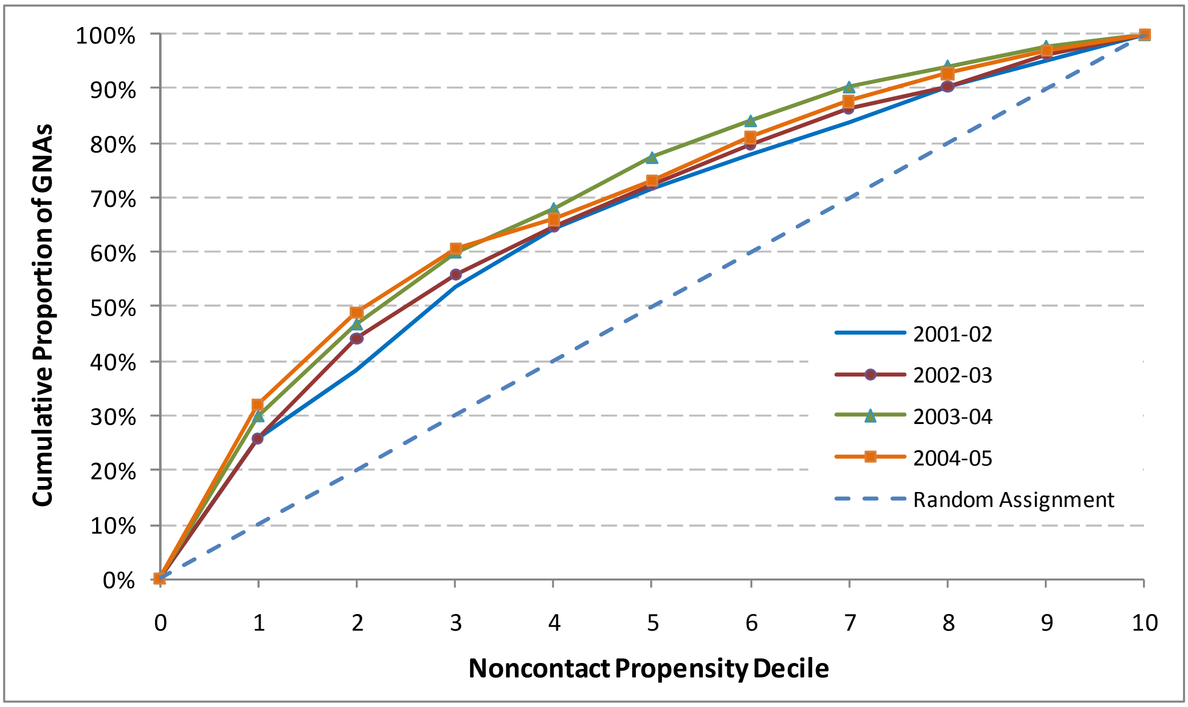

Figure 5: Gains chart for model predictions on 'build' datasets

Figure 5 suggests that all of the models performed similarly on the datasets they were built with. Around 55%-60% of GNAs occur within the first three deciles (30% of cases) ordered by noncontact propensity; a reasonable 'lift' above the baseline. This is, of course, not a very stringent test of model performance; one would hope that each model would fit the data it was trained on fairly well. However, it does indicate the 'best case' regarding how well the models might be expected to discriminate GNAs from other disposition codes if applied to a new dataset. It also gives a visual insight into just how consistent the models built across years appear to be in their predictive ability.

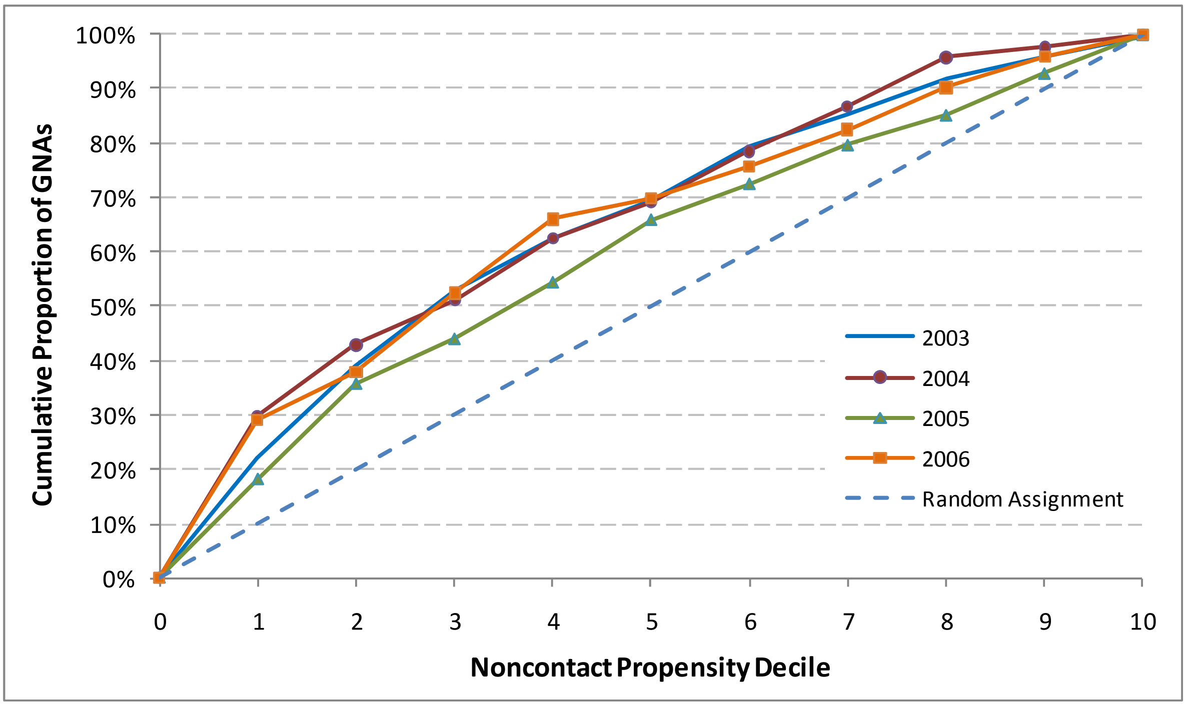

Figure 6: Gains chart for model predictions on 'test' datasets

When examining the predictions of the models on the follow-up datasets not used in their development (e.g., the model built using 2001-2002 data was applied to 2003) a small overall decline in performance can be seen, in addition to a greater variation across the models. Although this is to be expected, the results in Figure 6 do have implications for any NPS scheme relying on the models. For instance, the relatively poor result for 2005 suggests that an NPS scheme applied for that year would be expected to provide less of a reduction in noncontact bias than could have been achieved in other years.

Overall, the models do appear to have moderate predictive power when applied to a test survey setting. There is therefore potential for an NPS scheme based on them to achieve at least some reduction in noncontact bias.

4.5. A Proposed Noncontact Propensity Sampling (NPS) Procedure

A number of avenues could be taken to translate noncontact propensities into a modified sampling scheme, with the ideal being to simply count each individual's propensity score (i.e., the probability score calculable from logistic regression output) as a selection weight. Theoretically, since those with higher scores would occur in the sample in direct proportion to their likelihood of noncontact (i.e., five times as many people with a score of 0.50 would be selected than those with a score of 0.10), the resulting *contacted* sample would, on average, be equivalent to a simple random sample from a frame with perfect contact information. As discussed earlier, assuming the potential respondents within each propensity band are substitutable (i.e., the distribution of possible response for respondents and non-respondents is the same), such a scheme would eliminate noncontact bias.

Unfortunately, there are some practical limits to this approach. Specifically, as Table 29 outlines, the proportion of people with higher propensity scores is likely to be very small.

Table 29: Distribution of propensity scores in each modelled dataset

	Modelled Datasets			
Propensity Band	2001-02 (% col.)	2002-03 (% col.)	2003-04 (% col.)	2004-05 (% col.)
0.01 - 0.09	45	50	74	80
0.10 -0.19	40	36	20	15
0.20 - 0.29	11	10	4	3
0.30 - 0.39	3	3	2	1
0.40 - 0.49	1	1	0	0
0.50 - 0.99	0	0	1	0
Total	*100*	*100*	*100*	*100*

Note: The 2003-04 and 2004-05 datasets contained lower proportions of GNAs than the other datasets. Hence, a higher proportion of low noncontact propensity scores occur in those datasets.

This means that, as the requirement to oversample those with the same propensity scores increases, the number of people available for selection decreases. As Lee and Valliant (2007) note, this issue is common in propensity adjustment applications. Nevertheless, in some cases it may prohibit a 'like-for-like' or matched substitution being achieved for the most likely noncontacts. Furthermore, the propensity scores modelled relate to *reported* noncontact (i.e., GNAs). Thus, to eliminate *total* noncontact in the sample, an adjustment would need to be made to the sampling procedure. Where the adjustment requires an even greater number of high propensity units to be oversampled, as discussed later in this section, this may exacerbate the issue of matched substitution.

It is also important to note that each propensity score has an associated error, which further complicates the decision of exactly what sampling weight to assign an individual. A common approach to these practical problems in post-survey propensity weighting procedures is to apply sampling adjustments at the level of propensity score bands or strata, rather than the individual (Lee & Valliant, 2007; Little & Rubin, 1987). In essence, this represents a relaxation of the substitutability assumption such that sample units *within a stratum* are considered substitutable with respect to key survey variables (i.e., that they are noncontactable at random).

Strata can be determined by ranges of propensity scores (e.g., see Table 29, above) or by splitting the frame into equally sized groups ordered on propensity scores (e.g., the deciles in Figure 6). The most reasonable approach is ultimately a matter of judgement in each application situation, with an implicit trade-off being made between practically useful strata sizes and the variability of propensity scores within each chosen band (and the associated potential for noncontact bias reduction).

Once a stratification schema is chosen, sampling adjustment weights may be calculated using group averages and applied at the group level. Specifically, the first step would involve calculating the average noncontact propensity score to give an expected level of noncontact for each stratum. As alluded to earlier, since this figure will depend on the data used to develop the propensity model, the need may arise to adjust it to account for any anticipated differences between the modelled situation and the application situation. In particular, if there is good reason to expect that rates of total noncontact will be higher in the application situation, the expected level of noncontact for the stratum may need to be increased. There are some clear situations where this would be the case.

For example, if the application sample is to be taken from an older frame, more movement will have occurred and the expected noncontact rate will therefore be higher. Similarly, if the propensity models were built on *reported* noncontact data, the expected noncontact rate in the application situation may have to be increased to account for levels of unreported noncontact. Indeed, given the results of the noncontact reporting study presented in chapter 2, the latter issue is likely to occur in many potential NPS implementations. Specifically, most propensity models developed on prior survey data will systematically underestimate the total level of noncontact to be expected in a given propensity stratum. Moreover, the underestimation is likely to be worse in those strata with higher levels of modelled noncontact propensity. This represents a violation of a foundational assumption of the NPS scheme; that an unbiased predictor of noncontact propensity is employed.

In practice, it may be possible to resolve the underreporting bias in the propensity scores using empirical knowledge of reporting behaviour. For example, a propensity-level adjustment could be made via a model of the relationship between reported noncontact propensity and reporting rates. However, it is questionable whether many researchers will have the quantity of data necessary to do this in a robust way. Alternatively, then, adjustment may be made using estimates of reporting rates at the propensity stratum level obtained via techniques such as those developed in chapter 2 (see section 2.7, p. 47). For bands with a high level of predicted noncontact, the *Cross-Group Comparison* estimation procedure is likely to break down (as discussed in section 2.7.1, p. 51). Hence, although it is not ideal, the *Iceberg* procedure is recommended.

The result of any adjustments to the average propensity score for a given stratum will be an expected level of *total* noncontact for the group. The final step in translating this into a sampling adjustment rate is to calculate the rate of oversampling required to achieve the same number of contacted sample units as the original group size:

Equation 8: The oversampling rate for an NPS scheme band[15]

$$\text{Oversampling Rate} = \frac{\text{Expected Total Noncontact Rate}}{\text{1-Expected Total Noncontact Rate}}$$

[15] This equation is often referred to as an odds ratio in a wide range of other application situations.

This formula was derived from the general formula for the sum of an infinite geometric progression and in this particular application situation accounts for the fact that each additional unit sampled will itself have an associated chance of being a noncontact. Hence, to achieve a contacted sample the same size as the original group, each 'oversample' set must itself be oversampled according to the expected noncontact rate.

Given the required oversampling rate for each propensity stratum in the frame, the researcher can either select additional sample units relative to the originally intended stratum sample size, leading to an increased total sample size, or select the same number of people overall by readjusting the proportion of the total sample selected from each stratum. The latter approach would be expected to achieve a reduction in noncontact bias without increasing survey costs (except, possibly, follow-up contact costs). However, it would be expected to lead to fewer total achieved contacts and, so, may increase overall variance.

In summary, the noncontact propensity sampling (NPS) scheme outlined here would have four general steps:

1. Development of a propensity score using available prior data relating to the intended application frame.
2. Scoring of the sampling frame and assignment of sample units to propensity strata.
3. Estimation of total expected noncontact rate in each stratum. In many cases this is likely to involve adjustment of the propensity-based noncontact rate to account for underreporting of noncontact.
4. Calculation of the oversampling rate required in each stratum to achieve a contacted sample of the same size (or sample proportion) as the originally intended size.
5. Sample selection at random within propensity strata according to the sample size (or proportions) determined in step 4.

A spreadsheet containing a walk-through with example calculations for the proposed NPS scheme has been included on the thesis supplementary CD. Interested readers are directed to Appendix section A1.4, p. 169, for more information.

5. Simulated Performance of the NPS Scheme

5.1. Introduction

Chapter 4 noted that, although there is a clear theoretical basis for eliminating noncontact bias via propensity adjusted sampling, practical implementation considerations are likely to limit the amount of bias reduction achievable. Thus, an empirical test of the NPS scheme was undertaken to assess its likely practical effect in a general population survey setting (i.e., the context common to the studies in this thesis). In addition to examining effects on response distribution and survey estimates, the test aimed to see whether the NPS procedure improved the estimates produced by common post-survey weighting procedures.

One potential vehicle for such an investigation would have been an application of the NPS scheme in an ISSP survey. For example, twin samples could be taken; one under an SRS and one (or more) under a version of the NPS scheme, with the same survey administered to each sample. However, a one shot study provides no opportunity to examine performance of the scheme over a range of instances, including different base propensity models, survey topics or frame snapshots. Finally, because it would involve only one survey, it would be impossible to examine the effect of the NPS scheme on the variance of some metrics such as the response disposition breakdown or survey estimates after wave extrapolation. Traditional parametric approaches to estimating variances cannot be readily applied to these measures because it is not clear that fluctuations in their values over multiple samples are normally distributed.

Given the aims of the empirical test and the limitations of a one-shot survey approach, a simulation study was chosen as the most appropriate vehicle for investigation. Specifically, it was decided to resample existing survey data according to a bootstrap procedure (Efron & Tibshirani, 1993), as discussed below.

5.2. Simulating NPS Procedure Performance

As the requisite technology becomes cheaper, faster and more widely available, computer intensive statistical methods are increasingly employed to address complex estimation problems in a nonparametric way. In particular, resampling techniques such as cross validation, the jackknife and the bootstrap have become indispensible tools for such applications (e.g., see Efron, 1982). Although each involves resampling from an existing set of data, they vary in their approach to this and therefore have different strengths, weaknesses and computational requirements. In general, the bootstrap is considered to be the more robust procedure, although it typically requires more processing than methods such as the jackknife (Efron, 1979, 1982).

Diaconis and Efron (1983) give a short and very accessible account of the development, application, and theoretical foundation of the bootstrap, but readers interested in a detailed discussion of the procedure and its background are directed to Efron and Tibshirani (1993). Originally developed by Efron (1979), the bootstrap is a procedure for nonparametric estimation for a range of applied statistical problems by resampling with replacement from existing sample data. Essentially, the bootstrap treats the known empirical distribution for a variable (established from a sample) as a replacement for the unknown population distribution and generates empirical sampling distributions from it by resampling.

Chernick (2008) formally describes the procedure as follows:

> *"Given a sample of n independent identically distributed random vectors* X_1, $X_2, \ldots, X_n$ *and a real-valued estimator* $(X_1, X_2, \ldots, X_n)$ *(denoted by* $\hat{\theta}$*) of the parameter, a procedure to assess the accuracy of* $\hat{\theta}$ *is defined in terms of the empirical distribution function* F_n*. This empirical distribution function assigns probability mass* $1/n$ *to each observed value of the random vectors* X_i *for* $i=1,2,\ldots,n$*.*
>
> *The empirical distribution function is the maximum likelihood estimator of the distribution for the observations when no parametric assumptions are made. The bootstrap distribution for* $\hat{\theta} - \theta$ *is the distribution obtained by generating*

$\hat{\theta}'s$ *by sampling independently with replacement from the empirical distribution* F_n*. The bootstrap estimate of the standard error of* $\hat{\theta}$ *is then the standard deviation of the bootstrap distribution for* $\hat{\theta} - \theta$*.*

It should be noted here that almost any parameter of the bootstrap distribution can be used as a "bootstrap" estimate of the corresponding population parameter. We could consider the skewness, the kurtosis, the median, or the 95th percentile of the bootstrap distribution for $\hat{\theta}$*.*

Practical application of the technique usually requires generation of bootstrap samples or resamples (i.e., samples obtained by independently sampling with replacement from the empirical distribution). From the bootstrap sampling, a Monte Carlo approximation of the bootstrap estimate is obtained."
(Chernick, 2008, p. 9)

Typically, 1,000 or more resamples (replicates) are taken in practical applications. The statistic of interest is then calculated for each of these and used to build up an empirical sampling distribution that estimates the variance of that statistic. As Diaconis and Efron (1983) note, "*the distribution of [a statistic] for the bootstrap samples can be treated as if it were a distribution constructed from real samples; it gives an estimate of the statistical accuracy of the value of [a statistic] that was calculated for the original sample*." (p. 100). Furthermore, *"[t]he bootstrap has been tried on a large number of problems [related to variance estimation]... for which the correct answer is known. The estimate it gives is a good one for such problems, and it can be mathematically proved to work for similar problems*." (p. 108). A number of texts, including Chernick (2008) and Davison and Hinkley (1997) have examined the multitude of problems to which resampling methods can be applied.

As with all statistical methods, the bootstrap is not without its limitations. For example, in naive form it may give erroneous estimates for very small samples, extreme values, and survey samples to which a finite population correction factor would normally be applied (Chernick, 2008). None of the situations identified by Chernick apply to the application intended in this study; the survey datasets are from large samples (at least 2,000 units) drawn from a very large population (close to 3 million individuals), and the statistics under examination are common in bootstrap studies (means, variances and proportions).

Bootstrap Resampling Under a Complex Sampling Scheme

The application proposed here does deviate from typical applications in one respect; it involves resampling under a complex sample scheme. Mooney and Duval (1993) caution that much of the theoretical development of the bootstrap has occurred using the assumption of simple random sampling, to reduce mathematical complexity. Hence, it is possible that some of the positive attributes of bootstrap estimates established in prior studies under SRS assumptions fail to translate to other sampling schemes.

As outlined in a later section, this study involves stratified simple random resampling from pre-existing datasets. Rao and Wu (1988) examined the effect of this, and a number of other complex sampling schemes, on the reliability of bootstrap estimates for nonlinear statistics. They suggest the use of a rescaling procedure, dependent on the specific type of complex sample, to ensure the resulting variance estimator "*reduces to the standard unbiased variance estimator in the linear case*" (p. 231). However, in this study, a rescaling procedure is not required because the resampling relates to strata with large population sizes:

> *"For a stratified simple random sample without replacement (SRSWOR) design Rao and Wu suggest a rescaling procedure which matches the analytic formula for linear statistics. If the stratum population sizes are large, and if the bootstrap sample is a simple random sample with replacement (SRSWR) from each stratum of the same size as the stratum sample, then this procedure reduces to the naive bootstrap."* (Gray, Haslett, & Kuzmicich, 2004, p. 709).

Thus, the specific procedure employed here is naive bootstrap stratified resampling.

5.2.1. NPS Procedural Variations Simulated

In addition to testing how an NPS scheme would affect survey estimates and variances across a range of survey variables, the simulation study aimed to provide some insight into the effect of changes to NPS inputs on results. Furthermore, as alluded to earlier, it is unlikely that an NPS would be used in isolation to reduce postal survey bias. Rather, it would be employed in addition to other in-field procedures (e.g., follow-ups and incentives) and post-survey weighting techniques. Thus, another aim of the simulation was to test how the NPS might interact with the latter of these procedures to influence final survey estimates.

These aims required that multiple simulations be run over multiple survey datasets and that post-survey adjustment procedures be simulated in addition to the proposed modified sampling scheme. Details of the simulation methodology are outlined below.

Simulation over Varied Datasets

A separate simulation was performed on three of the four datasets for which propensity models had been developed (i.e., ISSP 2003, 2004, 2005)[16] in the preceding study. This enabled a comparison of performance across different time periods, propensity input datasets, in-field procedures employed and survey topics.

Section 3.3.1 (p. 66) details the differences between each study. However, by way of example, the survey in 2003 differed from the survey in 2004 in time, survey topic ("*Aspects of National Identity*" vs. "*Attitudes to Citizenship*") and number of follow-ups (three vs. two).

It is important to note that the variations that occurred across the surveys were determined by historical survey decisions and were not systematic. Hence, there are limits to what can be interpreted from any differences in NPS performance across the

[16] The dataset for 2006 was excluded because its original sample design (stratification by age band and sex) would have led to 60 resampling strata for the NPS simulation (3 age groups * 2 gender groups * 10 propensity deciles). Since 2,250 people were originally sampled, an average of 38 would fall into each resampling strata. Given an overall response rate of around 50%, this leaves only 19 units to be resampled per strata on average (the response rate is lower in high propensity bands). This may have led to unstable bootstrap estimates for survey only variables.

simulation studies. However, the comparisons did serve to answer the question of whether an NPS scheme at least exerts influence on response distributions and survey estimates in a consistent way (e.g., direction) across these varied situations.

Form of the NPS Procedure Used for Simulation

As outlined in chapter 4, two key decisions are required when translating propensity scores into an NPS survey adjustment mechanism. The first is which propensity strata will be used as the basis for adjustment. The second is whether any adjustment should be made to the expected noncontact rate in each stratum to account for known frame age differences or underreporting.

For the purposes of the simulation, a decile based stratification system was employed. In part, this was kept constant in an attempt to avoid the simulation analysis becoming too complex. The choice was also practically motivated; as Table 29 (p. 108) demonstrated, the distribution of propensity scores obtained from the logistic regression models was such that very few members of each sample would have fallen into the high propensity strata under a pure score-range based stratification. This would not have been conducive to simulation, which relies on having a reasonable pool of cases within each band to resample from. Decile stratification is also commonly employed in practical applications of propensity adjustment (e.g., see Lee & Valliant, 2007).

Thus, only the second of the key implementation decisions was varied. Here, two levels of underreporting adjustment were employed. One involved no adjustment to the expected noncontact rate (*no adjustment* treatment). The other (*stepped adjustment* treatment) assumed that the total noncontact rate would be 65% higher for the decile with the highest propensity scores and 20% higher for the decile with the lowest propensity scores, with the adjustment increasing in 5% increments for those deciles in between. This is likely to be closer to a real-world application of the NPS procedure. Hence, the comparison of NPS performance across the two adjustment scenarios was expected to show that stepped adjustment provides the greatest reduction in noncontact bias.

The step adjustment rates were established using the *Iceberg* method of estimation developed in section 2.7.1 (p. 51). Specifically, each of the datasets used in chapter 4 to build propensity models for the years 2003 to 2005 was split into propensity deciles according to the model scores relating to them. Then, noncontact adjustment rates for each decile were calculated using the ratio of GNA returns to 'responded' returns. Table 30 presents the ratio of GNA to 'responded' returns for each dataset, together with the average adjustment rate required to transform the reported GNAs in each decile into a total estimated number of noncontacts. The 'model' column represents a linear approximation to the averages (R^2=0.89) that smooths out the rates to create an evenly stepped set of adjustments. These rates were used in the simulation procedure.

Table 30: Estimated adjustment rates for noncontact underreporting

Propensity	**GNA Ratio (%)**			**Adjustment Rate (%)**	
Decile	2001-2	2002-3	2003-4	Average	Model
0 (highest)	55	52	41	64	65
1	26	40	25	72	60
2	27	22	19	54	55
3	20	17	12	59	50
4	14	13	13	52	45
5	12	14	8	52	40
6	11	12	8	47	35
7	11	7	5	40	30
8	8	10	5	29	25
9 (lowest)	8	6	2	29	20

Although there are other ways the adjustment rates might be established (e.g., direct application of the averages or judgemental selection according to an overall estimated underreporting rate), the linear step approach was chosen because it is simple, consistently follows the expected direction and trend of underreporting established in chapter 2, and is strongly tied to the decile-based averages.

It is worth noting that, as expected under the *Iceberg* procedure, the estimates established here are likely to be quite conservative. Overall, they lead to a predicted noncontact level across the whole dataset that is approximately 45% higher than indicated by reported GNAs. In comparison, the three-wave no-envelope message treatment presented in chapter 2, which is comparable to the studies in the datasets above, had a total overall noncontact rate estimated by the *Cross-Group Comparison* procedure to be 100% higher than reported (12% vs. 5.6%; see Table 20, p. 52, and Table 18, p. 46). Employing stepped adjustments that lead to a 100% underreporting adjustment would most likely have increased the bias reduction achieved by the NPS scheme. However, because the *Cross-Group Comparison* procedure breaks down at the stratum level, there was no way to empirically select or justify adjustment rates at that level. Thus, the more conservative approach was taken for the simulation.

Details of the Resampling Method

The same simulation steps were followed for all three of the ISSP datasets mentioned earlier, to enable comparisons across the simulation scenarios. In each case, three simulations were run; one simple random sample (SRS) based, two NPS based. All involved 1,000 samples (or replicates) of 1,500 cases and produced summary measures for each replicate that could then be aggregated to assess response distribution and sample statistic changes under different base sampling scenarios. The number of replicates and cases taken in the resampling procedure were considered large enough to achieve robust estimates for the statistics under examination within a reasonable computation time (the complete set of simulations took approximately one working day to run).

Interested readers will find the SAS code used to perform the simulations on the thesis supplementary CD. Appendix section A1.5, p. 170, provides further information regarding the files. The code is commented to aid readability, but at a very high level it performs the following simulation steps.

For each of the ISSP datasets, and for each simulation:

1. Establish the resampling strata and weights to be used in the resampling step. For the NPS simulation, this is achieved *a)* by scoring the frame associated with the ISSP dataset using the noncontact propensity model developed for it, *b)*

splitting the frame into deciles based on the propensity scores (and original sampling strata, where applicable), and then *c)* applying the NPS procedure to determine the proportion of the total sample to be sourced from each resampling stratum. For the SRS simulation either one resampling stratum is established (i.e., where the ISSP dataset was originally selected via an SRS) or resampling strata relating to original stratification variables are created, with the proportion to be sourced from each stratum set according to the proportion that would have been selected had an SRS been taken.

2. Apply the resampling strata assignment and weights established using the frame data to each case in the ISSP dataset. This is achieved via an ID match between cases on the frame and cases in the ISSP dataset.

3. Take 1,000 replicates of 1,500 cases by sampling at random with replacement from the ISSP dataset according to the resampling rate established for each resampling stratum. Random resampling with replacement from each stratum is achieved using the SAS SURVEYSELECT procedure.

4. For each replicate, calculate summary statistics (means, proportions, counts) for key frame, response disposition and wave, and survey variables. Also apply and calculate summary statistics for age/sex post-survey weighting. Add these to a table of summary results from the simulation run, along with an identifier for the replicate, for further summarisation as required.

5.2.2. Selection of Variables for Examination

Two criteria determined the selection of variables for examination. First, because the aim of the simulation study was to enable comparative analysis, only those variables available across all three survey datasets were included in the examination set. Second, any survey variables had to relate to items for which independent population data were available so that differences in estimates for them across the simulation scenarios could be benchmarked to indicate levels of nonresponse bias. The following variables met these criteria:

Variables from the frame or otherwise calculable for all cases

- Response disposition code
- Wave of response
- Age
- Maori descent (flag)
- Gender: male
- Roll type: General (flag)
- Occupation status
- Number of electors in household
- Average age of electors in household
- Number of different elector surnames in household
- % of male electors in household
- % of Maori electors in household
- % of electors on the general roll

Variables common to most surveys, for which population data existed

- Gender (flag)
- Age (and bands)
- Religion: None (flag)
- Religion: Christian (flag)
- Employment status: Part time (flag)
- Employment status: Full time (flag)
- Highest qualification achieved
- Individual income (band)
- Marital status: married (flag)
- Marital status: single (flag)
- Marital status: widowed (flag)
- Ethnicity: NZ European (flag)
- Ethnicity: NZ Maori (flag)
- Voted for National in 2002 (flag)
- Voted for Labour in 2002 (flag)
- Number of people in household (and bands)

For all except the voting variables, comparable population parameters from the NZ census (2001 and 2006) were available. For the voting variables, final counts from the 2002 election were available from the NZ electoral commission website (http://www.elections.org.nz).

5.2.3. Post-survey Bias Reduction Techniques Compared

As part of the simulation, results from the NPS scheme were compared with three post-survey estimate adjustment methods. Specifically, frame-based age/sex weighting, census-based age/sex weighting and wave extrapolation were performed for each survey variable, for each replicate returned. The frame-based weights were calculated using cell breakdowns from full frame information relating to each simulation dataset.

The census age/sex weights were based on the population cell breakdowns presented in Table 31. These were determined using publicly available census data from 2001 and 2006. Since the cell breakdowns did not change substantially across those years, and all of the survey datasets simulated were fielded within the period, a simple average of the 2001 and 2006 census data was taken to establish the proportion of the population in each cell.

Table 31: Proportion of the population in each census age / sex cell

Age Group	Male	Female	Total
18-30	0.11	0.12	0.23
31-40	0.10	0.11	0.21
41-50	0.10	0.10	0.20
51-60	0.08	0.08	0.15
61-70	0.05	0.05	0.10
71-80	0.03	0.04	0.07
81+	0.01	0.02	0.04
Total	*0.48*	*0.52*	*1.00*

Sources: Statistics New Zealand (2007d) for 2001 data. Statistics New Zealand (2007i) for 2006 data.

Finally, the wave-based procedure was performed via linear extrapolation of cumulative averages for each wave of response, for each survey variable (i.e., the same technique discussed and used in section 3.5.2, p. 77).

Of the three common post-weighting methods examined, the frame and wave-based procedures attempt to adjust for all sources of nonresponse. Conversely, the census procedure attempts to weight for all sources of error. Thus, the aim in combining these with the NPS scheme is to see whether the survey returns achieved under an NPS provide a better foundation than an SRS for untargeted post-survey weighting.

5.2.4. Hypothesised Effects of the NPS Scheme

The study sought to compare response distributions and survey estimates derived from the SRS and NPS simulations over different base datasets and implementation decisions. It also aimed to examine the effect of the NPS procedure on estimates adjusted by common post-survey weighting procedures. Given the findings on bias and nonresponse components presented in chapters 2 and 3, and the modelling undertaken in chapter 4, it was hypothesised that:

1. The NPS scheme would reduce the total proportion of valid responses and increase the proportion of GNAs and inactives across all simulation scenarios. This is because, as chapter 2 established, those more likely to be GNAs are also more likely to be inactives, and less likely to return a valid response. Nevertheless, the distribution of valid responses should be altered such that a higher proportion are from the high noncontact propensity deciles.

2. Survey estimates returned under the NPS scheme would be consistently closer to full frame values or census parameters compared to the SRS scheme. Furthermore, the variance of estimates would not be substantially increased by the NPS scheme. This is because, although the scheme would likely result in fewer valid responses from a given sample size, the fact that it is a form of stratification means its design effect should be less than 1.

3. The NPS scheme would make the post-survey adjustment techniques examined more stable (smaller, less variable weights) and improve the results of weighting such that final estimates would be consistently closer to census parameters than weighted SRS estimates.

4. The NPS results for the 2003 and 2004 simulations would outperform those from the 2005 simulation because of the relatively poor discriminatory power of the underlying noncontact propensity model for 2005.

5.3. The Effect on Response Distributions

Though only summary results are presented and discussed here (and in later sections of this chapter), the full data tables that contributed to them are supplied on the thesis supplementary CD in spreadsheet format. Appendix section A1.6, p. 172, provides further information for accessing those files.

As anticipated, the NPS scheme increased the proportion of GNAs and inactives, and decreased the proportion of valid responses. Table 32 shows that, on average across the three simulated surveys, valids were reduced from 53% to 51% under the stepped adjustment NPS scenario. Thus, employing an NPS scheme will result in slightly less raw survey data to work with (assuming a constant sample size).

Table 32: Effect of the NPS scheme on survey response†

Response Disposition	Base: SRS (% col.)	NPS Scenario: No Adjust. (% col.)	NPS Scenario: Stepped Adjust. (% col.)
Valid	52.9	52.0	51.0
Inactive	33.1	33.6	34.1
Gone, No Address	8.0	8.5	9.1
Active Refusal	3.7	3.6	3.6
*Ineligible**	*2.2*	2.2	2.2

† Figures represent averages over the three simulated surveys for 2003, 2004 and 2005

* Trends were not consistent across all three survey scenarios for the ineligible disposition code

Two points are worth noting with respect to the altered response disposition distribution. First, the NPS does not lead to a drastic reduction in valids. Furthermore researchers concerned about the scheme's impact on cooperation rates[17] are reminded that these can be adjusted according to an estimate of total noncontact derived using the procedures developed in chapter 2. For instance, given the data in Table 32, the average unadjusted cooperation rates for the three sampling scenarios are 58.9%, 58.1% and 57.5%, respectively. However, adjusting for unreported noncontact at a rate of 100% (12% estimated vs. 5.6% reported; see Table 20, p. 52, and Table 18, p. 46) results in cooperation rates of 64.6%, 64.3% and 64.0%. Thus, the NPS scheme does not have a substantial impact on this key survey response performance indicator.

Second, an examination of the valid groups under each scenario suggests that a higher proportion of the group come from high noncontact propensity deciles under the NPS schemes (see Table 33). This indicates that the scheme has its hypothesised effect and is likely to reduce survey bias.

[17] Calculated as Valids / (Total sample size – (Ineligibles + Noncontacts))

Table 33: Proportion of the valid group in each propensity decile[†]

Noncontact Propensity Decile	NPS Scenario		
	Base: SRS (% col.)	No Adjust. (% col.)	Stepped Adjust. (% col.)
0 (highest)	6	8	10
1	7	8	9
2	8	8	9
3	11	10	10
4	11	10	10
5	11	11	10
6	11	10	10
7	11	11	10
8	12	12	11
9 (lowest)	13	12	11
Total	*100*	*100*	*100*

[†] Figures represent averages over the three simulated surveys for 2003, 2004 and 2005

5.4. The Effect on Survey Estimates

In order to assess whether the NPS scheme was able to reduce survey bias due to noncontact, its effect on unweighted survey estimates was tested in two ways. The first involved a comparison on frame data. The second compared survey estimates to data from census or election returns.

Table 34 presents the results of the frame comparison, averaged across the three simulated surveys. As with the response disposition data above, average results are presented because the trends were consistent across all of the simulations. Specifically, in all but a few cases the NPS scheme led to a valid group that was more representative of the total frame than an SRS. This was achieved irrespective of whether the variables were part of the propensity models used in the NPS scheme (e.g., occupation and individual gender were not included in any of the propensity models underlying the NPS scenarios tested).

Table 34: Effect of the NPS scheme on estimates for frame variables†

Frame Variable	*Entire Frame*	**SRS** (Valids)	**NPS Scenario (Valids)** No Adj.	Stepped Adj.
Age (Mean)	*48.4*	50.9	50.5	**50.0**
Maori Descent (%)	*13.5*	11.0	11.4	**12.0**
Gender: Male (%)	*48.0*	45.4	45.4	**45.6**
Roll: General (%)	*92.6*	94.3	94.0	**93.6**
Occupation: Benefit (%)	*1.8*	1.1	1.1	1.1
Occupation: Employed (%)	*58.8*	61.2	61.1	**61.1**
Occupation: Homemaker (%)	*13.4*	14.6	14.5	**14.2**
Occupation: Not Stated (%)	*4.7*	3.1	3.2	**3.4**
Occupation: Retired (%)	*11.1*	12.4	12.1	**11.7**
Occupation: Student (%)	*7.9*	6.0	6.3	**6.7**
Occupation: Unemployed (%)	*2.3*	1.6	1.7	**1.8**
Household: Electors (Mean)	*3.0*	2.8	2.8	2.8
Household: Avg. Age (Mean)	*48.3*	50.2	49.8	**49.4**
Household: Surnames (Mean)	*2.0*	1.7	1.7	**1.8**
Household: Males (%)	*47.3*	46.3	46.4	**46.5**
Household: General Roll (%)	*92.6*	94.4	94.2	**93.8**
Household: Maori Descent (%)	*13.4*	10.8	11.2	**11.7**

† Figures represent averages over the three simulated surveys for 2003, 2004 and 2005

At least on these frame variables, then, it appears the NPS scheme consistently reduces noncontact nonresponse bias. Overall, across the variables and survey scenarios examined, it led to an average 28% reduction in absolute error between the returned valid group estimates and the known frame parameters.

An analysis of the standard deviations of the estimates from each simulation scenario, for each frame variable, suggests that the NPS procedure generally increases the variability of results under a 'constant sample size'[18] application. Although it does not do so for all variables, on average the stepped adjustment NPS procedure increased estimate variability by 4% compared to that for the comparative

[18] That is, where the NPS sample size is the same as would have been taken under a SRS scheme, rather than increasing the sample size to accommodate the requirements of oversampling likely noncontacts. See page 111 for a discussion of these possible approaches.

SRS scheme (e.g., an average standard deviation of 10 for the SRS scheme would increase to 10.4 under the NPS scheme). Overall, then, the increase is not large. That said, in the 2003 simulation two variables did increase in variation by 20%, while in the other simulations the maximum increase in variability for a variable was 11%. Thus, the scheme may have a substantive influence on variability in a small number of cases.

Turning to survey-only variables, Table 35 presents the results of an analysis to determine which scheme gave estimates closest to known census or election parameters.

The results are not as clear as for the frame variables. However, the NPS scheme did generally produce better point estimates across the variables and scenarios examined. Where the NPS1 (no adjustment) procedure outperformed the NPS2 (stepped adjustment) procedure, the estimates were typically very close between the two, so either would have produced a better estimate. Indeed, in many cases the estimates from all of the sampling scenarios were close; on average they differed by under a percentage point.

Furthermore, where the NPS procedure improved estimates, it led to an average 17% reduction in absolute error between the survey estimates and the census parameters across the simulation scenarios. Thus, although the NPS scheme was generally effective in reducing bias in many survey estimates, it cannot be said to have reduced it substantially in this analysis. In part, this result may be due to other biases inevitably present in a census-based comparison, such as measurement and coverage. Moreover, future development to improve the propensity modelling and underreporting adjustment processes may see greater levels of bias reduction achieved.

Table 35: 'Best scheme' results for survey estimates compared to census

	Scheme Resulting in Best Estimate		
Survey Variable	ISSP03	ISSP04	ISSP05
Gender (Male)	SRS	NPS2	NPS2
Age 20-34	NPS2	NPS2	NPS2
Age 35-49	SRS	SRS	NPS2
Age 50-64	NPS2	NPS2	NPS2
Age 65+	NPS2	NPS2	SRS
Not Religious	NPS2	NPS2	SRS
Christian	NPS2	NPS1	SRS
Employed Full Time	SRS	NPS2	SRS
Employed Part Time	SRS	SRS	NPS2
One Person Household	NPS2	NPS2	SRS
Two Person Household	NPS2	NPS2	NPS2
Three Person Household	NPS2	NPS1	SRS
Four Person Household	SRS	SRS	SRS
Five+ Person Household	SRS	NPS1	NPS1
Qualification Bachelor Degree+	NPS1	NPS1	SRS
Own Income <$20k	NPS2	NPS1	SRS
Own Income >$50k	NPS2	NPS1	SRS
Marital Status: Married	NPS2	NPS2	NPS2
Marital Status: Single	NPS2	NPS2	NPS2
Marital Status: Widowed	NPS1	NPS2	SRS
Ethnicity: NZ European	SRS	NPS2	NPS2
Ethnicity: NZ Maori	SRS	SRS	NPS2
Election '02 Vote: Labour	NPS1	SRS	N/A
Election '02 Vote: National	SRS	SRS	N/A
NPS best estimate in *x* of *y* cases	15 of 24	18 of 24	11 of 22

Note: SRS represents the Simple Random Sample scenario, NPS1 represents the 'No Adjustment' NPS scenario, and NPS2 represents the 'Stepped Adjustment' NPS scenario.

As for the frame variables, an analysis of the standard deviations of the estimates from each simulation scenario, for each survey variable, suggests that the NPS procedure generally increases the variability of results. Specifically, on average the scaled adjustment NPS procedure increased variability by 2% compared to that for the comparative SRS scheme (e.g., an average standard deviation of 10 for the SRS scheme would increase to 10.2 under the NPS scheme).

Again, then, the increase is not large. Moreover, the maximum increase for any one variable across all of the scenarios was 17%. Thus, the NPS scheme appears to have had even less of an effect on point estimate variation, as measured by the standard deviations of simulated estimates, for the survey variables than for the frame variables.

Together, these results lend moderate support to hypothesis 2, that the NPS scheme would consistently improve estimates without a substantive increase in estimate variability. Additionally, the mixed performance of the NPS procedure in the 2005 simulations is consistent with expectations given the relatively poor performance of the underlying propensity model for that scenario (as per hypothesis 4).

5.5. Interaction with Three Common Post-Survey Procedures

The final area of investigation involved the NPS scheme's interaction with three post-survey weighting procedures. First, age/sex adjustments were examined. Then, focus moved to the effect of the procedure on wave extrapolation.

These particular procedures were chosen because they are prevalent in postal survey practice, are relatively simple to implement and integrate into resampling simulations, and independent data were available at the levels required (i.e., both the frame data and New Zealand census parameters enabled cell calculations for age by sex). The aim was therefore not to undertake an exhaustive analysis of NPS scheme interaction with a wide variety of post-survey procedures, but rather to generate insights into the possible effect of an NPS implementation in a 'typical' survey practice situation.

Age/Sex Weighting

Table 36 presents the average weights applied to each age/sex band across the simulation studies based on either frame data or census parameters. Generally, the

NPS scheme led to lower weights for the younger age bands and higher weights for the older age bands, an effect expected given the shift in distribution of valids presented earlier. Overall, these changes counterbalanced each another such that the average weights across all bands were very similar. However, the NPS scheme did lead to slightly lower average weights overall and, more importantly, weights that were consistently closer to a baseline weight of 1 across the adjustment cells.

Table 36: Age/sex weights under each sampling scenario[†]

Weight Band		**Frame Weights**			**Census Weights**		
Age	Sex	SRS	NPS1	NPS2	SRS	NPS1	NPS2
18-29	F	1.41	1.31	1.21	1.58	1.48	1.38
18-29	M	1.67	1.56	1.42	1.84	1.75	1.62
30-39	F	1.00	0.98	0.95	1.02	1.01	0.98
30-39	M	1.31	1.29	1.26	1.40	1.38	1.35
40-49	F	0.89	0.91	0.93	0.82	0.83	0.85
40-49	M	1.05	1.05	1.07	1.10	1.11	1.13
50-59	F	0.97	1.00	1.03	0.84	0.86	0.90
50-59	M	0.93	0.94	0.95	0.92	0.94	0.95
60-69	F	0.77	0.77	0.80	0.67	0.68	0.70
60-69	M	0.89	0.89	0.90	0.74	0.74	0.75
70-79	F	0.85	0.89	0.93	0.91	0.96	0.99
70-79	M	0.73	0.78	0.81	0.67	0.71	0.73
80+	F	1.68	1.72	1.73	1.56	1.63	1.62
80+	M	1.11	1.09	1.11	0.80	0.77	0.78
Average weight		*1.09*	*1.08*	*1.08*	*1.06*	*1.06*	*1.05*
**Average deviation*		*0.23*	*0.20*	*0.18*	*0.30*	*0.28*	*0.25*

[†] Figures represent averages over the three simulated surveys for 2003, 2004 and 2005

* Average deviation from a weight of 1.0

Turning to the variability of cell weights across simulation replicates, there was no clear overall difference between the sampling scenarios. Specifically, although there were differences in the standard deviations between the scenarios at a cell level, these were associated with the shift in distribution of valids noted above (e.g., a shift to smaller/larger weights also led to a shift to smaller/larger standard deviations).

However, the scenarios all had very similar average standard deviations across the cells (0.18 for the frame weights and 0.19 to 0.20 for the census weights). Thus, despite returning lower numbers of valids overall, the NPS scheme did not increase the variability of weights established on those returns.

Given these results for the weights, one might expect that the NPS procedure would in turn improve weighted estimates. However, this was not the case. As tables 37 and 38 show, none of the sampling scenarios was clearly superior with respect to estimate accuracy compared to census data. Indeed, in many instances *none* of the weighted estimates were closer to census parameters than an unweighted SRS estimate (signalled by a dash '-' in the tables). Also, the estimates produced by the different sampling procedures post-weighting were very similar across most of the variables (see Appendix section A1.6, p. 172, for details); on average, they differed by just 0.3% for frame weighting and 0.4% for census weighting.

The poor performance of age/sex weighting under all sampling scenarios suggests that much of the problem lies with the reliability of this post-survey technique, rather than with the NPS scheme. Indeed, although it did not consistently reduce bias when combined with age/sex weighting, the NPS scheme did work to reduce variability in the simulated weighted point estimates in many cases (see Appendix section A1.6, p. 172, for the location of the data on the thesis supplementary CD). Furthermore, the scheme had no real impact on variability across the entire set of weighted survey variables (0.3% increase in standard deviations for frame weighting, 0.1% decrease for census weighting). Again, this is a positive result considering the NPS procedure results in fewer valid returns for a given sample size.

Table 37: 'Best scheme' results for survey estimates with frame weighting

	Scheme Resulting in Best Estimate		
Survey Variable	ISSP03	ISSP04	ISSP05
Gender (Male)	NPS2	SRS	SRS
Age 20-34	NPS2	NPS2	NPS2
Age 35-49	SRS	-	-
Age 50-64	NPS2	NPS1	SRS
Age 65+	SRS	NPS2	NPS2
Not Religious	NPS2	NPS1	-
Christian	NPS2	-	-
Employed Full Time	-	NPS2	-
Employed Part Time	-	SRS	NPS2
One Person Household	-	-	SRS
Two Person Household	NPS2	SRS	SRS
Three Person Household	SRS	-	-
Four Person Household	-	-	-
Five+ Person Household	-	-	-
Qualification Bachelor Degree+	-	-	-
Own Income <$20k	-	-	SRS
Own Income >$50k	-	-	SRS
Marital Status: Married	NPS2	NPS2	NPS2
Marital Status: Single	SRS	NPS2	NPS2
Marital Status: Widowed	SRS	SRS	SRS
Ethnicity: NZ European	SRS	NPS2	NPS2
Ethnicity: NZ Maori	-	-	SRS
Election '02 Vote: Labour	-	-	N/A
Election '02 Vote: National	SRS	-	N/A

Note: SRS represents the Simple Random Sample scenario, NPS1 represents the 'No Adjustment' NPS scenario, and NPS2 represents the 'Stepped Adjustment' NPS scenario.

Table 38: 'Best scheme' results for survey estimates with census weighting

	Scheme Resulting in Best Estimate		
Survey Variable	ISSP03	ISSP04	ISSP05
Gender (Male)	NPS1	SRS	SRS
Age 20-34	SRS	SRS	SRS
Age 35-49	NPS2	NPS1	-
Age 50-64	NPS2	NPS2	NPS2
Age 65+	SRS	SRS	-
Not Religious	SRS	-	-
Christian	SRS	-	-
Employed Full Time	-	-	-
Employed Part Time	-	SRS	NPS2
One Person Household	-	-	-
Two Person Household	NPS2	NPS2	SRS
Three Person Household	-	-	-
Four Person Household	-	-	-
Five+ Person Household	-	-	-
Qualification Bachelor Degree+	-	-	-
Own Income <$20k	-	-	-
Own Income >$50k	-	-	SRS
Marital Status: Married	NPS2	NPS2	NPS2
Marital Status: Single	NPS2	NPS2	NPS2
Marital Status: Widowed	NPS2	NPS2	SRS
Ethnicity: NZ European	-	NPS2	NPS2
Ethnicity: NZ Maori	-	-	SRS
Election '02 Vote: Labour	NPS1	-	N/A
Election '02 Vote: National	-	-	N/A

Note: SRS represents the Simple Random Sample scenario, NPS1 represents the 'No Adjustment' NPS scenario, and NPS2 represents the 'Stepped Adjustment' NPS scenario.

Wave-of-Response Extrapolation

Table 39 presents the results of wave-extrapolation based estimates across the simulation studies compared to census or election parameters.

Table 39: 'Best scheme' results for survey estimates with wave extrapolation

	Scheme Resulting in Best Estimate		
Survey Variable	ISSP03	ISSP04	ISSP05
Gender (Male)	NPS2	-	-
Age 20-34	NPS2	NPS1	NPS2
Age 35-49	-	-	-
Age 50-64	SRS	NPS2	NPS2
Age 65+	SRS	SRS	-
Not Religious	NPS2	NPS2	SRS
Christian	NPS2	NPS2	-
Employed Full Time	-	-	-
Employed Part Time	NPS1	SRS	NPS2
One Person Household	-	-	-
Two Person Household	SRS	NPS2	NPS1
Three Person Household	SRS	-	NPS1
Four Person Household	-	-	-
Five+ Person Household	-	-	-
Qualification Bachelor Degree+	-	-	-
Own Income <$20k	-	-	-
Own Income >$50k	NPS2	NPS1	-
Marital Status: Married	NPS2	NPS2	NPS2
Marital Status: Single	NPS2	NPS2	NPS2
Marital Status: Widowed	NPS1	SRS	-
Ethnicity: NZ European	-	-	SRS
Ethnicity: NZ Maori	-	-	-
Election '02 Vote: Labour	-	-	N/A
Election '02 Vote: National	-	-	N/A

Note: SRS represents the Simple Random Sample scenario, NPS1 represents the 'No Adjustment' NPS scenario, and NPS2 represents the 'Stepped Adjustment' NPS scenario.

As with the age/sex weighting, although the NPS scheme improved estimates in a number of cases (and in more cases than an extrapolated SRS), it did not lead to a consistent improvement in estimates across a wide range of the variables tested. This is not particularly surprising, given the poor performance of this post-survey adjustment technique in earlier analyses (see section 3.5.2, p. 77). Indeed, similar to the findings for age/sex adjustment, wave extrapolation failed to improve estimates under any sampling scenario in a substantial portion of cases (signified by a dash '-' in Table 39).

A critical factor in wave extrapolation is the slope of the regression line established across cumulative waves of response. Ideally, the slope would be zero for each variable, indicating no differences in respondents by wave and, if one is willing to assume late responders adequately represent nonresponders, suggesting that no nonresponse bias exists. One consequence of a sampling scheme that is able to return a more representative mix of valid responses may therefore be regression slopes that are closer to zero.

An examination of extrapolation equations from the simulations did not indicate that the NPS scheme leads to slopes that are closer to zero on average. In fact, the scheme appeared to slightly increase the gradient of the extrapolation lines, from a cross-simulation average deviation from zero of 7.3% for the SRS scheme to 7.6% for the NPS2 scheme. Furthermore, unlike the result for age/sex weighting, the NPS scheme increased the standard deviation of extrapolation gradients, from a cross-simulation average of 5.8% for the SRS scheme to 6.2% for the NPS2 scheme. The implication is that, when combined with wave extrapolation, the NPS scheme may lead to less stable adjusted estimates. This is probably because, although it returns a better demographic profile of valid returns overall, the scheme also alters the distribution of those returns across waves of response and exacerbates the differences between them.

For instance, in the 2005 survey simulations, 53.2% of all valid returns were received in the first wave under an SRS scheme, whereas 52.5% were received in the first wave under the NPS2 scheme. Similarly, 20.9% of all valid returns were received in the last (third) wave under an SRS scheme, whereas 21.6% were received under the NPS2 scheme.

Given the results for the post-survey adjustment procedures examined in this section, hypothesis 3 is not supported. Specifically, although the NPS scheme did increase the stability of age/sex weights and in some cases reduce variability in weighted estimates, it did not consistently lead to less biased results. However, as outlined earlier, the poor performance of the weighting techniques in general suggests it would be premature to assume this estimate inaccuracy is due to the NPS scheme. In fact, these findings reinforce the idea that researchers should expend effort minimising bias during the field period rather than relying on post-survey adjustments to improve estimates.

Nevertheless, it is possible that future research employing different post-survey adjustment approaches may achieve more positive results in combination with an NPS scheme. For example, other adjustment schemes may employ different base variables, differential adjustments for nonresponse components, or different post-stratification approaches such as raking or response propensity weighting.

5.6. A Promising Procedure

Overall, the results of the simulation study confirm that an NPS scheme can consistently reduce postal survey bias due to noncontact for a range of frame and survey items, at least in situations where other aspects of the survey sampling design are close to an SRS[19]. However, it cannot be said at this point that it reduces overall nonresponse bias substantively. Nevertheless, the scheme has a number of other positive attributes in that it does not appear to lead to large increases in estimate variability, has minimal 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. As a proof-of-concept, then, the general success of the prototype procedure developed and tested here suggests ongoing investigation and refinement of the technique is warranted.

[19] As noted earlier in section 4.4.2, p. 103, the estimated design effects for the simulation base survey datasets were close to 1.

Because the number of variables and surveys available for analysis in this particular piece of research was limited, it is not possible to comprehensively outline the range of items or survey situations for which the scheme may provide the greatest reduction in noncontact bias. Nevertheless, there is good reason to expect it will improve estimates for a number of variables of common interest to postal survey researchers. Specifically, in line with a conceptualisation of nonresponse bias that relates it to variables influencing individual response propensity (e.g., see Groves & Peytcheva, 2008, and Equation 9, below), bias reduction under the scheme will be greatest for those variables that correlate with noncontact propensity.

Equation 9: Nonresponse error incorporating response propensity[20]

$$\text{bias}(\bar{y}_r) = \text{Cov}(r_i, Y_i) + E\left[\left(\frac{m_s}{n_s}\right)(\bar{y}_r - \bar{Y})\right]$$

Where:

- $\bar{y}_r$ = Mean of the respondents within the s[th] sample for the variable of interest
- r_i = Probability of becoming a respondent
- Y_i = Values of the variable of interest
- m_s = Total number of nonrespondents in the s[th] sample
- n_s = Total number of sample members in the s[th] sample
- $\bar{Y}$ = Mean of the variable of interest

(source: Groves, Fowler et al., 2004, p. 182)

Those demographic variables identified as relating to movement in chapters 2 and 3, such as age, ethnicity, household composition and employment status, meet this criterion. Moreover, Table 26 (p. 77) presented a variety of ISSP survey variables that correlate with these demographics, including income, political engagement, level of social conservatism and views on indigenous issues. Thus, surveys that cover topics such as these are likely to benefit from the application of an NPS scheme, especially where age or household related sub-group comparisons are to be made.

[20] This equation is for a linear statistic, such as a mean or a proportion. The bias formula for nonlinear statistics are more complicated.

It is worth reiterating that the NPS scheme only aims to reduce noncontact bias. As such, the procedure is not intended as a panacea for substantive nonresponse issues related to refusal (passive or active) or ineligibility. Indeed, it is likely that, to the extent that propensity for noncontact and propensity for passive refusal positively covary for a given survey, the use of an NPS scheme will increase the proportion of net nonresponse bias attributable to passive refusal.

Evidence suggesting propensity for noncontact and propensity for passive refusal do positively covary, at least for the surveys examined in this research, was presented in chapter 2. However, the survey-dependent nature of passive refusal means that factors beyond those coincidentally related to noncontact will also influence propensity for passive refusal (e.g., survey topic, sponsor, questionnaire length, etc.). Thus, even where use of an NPS scheme does increase the proportion of net nonresponse bias attributable to passive refusal, it will enable a clearer view of the influence of these survey-dependent factors on nonresponse bias.

That is, whereas the noncontact reporting estimation techniques developed earlier in this research (see chapter 2) will help decompose the components of nonresponse *rates*, the NPS scheme is likely to assist researchers to at least partially decompose the components of postal survey nonresponse *bias*. Such decomposition mechanisms will become increasingly important as focus shifts away from interventions aimed solely at improving response rates and toward in-field procedures targeted directly at bias. Furthermore, the decomposition of postal survey bias components is likely to receive increased attention as many studies move to mixed mode designs.

6. Summary, Applications and Future Directions

6.1. Introduction

Nonresponse is of increasing concern to survey methodologists, who are facing general declines in response to survey requests. Indeed, the situation is such that the editors of a recent compilation of nonresponse research proposed that two key challenges facing methodologists at this juncture in history involve "*determining the circumstances under which nonresponse damages inference to the target population"* and identifying "*methods to alter the estimation process in the face of nonresponse to improve the quality of sample statistics"* (Groves et al., 2002, p. xiii).

In response to these challenges, practitioners and academics have begun to investigate the contribution to total survey bias of individual nonresponse components (e.g., refusal, noncontact, ineligibility). Furthermore, there has been a movement toward responsive survey designs that allocate increasingly limited survey resources to targeted interventions aimed at improving response or mitigating biasing effects at a component level. For example, a survey may include callbacks, incentives, refusal conversion techniques and mixed-mode contact strategies.

Although much has been done to better understand the causes and effects of nonresponse in telephone and face-to-face modes, relatively little is known about the components of postal survey nonresponse. In part, this is because it is difficult to separate out nonresponse components in that mode. Nevertheless, there is good reason to expect that the components contribute differentially to bias, just as they do in other modes. Indeed, as more postal surveys are deployed either as stand-alone vehicles or as part of mixed mode designs, researchers will require a better understanding of postal survey nonresponse components if they are to minimise overall survey bias and maximise effective use of survey resources.

This research therefore sought to examine the nature and extent of noncontact in the postal mode, to better understand its contribution to survey bias, and to explore the development of in-field interventions targeted at any bias associated with it.

The three overarching objectives of the research were to:

1. *Empirically estimate the levels of total noncontact present in the surveys examined and identify key correlates of both noncontact incidence and reporting*;
2. *Identify the direction and magnitude of postal survey bias introduced by noncontact and compare it to error introduced by other nonresponse components*;
3. *Investigate targeted in-field mechanisms for reducing postal survey bias introduced by noncontact.*

To achieve these objectives, a series of empirical studies was undertaken involving general population postal surveys fielded in New Zealand between 2001 and 2006. In addition to identifying a number of key features of the noncontact phenomenon, the research developed procedures for uncovering, estimating and adjusting for noncontact nonresponse that are directly applicable to postal survey practice.

6.2. Key Findings and Implications

6.2.1. Noncontact is Underreported and Systematic in Nature

Conceptually, postal noncontact was considered to be a survey-independent phenomenon related to individual propensity for movement, frame update processes, and household or individual propensities for notifying frame-keepers of changes. Furthermore, the level of noncontact reported to researchers was expected to relate to household propensity to return misaddressed mail and the attributes of the survey invitation.

A study embedded in a general population survey of 2,400 people confirmed many of these expectations (see section 2.3, p. 30). It did so by exploiting a unique frame update situation to identify addresses that were likely to be inaccurate at the time the survey was fielded and comparing these with ‘gone, no address’ (GNA) returns to the survey invitation. Independent frame information was also used to develop profiles of sample units more likely to change addresses and third parties more likely to report noncontact. Finally, the study tested the efficacy of a ‘please return’ message on the survey invitation envelope, aimed at increasing noncontact reporting rates.

Frame address inaccuracies were found to correlate with age, employment status and household composition. For instance, those who were young, living in multi-surname households, or who were students or beneficiaries were more likely to have changed address details. Furthermore, noncontact reporting related to household characteristics, such that those households more likely to contain movers were also less likely to report noncontact when it occurred.

Using a procedure developed as part of the study to estimate levels of unreported noncontact, it was found that follow-up mailings and envelope messages both significantly increased reporting by third parties. Moreover, results suggest that estimated total noncontact was as much as 400% higher than the reported level in a single-contact unmessaged study (2.8% vs. an estimated 12%). Indeed, even with three contacts and an envelope message, total noncontact was estimated to be 30% higher than reported (9.6% vs. an estimated 13%).

These findings have significant implications for survey practice. Specifically, noncontact appears to be drastically underestimated in standard postal surveys using frames such as an electoral roll. The cooperation rates reported for many postal studies are therefore likely to be understated. Moreover, the results also suggest noncontact is a much larger component of total postal survey nonresponse than typically acknowledged. Given widespread concern about declining survey response, this is important to know. Efforts aimed at understanding the reasons for declining response, identifying any associated bias, or developing tools to combat the problem, all require knowledge of the size and nature of nonresponse components. Both the envelope message technique and the notification rate estimation procedure developed as part of this study contribute to the development of that knowledge.

Finally, the interrelationships identified between mobility and noncontact, and household characteristics and reporting, present opportunities for targeted design interventions to be developed for this component of postal survey nonresponse. These might, for example, modify reporting propensities by the manipulation of survey features under the researcher’s control (e.g., the survey invitation) or incorporate expected noncontact propensities into the survey design (e.g., at the sampling phase) to reduce the effect of this potential error source on estimates.

6.2.2. Noncontact Contributes to Net Survey Bias

Notwithstanding the findings regarding the underreporting and systematic nature of noncontact, noncontact does not necessarily contribute error to survey estimates. Moreover, if it does, it is possible that any error is either the same as, or entirely offset by, that contributed by other nonresponse components. Therefore, an empirical study was undertaken to identify the direction and level of postal survey noncontact bias and to compare it to error introduced by other nonresponse components (see section 3.3, p. 65).

The study, which examined a selection of general population surveys fielded between 2001 and 2006, established estimates of bias due to noncontacts, active refusals, ineligibles, and inactives (respondents from whom no response at all was received). Multiple techniques for estimating error were employed, including benchmarking against population parameters, comparisons on individual-level frame data, and analysis of valid responses over successive waves of contact.

Steady trends in bias were identified across the surveys. Specifically, survey estimates were consistent in their direction of deviation from known population data on age, gender, Maori ethnicity, marital status, qualifications, income and household composition. As noted in the prior study, and in recent research published by *Statistics New Zealand* (2007j), many of these variables are known to relate to movement and noncontact. Furthermore, the trends in deviations identified persisted across variables for which individual frame data were also available (e.g., age, gender, Maori descent). Since a frame-level analysis eliminates the possibility for identified deviations to be due to coverage or measurement, this suggests that the deviations on related survey-only variables are also at least in part due to nonresponse.

At the component level, sample units for which a refusal or ineligible response was recorded differed substantially from those listed as a noncontact or inactive, on average, over a range of frame variables. Moreover, although the different component biases cancelled each other out to some degree, net nonresponse bias remained and was attributable to the noncontact and inactive groups. Indeed, an analysis of bias on the frame variables taking into account noncontact underreporting rates established in the prior study suggested that up to 40% of residual nonresponse bias after multiple follow-up contacts may be contributed by noncontact.

It was not possible to assess the degree of bias caused by noncontact on survey-only variables, because comparative population parameters were either unavailable or potentially confounded by coverage and measurement error. Furthermore, an attempt to assess bias magnitude by wave-of-response extrapolation proved too unreliable to generate any sound conclusions. Nevertheless, there were clear correlations between many of the frame variables for which nonresponse bias was known to exist, and a range of demographic and attitudinal survey-only items. Thus, it seems reasonable to conclude that noncontact bias affects a variety of variables.

Together, these findings point to a clear opportunity for methods targeted at reducing noncontact bias to improve final survey estimates for a range of items.

6.2.3. Practical Issues Limit the Options for Targeted In-Field Noncontact Interventions

In general, two approaches to bias reduction are proposed in the nonresponse literature; post-survey adjustment via techniques such as weighting or imputation, and in-field design interventions aimed at improving the distribution of responses. Although both are commonly employed, the success of post-survey approaches ultimately rests on the amount of data gathered during the field period and the validity of assumptions about the relationship between responders and nonresponders. Thus, where possible, researchers are advised to adopt the responsive design approach to fieldwork mentioned earlier, and to allocate resources to in-field interventions targeted at low-response groups.

With this in mind, an exploration of potential in-field mechanisms targeted at noncontact nonresponders was undertaken. A search of the literature for techniques that could be modified for such a purpose yielded four potential methods: finding and subsampling noncontacts, sampling movers from an independent source, substitution from within mover households, and sampling based on propensity to be a noncontact. Of these, the first three were found to have significant limitations in the postal mode, at least in a New Zealand context. For instance, evidence from an attempt to find and survey noncontacts to one study suggests this approach is unlikely to succeed in obtaining data for many noncontacted individuals, within the budgetary constraints of many postal surveys. Moreover, it appeared that an independent list of movers available to New Zealand researchers would suffer from significant coverage issues, and thus, would be unsuitable for substitution purposes. Similarly, a small study that attempted substitution from within noncontact households did not deliver an adequate or representative set of replacements.

However, the fourth option, noncontact propensity sampling (NPS), was found to have both a compelling theoretical foundation and potential for wide practical applicability. The procedure is based on the rationale that, because noncontact is survey independent, sample units with similar propensities for noncontact should be substitutable with respect to survey items. Hence, noncontact bias may theoretically be eliminated by altering the survey sampling weights based on individual propensity for noncontact. That is, potential respondents may be sampled in proportion to their likelihood of noncontact, with the aim of achieving a contacted sample that is equivalent, on average, to a random sample taken from a frame with no contact inaccuracies.

The practicality of an NPS scheme relies on researchers' ability to predict noncontact together with a clear procedure for turning these predictions into sampling weights. In order to further explore these factors, a propensity modelling study was undertaken using data from the six surveys examined in earlier work (see section 4.4, p. 99). As anticipated, a number of demographic and household variables previously identified as correlates of movement and noncontact were consistently retained in logistic regression models built to predict reported noncontact. Furthermore, the models, which related to different base datasets and field periods, each performed similarly with respect to predictive power. In addition to establishing that noncontact may be predicted using common frame-based variables, the consistency in the performance of the models lent support to the idea that noncontact propensity is a survey-independent phenomenon.

A strata-based procedure was adopted as the most practical means of translating propensity scores into adjusted sampling weights for postal noncontact, for two reasons. First, results from the modelling indicated that the distribution of propensity scores would require it. This has also been the experience of other researchers employing propensity adjustment for nonresponse to internet surveys or undercoverage in telephone samples. Second, the fact that noncontact is often underreported means further adjustments to the sampling weights must be made to take this into account. In many situations a strata-based procedure will be the most conducive to such an adjustment.

There are some clear advantages to this approach compared to the other targeted mechanisms explored. Specifically, an NPS scheme:

- Is more cost-effective than procedures that require follow-up of noncontacts. In particular, organisations that undertake multiple surveys from the same frame could expend effort building a noncontact propensity model which they could then apply across multiple surveys;
- Is founded on unambiguous and defensible assumptions;
- Allows use of a single frame for sourcing all sample units, thereby eliminating the potential for coverage error to be compounded across sub-samples;
- Maintains a probability-based sampling procedure that can be specified and documented, and potentially used in combination with other probability procedures.

As such, of the potential methods identified, the NPS scheme comes closest to the ideal of an in-field design intervention that is "*practical, cheap, effective, and statistically efficient*" (Kish & Hess, 1959, p. 17).

6.2.4. An NPS Scheme can Reduce Noncontact Bias

Although the NPS scheme appeared to hold the greatest potential of the alternatives examined, the limits of predictive models, along with a strata-based weighting scheme and the need to adjust for underreporting, was likely to mean it could not totally eliminate noncontact bias. Thus, an empirical test of the scheme was undertaken to assess its likely practical effect on estimates.

Specifically, a simulation study was carried out using bootstrap resampling of results from three surveys fielded between 2003 and 2005 (see section 5.2, p. 114, for details). Two different NPS scheme implementations were tested (no adjustment for underreporting and stepped adjustment for underreporting). Furthermore, a parallel simulation for each survey employing an SRS scheme was conducted for comparison. Each simulation run involved 1,000 sample replicates of size 1,500, with summary measures calculated for a range of frame and survey variables for which independent data existed. In addition, three common post-survey weighting procedures (frame age/sex weighing, census age/sex weighting and wave extrapolation weighting) were applied as part of the simulation to examine the effect of the NPS scheme on the adjusted estimates they produce.

The results of the simulation suggest that the NPS scheme altered response distributions such that a higher proportion of the sample generates inactive or GNA outcomes, while a lower proportion generates valid or refusal outcomes. This effect was most pronounced for the stepped adjustment version of the NPS scheme thought to be most representative of a real-world implementation. Although it was expected, the obvious consequence is that standard cooperation rates for NPS samples will be lower than a comparative SRS. However, the effect on this important survey metric was relatively small and becomes trivial if an appropriate adjustment for unreported noncontact is made (as outlined in chapter 2). Furthermore, what is ultimately important is whether or not the scheme results in better survey estimates.

In that regard, the NPS scheme did consistently produce a superior profile of valid responders than an SRS scheme when compared on independent frame data, with the stepped adjustment version generating the greatest reduction in nonresponse bias (an average 28% reduction in absolute error over the three simulated surveys). Furthermore, the NPS scheme produced survey item estimates closer to known census figures than the comparative SRS over a range of variables and survey periods. In particular, the scheme appeared to consistently improve estimates for age, religiosity, household size, qualifications, income, and marital status for two of the three simulated survey scenarios. However, the amount of bias reduction achieved was not as large as for the frame values; where it improved survey item estimates, the scheme led to an average 17% reduction in absolute error. The scheme also performed worst in the 2005 survey simulations, as expected given the lower predictive power of the propensity model developed for that period.

With respect to variability, although on average the NPS procedure increased the standard deviation of simulated point estimates, the effect was relatively small. For frame variables, an average 4% increase was observed (e.g., a standard deviation of 10.0 would increase to 10.4). For the survey variables examined, the increase was 2% on average. This is a positive result considering the scheme returns fewer valid responses than an SRS for a given initial sample size (1,500 in this case).

The scheme did not lead to substantive improvements in estimates when paired with frame age/sex weighing, census age/sex weighting or wave extrapolation weighting. But this result is more of a reflection of the shortcomings of these common post-survey weighting procedures than it is of the utility of the NPS scheme. Indeed, in many cases weighting or wave extrapolation combined with *either* sampling scheme (SRS or NPS) actually made the survey estimates worse.

Overall, the NPS shows promise as a targeted in-field mechanism for reducing noncontact bias in postal surveys and therefore warrants further developmental effort. It is likely to have the greatest impact where noncontact can be expected to be a nontrivial component of total nonresponse, where the sampling adjustment is based on a strong predictive model, and where the survey covers variables that are likely to covary with items known to relate to propensity for noncontact (including age, household composition, employment status and ethnicity).

Furthermore, given that the procedure reduces error due to noncontact but is not expected to affect error due to other nonresponse components (although it may alter the proportion contributing to net nonresponse bias), it will allow researchers to at least partially decompose the various facets of nonresponse bias. In turn, this is likely to contribute to the development of in-field procedures targeted at reducing bias due to these other sources. As more studies move to mixed mode designs incorporating self-administration by post, such development work is expected to become the focus of increased attention.

6.3. Potential Applications

The insights and procedures relating to noncontact contributed by this research have a range of potential practical applications. Three general domains to which the knowledge generated here could be applied are postal survey methodology, online survey methodology and organisational database management.

6.3.1. Postal survey methodology

Knowledge of the correlates of noncontact, methods for improving reporting or estimating the level of reporting, and procedures for reducing its associated bias will be useful across countries and disciplines of research. For example, depending on access restrictions, researchers in countries with voter or population registers such as Australia, Finland, Norway and the United Kingdom may be able to closely follow the process outlined here to develop country-specific estimates of noncontact and reporting rates, and models of noncontact propensity. These could then be applied to a variety of postal surveys in those locations. Indeed, even in countries without such frames, or for organisations without access to them, it is likely that minor variations on the approach presented here (e.g., using internal lists or other publically available frames) may be applied to gather more information about noncontact than is currently available. Certainly, the findings regarding envelope message effects on reporting rates and procedures for estimating total noncontact should be generally applicable to a wide variety of postal survey situations.

Organisations undertaking panel or longitudinal research would probably gain the most benefit from the implementation of an NPS scheme, since they will have ready access to prior response data and use a consistent frame over time. Moreover, the cost of developing a model could be amortised across surveys. Nevertheless, since noncontact is a survey independent phenomenon, there is also the potential for industry-level development of guidelines relating to noncontact reporting rates and key predictors of noncontact for commonly employed frames. If these were to be developed, individual researchers could make use of them at relatively little cost in their own one-shot studies.

6.3.2. Online survey methodology

Although the results of the empirical studies presented here cannot be directly applied to an online context, the self-administered and individualised nature of many online surveys means there are parallels between the online and postal modes. Furthermore, the relatively recent advent of online media means there are many aspects of methodology that require further investigation in this mode. One such aspect relates to the receipt of, and response to, email invitations. It is likely that several of the issues addressed in this project will also be faced by researchers examining nonresponse to online questionnaires (e.g., understanding reporting of noncontact, estimating total noncontact levels, identifying bias and developing targeted in-field interventions).

There are, of course, clear differences in an online setting because a certain amount of noncontact reporting is automatic (i.e., email bounce-back). Yet, there is still a decomposition problem with respect to inactive nonresponse due to abandoned email accounts that still accept messages, messages that are received but caught up in spam filters, or passive refusal by those who see the invitation. Thus, the general approach to identifying and mitigating the effects of noncontact developed here will be useful to researchers in that field of inquiry. For instance, technical methods exist for identifying when an email message is opened in certain situations. These might be employed to establish estimates of passive refusal versus unreported noncontact that could assist with cooperation calculation and help develop a better understanding of online survey nonresponse.

It is also worth noting that some researchers are employing postal invitations in online studies to overcome coverage problems with email frames (if these are available at all). In such cases, many of the findings and procedures presented here may be directly applicable.

6.3.3. Organisational database management

Postal survey noncontact represents a special case of a much broader issue – mail nonreceipt caused by individual physical contact information inaccuracies due to population movement. This affects a range of activities, including corporate communications with customers, governmental notifications or requests to citizens and organisational messages to members or subscribers. In all of these endeavours the costs associated with noncontact may be reduced by targeted list maintenance activities focusing on those records most likely to change. Many of the procedures developed for this project should be helpful in this regard. For example, organisations undertaking ongoing data cleaning exercises could routinely incorporate a prominent envelope message on their communications to better identify invalid addresses.

Furthermore, initial individual data capture requirements might include data known to be predictive of noncontact propensity (e.g., household composition, employment status). Models built on such data would have a range of uses, including predicting likelihood of membership churn for geographically focused services or selecting nonrespondent individuals for follow-up by more expensive alternative contact methods.

Turning to external lists, organisations may also find use for the results regarding noncontact reporting rates in their data purchase decisions. For instance, many publicly available lists have ‘per record’ costs based on a number of factors, one of which is the contact rate for the list. Given these rates are very likely to *overestimate* the true contact rate, organisations could use knowledge of noncontact underreporting correlates to better assess the comparative costs of alternatives.

6.4. Limitations and Directions for Future Research

While effort was made in the research to achieve robust results by examining multiple methods and datasets, there were inevitably limits to the scope of investigation that could be undertaken. Thus, there are a number of aspects of the research presented in this thesis that would benefit from replication and extension.

Specifically, all of the datasets employed relate to general population surveys of individuals undertaken in New Zealand and sponsored by *Massey University*. Hence, it is possible that some results may vary in different countries and for different sponsorship organisations. For instance, the dynamics of population movement, frame maintenance and postal service efficiency could alter the correlates of noncontact incidence and reporting rates in different settings. Furthermore, surveys sponsored by different organisations may not achieve the same levels of response to survey requests or envelope prompts. Additional research is therefore required to assess the effect of changes in these factors.

Similarly, the bias analysis in this research focused on simple linear statistics (e.g., means and proportions) and was based on data obtained from surveys with simple sampling designs (i.e., where stratification was employed in base surveys, the estimated design effects were generally close to 1). As such, it is unknown how bias in nonlinear statistics, or for linear statistics in surveys with more complex base designs, would be affected by noncontact nonresponse or the NPS schemes explored here.

There was also limited examination undertaken of potential noncontact reporting rate influencers. It is possible that future research may lead to significant improvements in reporting cues. For instance, there are likely to be interactions between the existence of an envelope message and other attributes of the invitation package (e.g., logos, envelope size, bulk of package, etc.) that alter the achieved reporting rate. Furthermore, variations in message wording may lead to improved response. With regard to this, additional research incorporating diffusion of responsibility theory may bear fruit.

Further research on, and improvements in, noncontact reporting rate stimuli are also likely to have a flow-on effect for the continued development of total noncontact estimation procedures. Such development will be important because estimates of underreporting are a critical input for targeted in-field interventions such as an NPS scheme. Although it was not examined further in this research because of sample size limitations, an ideal progression of development would lead to household-level models of reporting propensity. Since these could be incorporated at an individual level along with survey response data, they may lead to improvements in the prediction of noncontact propensity that would, in turn, improve the bias-reducing effects of an NPS scheme.

Other areas that could be explored with the aim of improving NPS scheme performance include frame data augmentation (e.g., small area census data may provide additional predictive variables that would realistically be available to a variety of researchers), the incorporation of frame snapshot age into estimates of noncontact propensity, and the development of propensity models using techniques other than logistic regression (e.g., discriminant analysis, Bayesian classification or neural networks). To the extent that exploration of these approaches leads to models with greater predictive power, gains in noncontact bias reduction should also be achieved.

An alternative approach to predicting noncontact, not explored here, may be to move away from the use of prior survey response data entirely. For example, it may be possible to develop models of movement using publicly available data, which could then be applied to survey frames as a proxy for expected noncontact. In New Zealand, this might be achieved by the use of limited individual-level census record sets such as the *Confidentialised Unit Record Files* (Statistics New Zealand, 2007a), which include a range of demographic variables and a length of residence indicator. However, although such an approach would circumvent the problem of noncontact underreporting, it could potentially introduce issues relating to discrepancies between the model development data and variables available on the survey frame.

Moving beyond the specific issues and approaches that were the focus of this research, there are a number of related areas that could build on the findings presented here. For example, knowledge of the bias contributions of the various nonresponse components is likely to be useful in the development of post-survey adjustment procedures that account for differences; the second approach to postal survey bias reduction recommended by Mayer and Pratt.

> *"Inasmuch as the biases tend to be offsetting for certain characteristics, the researcher who has carefully segmented nonresponse by source could minimize total nonresponse bias by (1) controlling the relative sizes of offsetting nonresponse segments, or by (2) applying differential weights based on the relative sizes of these segments."* (Mayer & Pratt, 1966, p. 644)

Such post-adjustment measures might specifically include items in the questionnaire (e.g., a question on length of residence or recency of movement) to facilitate differential weighting for noncontact. Certainly, the poor performance of the common post-survey weighting procedures examined as part of the nonresponse error (chapter 3) and NPS simulation (chapter 5) studies suggests work examining the bias-reducing efficacy of alternative methods is needed.

Another area potentially worth exploration is the effect of moving from individual to household level selection for postal surveys. For instance, many of the techniques employed by telephone researchers to generate pseudo-random samples of the population might be applicable to a postal setting. If so, the problem of noncontact may be avoided completely. However, other significant issues are likely to arise from this approach that could outweigh any benefits gained, such as a reduction in cooperation, distortions in sample representativeness and problems in determining selection weights.

Finally, coverage is another source of postal survey error that might be addressed via the sampling adjustment technique explored in this research. Certainly, researchers have employed post-survey propensity weighting in the telephone mode in an attempt to reduce coverage bias (e.g., Duncan & Stasny, 2001). Furthermore, it seems reasonable that at least some of the noncoverage in frames such as the electoral roll would be due to population movement (e.g., when a noncontact return is used to remove a record from the frame). Thus, there are likely to be a number of parallels between noncontact and noncoverage in the postal mode that mean advances in targeting and reducing the bias in one can be applied in some form to the other.

7. References

Anthony, B. A. (2002). *Performing logistic regression on survey data with the new SURVEYLOGISTIC procedure (Paper 258-27)*. Paper presented at the SUGI27. Retrieved 01 February 2008, from http://www2.sas.com/proceedings/sugi27/p258-27.pdf.

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research, 14*(3), 396-402.

Asch, D. A., Jedrziewski, M. K., & Christakis, N. A. (1997). Response rates to mail surveys published in medical journals. *Journal of Clinical Epidemiology, 50*(10), 1129-1136.

Baisden, K. L., & Hu, P. (2006). *The enigma of survey data analysis: Comparison of SAS® survey procedures and SUDAAN procedures (Paper 194-31)*. Paper presented at the SUGI31. Retrieved 01 February 2008, from http://www2.sas.com/proceedings/sugi31/194-31.pdf.

Battaglia, M. P., Izrael, D., Hoaglin, D. C., & Frankel, M. R. (2004). Tips and tricks for raking survey data (A.K.A. sample balancing) [Electronic Version]. *Paper presented at the annual meeting of the American Association for Public Opinion Research*. Retrieved May 11 from http://www.amstat.org/Sections/Srms/Proceedings/y2004/files/Jsm2004-000074.pdf.

Bednall, D., & Shaw, M. (2003). Changing response rates in Australian market research. *Australasian Journal of Market Research, 11*(1), 31-41.

Best, S. J., & Radcliff, B. (2005). *Polling America: An encyclopedia of public opinion* (Vol. 1). Westport, Connecticut: Greenwood Press.

Braunsberger, K., Gates, R., & Ortinau, D. J. (2005). Prospective respondent integrity behavior in replying to direct mail questionnaires: A contributor in overestimating nonresponse rates. *Journal of Business Research, 58*(3), 260-267.

Brennan, M., & Hoek, J. (1992). The behavior of respondents, nonrespondents, and refusers across mail surveys. *Public Opinion Quarterly, 56*(4), 530-535.

Cham, J. (2001). Newton's three laws of graduation. *Piled Higher & Deeper: A Grad Student Comic Strip.* Retrieved 03 June, 2008, from http://www.phdcomics.com/comics/archive.php?comicid=221

Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and researchers* (2nd ed.). New Jersey: John Wiley & Sons, INC.

Churchill, G., & Iacobucci, D. (2005). *Marketing research: Methodological foundations* (9th ed.). Mason, Ohio: Thomson South-Western.

Clausen, J., & Ford, R. (1947). Controlling bias in mail questionnaires. *Journal of the American Statistical Association, 42*(240), 497-511.

Colombo, R. (2000). A model for diagnosing and reducing nonresponse bias. *Journal of Advertising Research, 40*(1/2), 85-93.

Couper, M. P. (2000). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review, 18*(4), 384-396.

Craighill, P. M., & Dimock, M. (2005). Tough calls: Potential nonresponse bias from hard-to-reach respondents. *Public Opinion Pros. Accessed via http://www.publicopinionpros.com/from_field/2005/aug/craighill.asp on 12-03-2007*.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69*(1), 87-98.

Czajka, J. L., Hirabayashi, S. M., Little, R. J. A., & Rubin, D. B. (1992). Projecting from advance data using propensity modeling: An application to income and tax statistics. *Journal of Business & Economic Statistics, 10*(2), 117-131.

Davison, A., & Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

de Leeuw, E. (1999). Preface. *Journal of Official Statistics, 15*(2), 127-128.

de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 41-54). New York: John Wiley & Sons, Inc.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American, 248*(5), 116-130.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons, Inc.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method (2007 update with new internet, visual, and mixed-mode guide)* (2nd ed.). New York: John Wiley & Sons, Inc.

Dillman, D. A., & Carley-Baxter, L. (2000). Structural determinants of mail survey response rates over a 12-year period, 1988-1999 [Electronic Version]. *Proceedings of the Section on Survey Methods, The American Statistical Association.* Retrieved 01 July 2006 from http://survey.sesrc.wsu.edu/dillman/papers.htm.

Dillman, D. A., & Miller, K. J. (1998). Response rates, data quality, and cost feasibility for optically scannable mail surveys by small research centers. In M. P. Couper, R. P. Baker, J. Bethlehem, C. N. F. Clark, J. Martin, W. L. Nicholls &

J. M. O'Reilly (Eds.), *Computer-assisted survey information collection* (pp. 476-497). New York: Wiley.

Dommeyer, C. J., Elganayan, D., & Umans, C. (1991). Increasing mail survey response with an envelope teaser. *Journal of the Market Research Society, 33*(2), 137-140.

Duncan, K. B., & Stasny, E. A. (2001). Using propensity scores to control coverage bias in telephone surveys. *Survey Methodology, 27*(2), 121-130.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: Systematic review. *BMJ (British Medical Journal), 324*, 1183-1192.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics, 7*(1), 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other resampling plans*. Cambridge: Society for Industrial and Applied Mathematics.

Efron, B., & Tibshirani, R. (1993). *An introduction to the Bootstrap*. New York: Chapman & Hall/CRC.

Elections New Zealand. (2005). How to enrol. Retrieved 20 November, 2006, from http://www.elections.org.nz/how_to_enrol.html

Ellis, R., Endo, C., & Armer, J. (1970). The use of potential nonrespondents for studying nonresponse bias. *The Pacific Sociological Review, 13*(2), 103-109.

Esslemont, D., & Lambourne, P. (1992). The effect of wrongly addressed mail on mail survey response rates. *Marketing Bulletin, 3*, 53-55.

Filion, F. L. (1975). Estimating bias due to nonresponse in mail surveys. *Public Opinion Quarterly, 39*(4), 482-492.

Filion, F. L. (1976). Exploring and correcting for nonresponse bias using follow-ups of nonrespondents. *The Pacific Sociological Review, 19*(3), 401-408.

Fisher, M. R. (1996). Estimating the effect of nonresponse bias on angler surveys. *Transactions of the American Fisheries Society, 125*(1), 118–126.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rates: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*(4), 467-491.

Frankel, L. R. (1982). *On the definition of response rates: A special report of the CASRO Task Force on completion rates.* New York: Council of American Survey Research Organizations (CASRO).

Fuller, C. H. (1974). Weighting to adjust for survey nonresponse. *Public Opinion Quarterly, 38*(2), 239-246.

Gendall, P., Hoek, J., & Finn, A. (2005). The behaviour of mail survey non-respondents. *Australasian Journal of Market and Social Research, 13*(2), 39-50.

Goksel, H., Judkins, D. R., & Mosher, W. D. (1992). Nonresponse adjustments for a telephone follow-up to a national in-person survey. *Journal of Official Statistics, 8*(4), 417-431.

Gray, A., Haslett, S., & Kuzmicich, G. (2004). Confidence intervals for proportions estimated from complex sample designs. *Journal of Official Statistics, 20*(4), 705-723.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons, Inc.

Groves, R. M. (2004). *Survey errors and survey costs*. New Jersey: John Wiley & Sons, Inc.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675.

Groves, R. M., Benson, G., Mosher, W., Rosenbaum, J., Granda, P., Axinn, W., Lepkowski, J., & Chandra, A. (2005). Plan and operation of Cycle 6 of the National Survey of Family Growth. *Vital Health Stat 1, 42*, 1-86.

Groves, R. M., & Brick, M. J. (2006). Practical tools for nonresponse bias studies, *Notes for a one-day short course sponsored by the Joint Program in Survey Methodology.* Presented on May 21 at the American Association of Public Opinion Research Conference, Montreal, Canada.

Groves, R. M., & Couper, M. P. (1995). Theoretical motivation for post-survey nonresponse adjustment in household surveys. *Journal of Official Statistics, 11*(1), 93-106.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2002). *Survey nonresponse*. New York: John Wiley & Sons, Inc.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New Jersey: Jon Wiley & Sons, Inc.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(3), 439-457.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72*(2), 167-189.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Hair, J. F., Bush, R. P., & Ortinau, D. J. (2006). *Marketing research within a changing information environment* (3rd ed.). New York: McGraw-Hill/Irwin.

Healey, B., & Gendall, P. (2005). Understanding mail survey non-contact: An examination of misaddressed survey invitation returns. *Australasian Journal of Market and Social Research, 13*(1), 37-45.

Hoek, J. (2006). RE: GNA study documents. Email regarding administration of a GNA follow-up study in conjunction with a postal survey relating to Attitudes to Advertising. In B. Healey (Ed.). Palmerston North: Personal Communication.

Holt, D., & Elliot, D. (1991). Methods of weighting for unit non-response. *The Statistician, 40*(3), 333-342.

Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression* (2nd ed.). New York: Wiley-Interscience.

Hox, J., & de Leeuw, E. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity, 28*, 329-344.

Hutt. (1982). Wrongly addressed mail. *Survey Methods Newsletter, 2*.

International Social Survey Programme. (2003). Study descriptions for participating countries. *International Social Survey Programme 2003: National Identity II (ISSP 2003)* Retrieved 29 February, 2008, from http://www.za.uni-koeln.de/data/en/issp/nspub/ZA3910_StudyDescriptions.pdf

International Social Survey Programme. (2004). Study descriptions for participating countries. *International Social Survey Programme 2004: Citizenship (ISSP 2004)* Retrieved 29 February, 2008, from http://www.za.uni-koeln.de/data/en/issp/nspub/ZA3950_StudyDescriptions.pdf

International Social Survey Programme. (2005). Study descriptions for participating countries. *International Social Survey Programme 2005: Work Orientation III (ISSP 2005).* Retrieved 29 February, 2008, from https://info1.za.gesis.org/dbksearch12/download.asp?id=12748

Jones, W., & Lang, J. (1980). Sample composition bias and response bias in a mail survey: A comparison of inducement methods. *Journal of Marketing Research, 17*(1), 69-76.

Kalton, G. (1983). *Compensating for missing survey data*. Ann Arbor: Survey Research Center, University of Michigan.

Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*(4), 440-453.

Kennedy, J. (1993). A Comparison of Telephone Survey Respondent Selection Procedures. [Electronic Version]. *Presented at the American Association for*

Public Opinion Research (AAPOR) Annual Meeting. Retrieved 29 November 2007 from http://www.indiana.edu/~csr/aapor93.html.

Kessler, R. C., Little, R. J. A., & Groves, R. M. (1995). Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiologic Reviews, 17*(1), 192-204.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association, 44*(247), 380-387.

Kish, L. (1995). *Survey sampling* (Wiley Classics Library ed.). New York: John Wiley & Sons.

Kish, L., & Hess, I. (1959). A "replacement" procedure for reducing the bias of nonresponse. *The American Statistician, 13*(4), 17-19.

Koch, A., & Porst, R. (Eds.). (1998). *Nonresponse in survey research. Proceedings of the eighth international workshop on household survey nonresponse.* (Vol. Spezial No. 4). Mannheim: ZUMA-Nachrichten.

Lankford, S. V., Buxton, B. P., Hetzler, R., & Little, J. R. (1995). Response bias and wave analysis of mailed questionnaires in tourism impact assessments. *Journal of Travel Research, 33*(4), 8-13.

Lavrakas, P. J. (1996). To err is human. *Marketing Research, 8*(1), 30-36.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics, 22*(2), 329-349.

Lee, S., & Valliant, R. (2007). Chapter 8: Weighting telephone samples using propensity scores. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 170-183). Published Online: John Wiley & Sons, Inc.

Lin, I.-F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly, 59*(2), 236-258.

Little, R. J. A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association, 88*(423), 1001-1012.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Little, R. J. A., & Vartivarian, S. L. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology, 31*(2), 161-168.

Lynn, P. (1996). Weighting for non-response. In R. Banks (Ed.), *Survey and Statistical Computing* (pp. 205–214). London: Association for Survey Computing.

Lynn, P. (2003). PEDAKSI: Methodology for collecting data about survey non-respondents. *Quality and Quantity, 37*(3), 239-261.

Lynn, P. (2006). Editorial: Attrition and non-response. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(3), 393-394.

Lynn, P., & Clarke, P. (2002). Separating refusal bias and non-contact bias: evidence from UK national surveys. *Journal of the Royal Statistical Society Series D (The Statistician), 51*(3), 319-333.

Mangione, T. W. (1995). *Mail surveys: Improving the quality*. Thousand Oaks, CA: Sage Publications Inc.

Matthews, C. (2006). Survey followup. Email regarding outcome of an attempt to follow up GNA returns to the first wave of a postal survey. In B. Healey (Ed.). Palmerston North: Personal Communication.

Mayer, C., & Pratt, R. (1966). A note on nonresponse in a mail survey. *Public Opinion Quarterly, 30*(4), 637-646.

McKnight, P. E., McKnight, K. M., Sidani, S., & Fiqueredo, A. J. (2007). *Missing data: A gentile introduction*. New York: The Guilford Press.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference* Newbury Park, CA: Sage Publications.

Moore, D., & Tarnai, J. (2002). Evaluating nonresponse error in mail surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 197-211). New York: John Wiley & Sons, Inc.

Morwitz, V. G. (2005). The effect of survey measurement on respondent behaviour. *Applied Stochastic Models in Business and Industry, 21*(4-5), 451-455.

Neff, J. (2006). They hear you knocking, but you can't come in. *Advertising Age, 77*(40), pp. 1, 49.

New Zealand Electoral Enrolment Centre. (2005). Election day countdown to getting enrolled. Retrieved 26 Feb 2007, from http://www.elections.org.nz/news/enrol-for-election-day.html

New Zealand Post Limited. (2008). New movers mailing list. Retrieved 11 February, 2008, from http://www.nzpost.co.nz/Cultures/en-NZ/Business/DirectMarketing/DMResources/NewMoversMailingList.htm

Oh, H. L., & Scheuren, F. S. (1983). Weighting adjustments for unit non-response. In W. Madow, H. Nisselson & I. Olkin (Eds.), *Incomplete Data in Sample Surveys.* (Vol. 2. Theory and Bibliographies., pp. 143-184). New York: Academic Press, Inc.

Pace, C. (1939). Factors influencing questionnaire returns from former university students. *Journal of Applied Psychology, 23*(3), 388–397.

Rao, J. N. K., & Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association, 83*(401), 231-241.

Reid, S. (1941). *The classroom audience of network school broadcasts*. Columbus, Ohio: Ohio State University.

Reid, S. (1942). Respondents and non-respondents to mail questionnaires. *Educational Research Bulletin, 21*(4), 87-96.

Rollins, M. (1940). The practical use of repeated questionnaire waves. *Journal of Applied Psychology, 24*, 770-772.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association, 82*(398), 387-394.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Shaw, M., Bednall, D., & Hall, J. (2002). A proposal for a comprehensive response-rate measure (CRRM) for survey research. *Journal of Marketing Management, 18*(5/6), 533-554.

Shettle, C., & Mooney, G. (1999). Monetary incentives in US government surveys. *Journal of Official Statistics, 15*(2), 231-250.

Shuttleworth, F. (1940). Sampling errors involved in incomplete returns to mail questionnaires. *Annual meeting of the American Psychological Association, Pennsylvania State College, and summarized in Psychological Bulletin, 37*, 437.

Singer, E. (2006). Nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 637-645.

Sosdian, C. P., & Sharp, L. M. (1980). Nonresponse in mail surveys: Access failure or respondent resistance. *Public Opinion Quarterly, 44*(3), 396-402.

Squire, P. (1988). Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly, 52*(1), 125-133.

Stanton, F. (1939). Notes on the validity of mad questionnaire returns. *Journal of Applied Psychology, 23*, 95-104.

Statistics New Zealand. (2007a). Confidentialised unit record files. *Products and Services.* Retrieved 10 April, 2008, from http://www.stats.govt.nz/products-and-services/microdata-access/confidentialised-unit-record-files/default.htm

Statistics New Zealand. (2007b). December 2006 year (Tables). *External Migration.* Retrieved 12 February, 2008, from http://www.stats.govt.nz/store/2007/02/external-migration-dec06yr-hotp.htm?page=para017Master

Statistics New Zealand. (2007c). Ethnicity. *2006 Census Information About Data.* Retrieved 27 May, 2008, from http://www.stats.govt.nz/census/2006-census-information-about-data/information-by-variable/ethnicity.htm

Statistics New Zealand. (2007d). National summary tables. *2001 Census: National Summary (2001) - Reference Report.* Retrieved 13 July, 2007, from http://www.stats.govt.nz/census/2001-census-data/2001-national-summary/default.htm

Statistics New Zealand. (2007e). Profile of New Zealander responses, Ethnicity question: 2006 Census. Retrieved 27 May, 2008, from http://www.stats.govt.nz/NR/rdonlyres/EA0F8124-619C-47B3-ADB7-CBB28F44AE85/0/ProfileofNewZealanderCensus2006.pdf

Statistics New Zealand. (2007f). QuickStats about culture and identity (Tables). *2006 Census Data.* Retrieved 13 July, 2007, from http://www.stats.govt.nz/census/2006-census-data/quickstats-about-culture-identity/quickstats-about-culture-and-identity.htm?page=para014Master

Statistics New Zealand. (2007g). QuickStats about housing (Tables). *2006 Census Data.* Retrieved 13 July, 2007, from http://www.stats.govt.nz/census/2006-census-data/quickstats-about-housing/quickstats-about-housing-revised2.htm?page=para013Master

Statistics New Zealand. (2007h). QuickStats about incomes (Tables). *2006 Census Data.* Retrieved 13 July, 2007, from http://www.stats.govt.nz/census/2006-census-data/quickstats-about-incomes/quickstats-about-incomes.htm?page=para032Master

Statistics New Zealand. (2007i). QuickStats national highlights (Tables). *2006 Census Data.* Retrieved 13 July, 2007, from http://www.stats.govt.nz/census/2006-census-data/national-highlights/2006-census-quickstats-national-highlights-revised.htm?page=para014Master

Statistics New Zealand. (2007j). Survey of dynamics and motivation for migration in New Zealand: March 2007 quarter. *Survey of Dynamics and Motivation for Migration in New Zealand.* Retrieved 12 February, 2008, from http://www.stats.govt.nz/products-and-services/hot-off-the-press/survey-of-dynamics-and-motivations-for-migration-in-nz/survey-of-dynamics-and-motivation-for-migration-in-nz-mar07qtr-hotp.htm?page=para004Master

Statistics New Zealand. (2007k). Years at usual residence (Tables). *QuickStats About Population Mobility.* Retrieved 12 February, 2008, from http://www.stats.govt.nz/census/2006-census-data/quickstats-about-population-mobility/quickstats-about-population-mobility.htm?page=para018Master

Stinchcombe, A., Jones, C., & Sheatsley, P. (1981). Nonresponse bias for attitude questions. *Public Opinion Quarterly, 45*(3), 359-375.

Suchman, E., & McCandless, B. (1940). Who answers questionnaires? *Journal of Applied Psychology, 24*, 758-769.

The American Association for Public Opinion Research. (2008). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* (5th ed.). Lenexa, Kansas: AAPOR.

Thompson, F. (2007). New Zealand's Electoral Roll database: A PostgreSQL case study. *Paper presented at the PostgreSQL Conference 2007.* Retrieved 01 March, 2008, from http://www.pgcon.org/2007/schedule/attachments/25-Slides

Thompson, S. K., & Seber, G. A. F. (1996). *Adaptive sampling*: Wiley New York.

Toops, H. (1926). The returns from follow-up letters to questionnaires. *Journal of Applied Psychology, 10*, 92-101.

Wiseman, F., & Billington, M. (1984). Comment on a standard definition of response rates. *Journal of Marketing Research, 21*(3), 336-338.

Woodburn, R. L. (1991). Using auxiliary information to investigate nonresponse bias, *Proceedings of the Section on Survey Research Methods* (pp. 278-283). Washington, DC: American Statistical Association.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior - a meta-analysis. *Public Opinion Quarterly, 55*(4), 613-639.

Appendix 1: Information on the Thesis Supplementary CD

A1.1. Workings for Total Noncontact Estimates

A spreadsheet containing data and calculations relating to the total noncontact estimation procedures discussed in section 2.7 (p. 47) can be found in the following file on the supplementary CD attached to the inside back cover of this thesis:
A1-1_Workings_for_Total_Noncontact_Estimates.xls

The file is in Microsoft Excel 1997-2003 format.

A1.2. Copies of ISSP Questionnaires from 2001 to 2006

Scanned images of the *International Social Survey Programme* (ISSP) questionnaires which contributed data to a number of studies in this thesis can be found on the supplementary CD in the following files:

A1.2.1. ISSP 2001: Social Networks in New Zealand

File: A1-2-1_ISSP2001_Questionnaire.pdf

A1.2.2. ISSP 2002: The Roles of Men and Women in Society

File: A1-2-2_ISSP2002_Questionnaire.pdf

A1.2.3. ISSP 2003: Aspects of National Identity

File: A1-2-3_ISSP2003_Questionnaire.pdf

A1.2.4. ISSP 2004: New Zealanders' Attitudes to Citizenship

File: A1-2-4_ISSP2004_Questionnaire.pdf

A1.2.5. ISSP 2005: New Zealanders' Attitudes to Work

File: A1-2-5_ISSP2005_Questionnaire.pdf

A1.2.6. ISSP 2006: The Role of Government

File: A1-2-6_ISSP2006_Questionnaire.pdf

These files are in Adobe PDF format.

A1.3. Copies of Census Forms from 2001 and 2006

Scanned images of the Census forms fielded in 2001 and 2006 can be found on the supplementary CD in the following files:

A1.3.1. Census 2001: Individual Form

File: A1-3-1_CENSUS2001_form_individual.pdf

A1.3.2. Census 2001: Dwelling Form

File: A1-3-2_CENSUS2001_form_dwelling.pdf

A1.3.3. Census 2006: Individual Form

File: A1-3-3_CENSUS2006_form_individual.pdf

A1.3.4. Census 2006: Dwelling Form

File: A1-3-4_CENSUS2006_form_dwelling.pdf

These files are in Adobe PDF format.

A1.4. Walk-through of Calculation Steps in the Proposed NPS Scheme

A spreadsheet containing example data and calculations relating to the proposed noncontact propensity sampling procedure discussed in section 4.5 (p. 107) can be found in the following file on the supplementary CD:

File: A1-4_NPS_key_calculation_walkthrough.xls

The file is in Microsoft Excel 1997-2003 format.

A1.5. Modelling and Simulation SAS Code

The modules included on the supplementary CD were developed to achieve four main objectives:

1. Reformat each contributory dataset (both historic survey datasets and frame datasets) for use in exploratory data analysis (EDA), logistic regression modelling and simulation. This was essentially a data cleaning, summarisation and standardisation exercise to achieve consistency in variable names/values and ensure all required calculated fields (e.g., household composition variables derived from summaries of frame information) were in place before their use in the EDA, logistic modelling and simulation modules. Modules 1-3 cover this step.

2. Undertake survey response and demographic EDA on the historic survey datasets and their associated frame data. This is done to assess whether the different response groups (valids, GNAs, refusers, etc) differ in profile on various frame variables (e.g., frame age, frame gender) and whether those frame variables are in turn correlated with any survey variables (e.g., reported marital status, reported disability, reported income). The output enabled the discussion of relationships between survey response category and bias in survey estimates in chapter 3. Module 4 covers this step.

3. Undertake logistic regression modelling (including EDA of potential predictor variables) using frame variables and associated historic response information. The overall aim was to develop models to predict the likelihood a given 'new' frame record would be a noncontact. These models formed a key input to the Noncontact Propensity Sampling (NPS) mechanism employed in the simulation module. Modules 5 and 6 cover this step.

4. Undertake a bootstrap simulation of survey response and results under SRS (original) and proposed NPS sampling schemes. The NPS mechanism used the models developed earlier to break the simulated population into noncontact propensity strata (using frame information that would be known prior to selection) which were then resampled at different rates. The overall aim was to assess

whether an NPS sampling scheme would generate samples that return survey estimates that are less biased (closer to known population information) and less variable than an SRS scheme. The interaction of the sampling schemes with various post-survey adjustment procedures were also simulated. Module 7 covers this step.

Each of the individual modules of SAS code developed to achieve the above aims is included on the supplementary CD in the following files. Each module contains a description at the beginning of the file outlining the main functions performed. Comments are included throughout to aid readability. Although the file format is '.sas', the files only contain raw text and can safely be opened in any text editor (e.g., MS Word, Notepad, Wordpad).

A1.5.1. Module 1: Create General Resources

A1-5-1_PHD_create_general_resources.sas

A1.5.2. Module 2: Standardise Survey Sets & Assign Selection Weights

A1-5-2_PHD_standardise_surveysets_and_assign_selection_weights.sas

A1.5.3. Module 3: Clean up Survey Response Data

A1-5-3_clean_up_survey_response_data.sas

A1.5.4. Module 4: General Nonresponse EDA

A1-5-4_PHD_nonresponse_general_eda.sas

A1.5.5. Module 5: Pre-modelling Variable Screening EDA

A1-5-5_PHD_ nonresponse_modelling_discrim_eda.sas

A1.5.6. Module 6: Logistic Modelling on Prior Datasets

A1-5-6_PHD_nonresponse_logistic_modelling.sas

A1.5.7. Module 6: Simulation of the NPS Scheme

A1-5-7_PHD_nonresponse_simulation.sas

A1.6. Detailed Result Tables for the Simulation Study

A spreadsheet containing the variable and simulation scenario-level results that contributed to tables 32 to 39 in chapter 5 can be found in the following file:
File: A1-6_simulation_summaries.xls

The spreadsheet contains a number of workbooks and these are labelled according to the tables they contributed to.

The file is in Microsoft Excel 1997-2003 format.

Appendix 2: Sources of Census Figures

A2.1. Notes on Census Data Sources and Calculations

The census parameters against which survey estimates are compared in chapters 3 and 5 were all sourced from data tables publicly available from *Statistics New Zealand*'s website (www.stats.govt.nz). Table 40 presents citations for each of the variables employed. Full bibliographic details can be found in the references (p. 157).

For three of the variables, it was not possible to limit the population figures to those aged 20 years or older (see the 'Base' column in the table), because the official statistics were not available with age breakdowns. Furthermore, the electoral roll, and each ISSP survey sample taken from it, covers those 18 years or older. Hence, all variables except those relating to household size differ by 2 to 3 years in their comparative bases between the census and survey sets.

Table 40: Sources for individual census parameters

		Source of Census Data	
Variable	**Base**	2001	2006
% Male	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% 20-34 Years old	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% 65+ Years old	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% Maori	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007f)
% Marital: Single	15+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% Bach/PG Qual	15+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% Income <$20k	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007h)
% Income > $50k	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007h)
% Not Religious	20+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007f)
% Empl. Fulltime	15+	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007i)
% 1 Person HH	All	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007g)
% 5+ Person HH	All	(Statistics New Zealand, 2007d)	(Statistics New Zealand, 2007g)

The discrepancy in bases means that, in some cases, survey estimates that appear to match census figures on average would actually slightly under or overestimate the population parameter if the bases were equal. In particular, this is likely to be the case for ethnicity (% Maori) and religiosity (% Not Religious), since they covary with age (Maori have a lower life expectancy and their population is skewed toward the young).

For instance, for the ‘% Not Religious’ variable, a shift in base from 20+ to 15+ causes the population percentage to change from 28% to 29% for 2001, and 33% to 35% for 2006. All the survey estimates reported in Table 22 are below the latter figure, despite the fact that the surveys for 2005 and 2006 effectively oversampled younger people.

Turning to ethnicity, if the base for ‘% Maori’ is shifted from 20+ to 15+, the proportion rises from 11% to 12% in both 2001 and 2006. All but one of the survey estimates reported in Table 22 (p. 71) are below 12%, and the one that is not comes from a sample that deliberately overrepresented Maori. Indeed, there are also other reasons why measurement error may contribute to a smaller difference between the survey and census data than really exists for this variable. For example, classification methods for ethnicity changed between 2001 and 2006 (see Statistics New Zealand, 2007c), such that in 2006 the class ‘New Zealander’ was explicitly reported where, in the past, it was subsumed within the ‘European’ category. This occurred at a time when there was public discussion about the term ‘New Zealander’ in the months leading up to the 2006 field period. In 2001, 2.4% of people identified with the write-in ‘New Zealander’ ethnicity category, whereas 11.1% did so in 2006. To the extent that this category was used by some people who would normally have reported only Maori ethnicity (Statistics New Zealand, 2007e), the reported proportion of people with Maori ethnicity in the population for 2006 may have been reduced.

Furthermore, while the census employs a categorisation schema that counts people under multiple categories if they signal multiple ethnicities, the calculations for the ISSP survey were developed under single ethnicity prioritisation scheme (i.e., where multiple ethnicities were signalled, only one was used and Maori was given priority in selection). Had a multiple ethnicity scheme been employed for the ISSP, the gap between the census and ISSP figures is likely to have been larger. It would be possible to recode the ethnicity variable in the ISSP to make it comparable to the census classification mechanism. However, since it is unlikely to change the conclusions drawn in the studies in the thesis, this has not been done.

Other variables that are likely to be subject to measurement error are ‘Highest Qualification’ and ‘Marital Status’. Specifically, changes in the educational system in New Zealand between 2001 and 2006 meant that the classification scheme for highest educational attainment changed between the two census instances (see Appendix section A1.3, p. 169 for copies of the forms). Moreover, while qualifications beyond high school were required to be written in on the census form (and were subsequently coded), the ISSP surveys had coded options for post-high school qualifications. Thus, differences in question wording may have led to discrepancies between the census and ISSP results reported.

For ‘Marital Status’, changes in legislation regarding civil unions also led to a change in census question formats between 2001 and 2006. Furthermore, there were some differences between the ISSP question and the 2001 census format (the number of categories were the same, but the wording was simplified and they were presented in a different order in the ISSP). Thus, variations in question wording may have led to discrepancies between the census and ISSP results reported.

Of note is that, although measurement error is likely to be an issue in the ‘Highest Qualification’ and ‘Marital Status’ questions, the direction of bias was at least consistent across all of the surveys examined. Furthermore, the NPS scheme detailed in chapters 4 and 5 did have some success in moving the ‘Marital Status’ ISSP estimates closer to the census parameters (results were mixed for the ‘Highest Qualification’ variable). Hence, at least some of the bias in ‘Martial Status’ is likely to have been due to noncontact nonresponse.

Appendix 3: Logistic Regression Models

A3.1. Detailed Logistic Regression Model Specifications

The top-level specifications for each noncontact propensity model developed using historic survey datasets are presented below. The '*Std Error (Adj)*' figures under each '*Analysis of Maximum Likelihood Estimates*' section relate to parameter estimate standard errors calculated taking into account design complexities (i.e., stratification). These were generated using the SAS SURVEYLOGISTIC procedure.

Note: Complete specifications (including odds ratio point estimates) are contained in the following spreadsheet file on the thesis supplementary CD:
A3-1_ModelFINAL_GNA_specs.xls.

A3.1.1. 2003 Model: Built on 2001 and 2002 data

Number of Observations Used	4233
Sum of Weights Used	4232.821
Probability modeled is	GNA_flag=1

Model Fit Statistics

Max-rescaled R-Square	0.0915	Somers' D	0.356
Percent Concordant	67.2	Gamma	0.361
Percent Discordant	31.6	Tau-a	0.08
Percent Tied	1.2	c	0.678
Pairs	2022516		

Testing Global Null Hypothesis: BETA=0

Test	*Chi-Square*	*DF*	*Pr > ChiSq*
Likelihood Ratio	213.845	32	<.0001
Score	232.4155	32	<.0001
Wald	201.9792	32	<.0001

Analysis of Effects

Effect	*Wald Chi-Square*	*DF*	*Pr > ChiSq*
HH_prop_maoridesc	9.4976	1	0.0021
frame_agegrp	46.0576	12	<.0001
HH_electorsgrp	10.4881	5	0.0625
frame_dwellsplit_fla	6.7273	1	0.0095
frame_postal_adrs_ty	46.6313	5	<.0001
HH_surnamesgrp	33.3996	4	<.0001
HH_Prop_malegrp	6.8978	4	0.1414

Analysis of Maximum Likelihood Estimates

Parameter		*Estimate*	*DF*	*Std Error*	*Std Error (Adj)*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept		-1.0858	1	0.1819	0.1776	37.3874	<.0001
HH_prop_maoridesc		0.465	1	0.1533	0.1509	9.4976	0.0021
frame_agegrp	18-24	0.5909	1	0.1378	0.1387	18.1474	<.0001
frame_agegrp	25-29	0.6231	1	0.1349	0.1357	21.0965	<.0001
frame_agegrp	30-34	0.364	1	0.1394	0.1409	6.6775	0.0098
frame_agegrp	35-39	0.071	1	0.1401	0.143	0.2463	0.6197
frame_agegrp	40-44	-0.0295	1	0.1484	0.1493	0.039	0.8434
frame_agegrp	45-49	-0.251	1	0.1659	0.1663	2.2777	0.1312
frame_agegrp	50-54	-0.0936	1	0.1676	0.1679	0.3106	0.5773
frame_agegrp	55-59	-0.3237	1	0.1969	0.1977	2.6815	0.1015
frame_agegrp	60-64	-0.3408	1	0.2076	0.2101	2.6305	0.1048
frame_agegrp	65-69	-0.00427	1	0.2151	0.2135	0.0004	0.984
frame_agegrp	70-74	-0.2596	1	0.2435	0.2459	1.1148	0.291
frame_agegrp	75-79	-0.2939	1	0.2665	0.2678	1.2044	0.2724
HH_electorsgrp	1	0.6383	1	0.2246	0.218	8.5715	0.0034
HH_electorsgrp	2	0.1001	1	0.1791	0.1795	0.311	0.5771
HH_electorsgrp	3	-0.0417	1	0.1699	0.1654	0.0636	0.8009
HH_electorsgrp	4	-0.2803	1	0.1725	0.175	2.5658	0.1092
HH_electorsgrp	5	-0.1722	1	0.2432	0.2373	0.5266	0.4681
frame_dwellsplit_fla	0	-0.152	1	0.0585	0.0586	6.7273	0.0095
frame_postal_adrs_ty	-	-0.6082	1	0.1634	0.1583	14.7673	0.0001
frame_postal_adrs_ty	B	0.05	1	0.6069	0.5882	0.0072	0.9323
frame_postal_adrs_ty	C	-0.00736	1	0.4245	0.4086	0.0003	0.9856
frame_postal_adrs_ty	N	0.4147	1	0.2666	0.2614	2.5175	0.1126
frame_postal_adrs_ty	P	0.0452	1	0.2057	0.2028	0.0497	0.8237
HH_surnamesgrp	1	-0.777	1	0.1567	0.1538	25.512	<.0001
HH_surnamesgrp	2	-0.4584	1	0.1556	0.1563	8.5975	0.0034
HH_surnamesgrp	3	0.326	1	0.1819	0.1842	3.1321	0.0768
HH_surnamesgrp	4	0.3262	1	0.2753	0.2834	1.3248	0.2497
HH_Prop_malegrp	1 - 0%	-0.0314	1	0.1485	0.1464	0.0459	0.8303
HH_Prop_malegrp	2 - 1-44%	-0.0597	1	0.1705	0.1638	0.1326	0.7157
HH_Prop_malegrp	3 - 45-55%	0.0212	1	0.1447	0.1459	0.0211	0.8846
HH_Prop_malegrp	4 - 56-99%	-0.2589	1	0.1768	0.1736	2.2255	0.1358

A3.1.2. 2004 Model: Built on 2002 and 2003 data

Number of Observations Used	4202
Sum of Weights Used	4201.283
Probability modeled is	GNA_flag=1

Model Fit Statistics

Max-rescaled R-Square	0.1003	Somers' D	0.388
Percent Concordant	68.8	Gamma	0.392
Percent Discordant	30.1	Tau-a	0.082
Percent Tied	1.1	c	0.694
Pairs	1876552		

Testing Global Null Hypothesis: BETA=0

Test	*Chi-Square*	*DF*	*Pr > ChiSq*
Likelihood Ratio	224.9473	31	<.0001
Score	235.1146	31	<.0001
Wald	198.4102	31	<.0001

Analysis of Effects

Effect	*Wald Chi-Square*	*DF*	*Pr > ChiSq*
frame_age	53.6663	12	<.0001
HH_surnamesgrp	34.4438	4	<.0001
HH_electorsgrp	11.9076	5	0.0361
HH_Prop_malegrp	8.3543	4	0.0794
frame_dwellsplit_fla	5.5713	1	0.0183
frame_postal_adrs_ty	76.2245	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		*Estimate*	*DF*	*Std Error*	*Std Error (Adj)*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept		-1.0614	1	0.1828	0.1781	35.516	<.0001
frame_agegrp	18-24	0.7584	1	0.1422	0.1455	27.1682	<.0001
frame_agegrp	25-29	0.5942	1	0.1419	0.1434	17.1661	<.0001
frame_agegrp	30-34	0.3862	1	0.1442	0.1469	6.9169	0.0085
frame_agegrp	35-39	0.1842	1	0.1434	0.1447	1.6199	0.2031
frame_agegrp	40-44	-0.0814	1	0.1529	0.1557	0.2731	0.6012
frame_agegrp	45-49	-0.5325	1	0.1894	0.1913	7.7508	0.0054
frame_agegrp	50-54	0.0629	1	0.1657	0.1663	0.1428	0.7055
frame_agegrp	55-59	-0.0674	1	0.1904	0.1881	0.1284	0.7201
frame_agegrp	60-64	-0.0787	1	0.204	0.2072	0.1445	0.7039
frame_agegrp	65-69	-0.211	1	0.232	0.2332	0.8182	0.3657
frame_agegrp	70-74	-0.401	1	0.2559	0.2591	2.3952	0.1217
frame_agegrp	75-79	-0.3225	1	0.2838	0.2848	1.2822	0.2575
HH_surnamesgrp	1	-0.7926	1	0.156	0.1581	25.1199	<.0001
HH_surnamesgrp	2	-0.251	1	0.1509	0.1546	2.638	0.1043
HH_surnamesgrp	3	0.2051	1	0.1828	0.1892	1.1753	0.2783
HH_surnamesgrp	4	0.4252	1	0.2573	0.2611	2.6519	0.1034
HH_electorsgrp	1	0.5294	1	0.2272	0.2205	5.7652	0.0163
HH_electorsgrp	2	-0.0873	1	0.1725	0.1659	0.2767	0.5988
HH_electorsgrp	3	0.0396	1	0.1764	0.1765	0.0503	0.8225

HH_electorsgrp	4	-0.0398	1	0.1644	0.1642	0.0588	0.8085
HH_electorsgrp	5	0.2931	1	0.2294	0.2165	1.8316	0.1759
HH_Prop_malegrp	1 - 0%	0.1764	1	0.1488	0.142	1.5435	0.2141
HH_Prop_malegrp	2 - 1-44%	-0.3756	1	0.1776	0.1672	5.0462	0.0247
HH_Prop_malegrp	3 - 45-55%	0.1211	1	0.1438	0.1464	0.684	0.4082
HH_Prop_malegrp	4 - 56-99%	-0.3423	1	0.1764	0.1704	4.0326	0.0446
frame_dwellsplit_fla	0	-0.1447	1	0.0606	0.0613	5.5713	0.0183
frame_postal_adrs_ty	-	-0.7716	1	0.1641	0.1589	23.5947	<.0001
frame_postal_adrs_ty	B	0.06	1	0.6033	0.5722	0.011	0.9165
frame_postal_adrs_ty	C	0.2851	1	0.4167	0.4063	0.4926	0.4828
frame_postal_adrs_ty	N	0.0145	1	0.2866	0.2898	0.0025	0.96
frame_postal_adrs_ty	P	0.1603	1	0.2044	0.2012	0.635	0.4255

A3.1.3. 2005 Model: Built on 2003 and 2004 data

Number of Observations Used	4607
Sum of Weights Used	4607.273
Probability modeled is	GNA_flag=1

Model Fit Statistics

Max-rescaled R-Square	0.1283	Somers' D	0.452
Percent Concordant	72	Gamma	0.457
Percent Discordant	26.8	Tau-a	0.069
Percent Tied	1.1	c	0.726
Pairs	1625470		

Testing Global Null Hypothesis: BETA=0

Test	*Chi-Square*	*DF*	*Pr > ChiSq*
Likelihood Ratio	266.1578	34	<.0001
Score	321.7822	34	<.0001
Wald	258.6921	34	<.0001

Analysis of Effects

Effect	*Wald Chi-Square*	*DF*	*Pr > ChiSq*
frame_agegrp	34.8264	12	0.0005
frame_postal_adrs_ty	14.347	6	0.026
frame_dwellpost_diff	27.2495	1	<.0001
frame_dwellsplit_fla	5.3452	1	0.0208
HH_electorsgrp	9.5521	5	0.089
HH_Prop_malegrp	11.2014	4	0.0244
HH_surnamesgrp	28.8049	4	<.0001
frame_rolltype_char	3.9159	1	0.0478

Analysis of Maximum Likelihood Estimates

Parameter		*Estimate*	*DF*	*Std Error*	*Std Error (Adj)*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept		-1.2663	1	0.2076	0.2008	39.7677	<.0001
frame_agegrp	18-24	0.6091	1	0.1622	0.1668	13.3349	0.0003
frame_agegrp	25-29	0.6178	1	0.1603	0.1626	14.4325	0.0001
frame_agegrp	30-34	0.4048	1	0.1617	0.1608	6.3382	0.0118
frame_agegrp	35-39	0.2051	1	0.165	0.1675	1.4995	0.2207
frame_agegrp	40-44	0.0303	1	0.1692	0.177	0.0293	0.8642
frame_agegrp	45-49	-0.3282	1	0.2024	0.2077	2.4978	0.114
frame_agegrp	50-54	0.0275	1	0.196	0.202	0.0185	0.8918
frame_agegrp	55-59	-0.2345	1	0.2276	0.2209	1.1273	0.2884
frame_agegrp	60-64	0.1043	1	0.2255	0.2339	0.1988	0.6557
frame_agegrp	65-69	-0.2775	1	0.282	0.2923	0.9017	0.3423
frame_agegrp	70-74	-0.3847	1	0.3061	0.3192	1.4524	0.2281
frame_agegrp	75-79	-0.6027	1	0.3866	0.406	2.2036	0.1377
frame_postal_adrs_ty	-	-0.3593	1	0.2447	0.234	2.3576	0.1247
frame_postal_adrs_ty	B	0.0561	1	0.5569	0.5547	0.0102	0.9194
frame_postal_adrs_ty	C	0.0808	1	0.4623	0.4991	0.0262	0.8714
frame_postal_adrs_ty	F	0.953	1	0.3306	0.3004	10.0635	0.0015
frame_postal_adrs_ty	N	-0.297	1	0.3972	0.4312	0.4744	0.491

frame_postal_adrs_ty	P	-0.2487	1	0.2047	0.2097	1.4067	0.2356
frame_dwellpost_diff	0	-0.5877	1	0.1206	0.1126	27.2495	<.0001
frame_dwellsplit_fla	0	-0.16	1	0.0678	0.0692	5.3452	0.0208
HH_electorsgrp	1	0.3164	1	0.2395	0.2476	1.6329	0.2013
HH_electorsgrp	2	-0.3604	1	0.1729	0.1717	4.4077	0.0358
HH_electorsgrp	3	0.0677	1	0.1719	0.175	0.1494	0.6991
HH_electorsgrp	4	0.0708	1	0.1676	0.1761	0.1615	0.6878
HH_electorsgrp	5	0.1142	1	0.2472	0.2494	0.2098	0.647
HH_Prop_malegrp	1 - 0%	0.1013	1	0.1644	0.1651	0.3765	0.5395
HH_Prop_malegrp	2 - 1-44%	-0.5463	1	0.1804	0.1862	8.6109	0.0033
HH_Prop_malegrp	3 - 45-55%	0.0805	1	0.1531	0.1606	0.2514	0.6161
HH_Prop_malegrp	4 - 56-99%	-0.0953	1	0.171	0.1759	0.2935	0.588
HH_surnamesgrp	1	-0.7053	1	0.1576	0.1578	19.9874	<.0001
HH_surnamesgrp	2	-0.0197	1	0.1469	0.1455	0.0184	0.8921
HH_surnamesgrp	3	0.2654	1	0.1789	0.1793	2.1917	0.1388
HH_surnamesgrp	4	0.0964	1	0.266	0.282	0.1168	0.7325
frame_rolltype_char	G	-0.1605	1	0.0938	0.0811	3.9159	0.0478

A3.1.4. 2006 Model: Built on 2004 and 2005 data

Number of Observations Used	4857
Sum of Weights Used	4857.71
Probability modeled is	GNA_flag=1

Model Fit Statistics

Max-rescaled R-Square	0.1117	Somers' D	0.423
Percent Concordant	70.5	Gamma	0.429
Percent Discordant	28.2	Tau-a	0.057
Percent Tied	1.3	c	0.712
Pairs	1577450		

Testing Global Null Hypothesis: BETA=0

Test	*Chi-Square*	*DF*	*Pr > ChiSq*
Likelihood Ratio	222.4943	40	<.0001
Score	277.5111	40	<.0001
Wald	231.7819	40	<.0001

Analysis of Effects

Effect	*Wald Chi-Square*	*DF*	*Pr > ChiSq*
frame_agegrp	24.9062	12	0.0153
frame_postal_adrs_ty	12.3721	6	0.0542
frame_dwellpost_diff	16.2716	1	<.0001
frame_dwellsplit_fla	4.5031	1	0.0338
frame_employstatus	21.1258	6	0.0017
HH_electorsgrp	10.1701	5	0.0706
HH_Prop_malegrp	13.9249	4	0.0075
HH_surnamesgrp	26.8574	4	<.0001
frame_rolltype_char	6.1781	1	0.0129

Analysis of Maximum Likelihood Estimates

Parameter		*Estimate*	*DF*	*Std Error*	*Std Error (Adj)*	*Wald Chi-Square*	*Pr > ChiSq*
Intercept		-1.5036	1	0.2611	0.2706	30.8836	<.0001
frame_agegrp	18-24	0.4039	1	0.1966	0.2036	3.9363	0.0473
frame_agegrp	25-29	0.2262	1	0.1838	0.1763	1.6459	0.1995
frame_agegrp	30-34	0.2666	1	0.1674	0.1596	2.7918	0.0947
frame_agegrp	35-39	-0.1389	1	0.1877	0.1943	0.511	0.4747
frame_agegrp	40-44	0.0987	1	0.1722	0.1838	0.2883	0.5913
frame_agegrp	45-49	-0.2852	1	0.2032	0.2165	1.735	0.1878
frame_agegrp	50-54	-0.4145	1	0.2356	0.2514	2.7178	0.0992
frame_agegrp	55-59	-0.2954	1	0.2364	0.2296	1.6561	0.1981
frame_agegrp	60-64	-0.6287	1	0.3018	0.307	4.1954	0.0405
frame_agegrp	65-69	-0.3017	1	0.2906	0.2902	1.0812	0.2984
frame_agegrp	70-74	0.1968	1	0.298	0.2905	0.4588	0.4982
frame_agegrp	75-79	0.6461	1	0.2882	0.2736	5.5754	0.0182
frame_postal_adrs_ty	-	-0.0332	1	0.2752	0.2807	0.0139	0.906
frame_postal_adrs_ty	B	-0.2407	1	0.6159	0.6986	0.1188	0.7304
frame_postal_adrs_ty	C	-0.0582	1	0.6121	0.597	0.0095	0.9223
frame_postal_adrs_ty	F	0.8862	1	0.3241	0.32	7.6673	0.0056
frame_postal_adrs_ty	N	-0.3791	1	0.9585	1.053	0.1296	0.7188

frame_postal_adrs_ty	P	-0.1692	1	0.2782	0.2947	0.3295	0.566
frame_dwellpost_diff	0	-0.4904	1	0.1247	0.1216	16.2716	<.0001
frame_dwellsplit_fla	0	-0.1451	1	0.0685	0.0684	4.5031	0.0338
frame_employstatus	BENEFIT	-0.472	1	0.3888	0.3806	1.5377	0.215
frame_employstatus	EMPLOYED	-0.0475	1	0.1256	0.127	0.1396	0.7086
frame_employstatus	HOMEMKR	0.2993	1	0.1707	0.1717	3.0375	0.0814
frame_employstatus	NOTSTATED	0.2344	1	0.2137	0.2151	1.1876	0.2758
frame_employstatus	RETIRED	-0.5793	1	0.2674	0.2539	5.2067	0.0225
frame_employstatus	STUDENT	-0.2331	1	0.2093	0.2053	1.2895	0.2561
HH_electorsgrp	1	-0.00284	1	0.2522	0.2559	0.0001	0.9911
HH_electorsgrp	2	-0.3271	1	0.1827	0.1854	3.1134	0.0776
HH_electorsgrp	3	0.4782	1	0.168	0.165	8.3941	0.0038
HH_electorsgrp	4	-0.0879	1	0.1827	0.1873	0.2203	0.6388
HH_electorsgrp	5	0.068	1	0.2474	0.2521	0.0728	0.7874
HH_Prop_malegrp	1 - 0%	-0.1133	1	0.1739	0.167	0.4606	0.4973
HH_Prop_malegrp	2 - 1-44%	-0.3812	1	0.1679	0.1719	4.9169	0.0266
HH_Prop_malegrp	3 - 45-55%	0.0561	1	0.1618	0.1659	0.1144	0.7352
HH_Prop_malegrp	4 - 56-99%	-0.1743	1	0.1671	0.1644	1.1232	0.2892
HH_surnamesgrp	1	-0.7691	1	0.162	0.1619	22.5634	<.0001
HH_surnamesgrp	2	-0.1536	1	0.1482	0.1482	1.0743	0.3
HH_surnamesgrp	3	0.0938	1	0.1737	0.1755	0.2856	0.593
HH_surnamesgrp	4	0.0221	1	0.2754	0.2778	0.0063	0.9367
frame_rolltype_char	G	-0.2342	1	0.0928	0.0942	6.1781	0.0129