An Investigation into the Feasibility of an Online National Probability Panel Study in New Zealand

Lara M. Greaves

Table of Contents

Executive Summary

This report was conducted to investigate the feasibility of setting up a national probability sample online survey panel for academic research purposes. The report reviews existing panels, both internationally and nationally, and reviews practical considerations for setting up a national probability panel study, including the pilot study, sampling, response rates, attrition, size, and accessibility issues. The report concludes that setting up a national probability sample online survey panel is feasible, with a comprehensive pilot study and careful consideration of the areas reviewed below.

Section 1 starts with a summary of what a panel study is before describing the reasons why a probability panel is superior to a non-probability panel. In short, studies have shown that non-probability panels may appear representative in terms of demographics, however, they are typically inaccurate when it comes to social or psychological indicators. This is largely due to biases in who chooses to opt in to an online panel survey. This sections also provides a review of the existing and successful national probability panel studies in Australia, North America, and Europe. Additionally, this section describes the wide variety of non-probability panels currently in operation in New Zealand in the market research sector.

Section 2 describes the practical considerations involved in setting up a national probability online panel. Firstly, it is widely recommended that researchers conduct a pilot study, as judgement errors later in the process can prove to be costly. The report reviews past pilot studies internationally and common factors that are tested in pilot studies.

Secondly, past research has used a variety of sampling methods. However, given the New Zealand context mail (electoral roll) and phone (random digit dialing) are explored as the most viable options in the report. Ultimately, both have non-coverage issues, however, electoral roll based mail sampling is recommended due to cost and the existing knowledge of the research team.

Thirdly, the report reviews the sample size and response rates of past surveys. In summary, around 3,000 participants for the main study would be desirable. Such details would be figured out through the comprehensive pilot study.

Fourth, the report describes the past research on attrition and the different types of panel attrition that past studies have found. The techniques used to mitigate attrition and fatigue are discussed. It is often cheaper and more efficient to maintain panelists rather than recruit more, however, the report provides an overview of how researchers might ‘top up’ the sample. If stratified sampling is used from the electoral roll, this process would likely be easier and more cost effective than procedures other panels have had to use due to their methods.

Fifth, the report briefly covers the other, smaller, perhaps ‘miscellaneous’ factors involved in survey administration. This includes ethics, websites and software, frequency and length of surveys, costs and costs for clients.

Sixth, incentives are reviewed as a method to help response rates and retention. The difference between contingent (promised) and non-contingent (prepaid) incentives are discussed. It is recommended that the panel incentivize response. Furthermore, due to cost differences, contingent incentives seem to be the best option.

Finally, the report considers accessibility issues that may bias the sample in some way. Namely, the non-participation of those who do not speak English and those who either do not have internet access or do not like using the internet (and are therefore unlikely to join an online panel). While there are relatively few non-English speakers there is a considerable offline population. Thus, ways to include the offline population are discussed. In summary, the report recommends contacting offline participants via post (more likely, given the research team's area of expertise) or phone.

1. Review of Existing Studies

1.1 What is a Panel Study?

A panel refers to a group of respondents that have agreed to be contacted to participate in multiple surveys over time. This type of study has many advantages over traditional one-off studies. Firstly, it allows researchers to track changes in individual's attitudes and behaviours over time if they wish to. Secondly, because panelists fill out a number of surveys it allows researchers to collect a considerable amount of information on the participants. These participants are matched across surveys, which opens up a broad range of research questions, while minimizing participant fatigue (a common problem with surveys, where data quality suffers). Thirdly, as panelists join the study and trust is built between the participants and the researcher, panelists may allow the researcher to access external data, for example, in our case, the Integrated Data Infrastructure (IDI). Furthermore, this kind of research is cost effective, since one does not have to draw a whole new sample and wear the costs associated with nonresponse for each survey (as once participants are recruited to the panel response rates tend to be high).

1.2 Why do a National Probability Sample Panel Study over a Non-probability one?

Online research has become popular in recent years due to its low cost, quick turnaround times, and the increase of internet coverage across many nations. Over the past two decades, this has meant that many market research firms have created panels using non-probability (convenience) samples that pay participants for completing surveys (see Callegaro, Villar, Yeager, & Krosnick, 2014; Postoaca, 2006). Participants in non-probability panel surveys opt-in to these panels at the end of an online, phone or postal survey, or more commonly, through clicking on online advertising (Callegaro et al., 2014). However, this sample is limited in that participants always (a) already have a computer and the internet and (b) are comfortable enough with using the internet to join the panel.

Firms convenience sample because there is no comprehensive sample frame for the internet (i.e., no list of every person on the internet and their contact details or email). Thus, an issue for this research is how representative the results of these panels are, basically, whether the results from these online panels can be generalized to the population (external validity; Callegaro et al., 2014). Research finds that these non-probability panel members tend to be more likely to be unemployed and are far heavier internet users than the population (e.g., Baim et al., 2009; Williams, 2012). There is also evidence to suggest that more than 30% of online panelists for non-probability samples belong to more than one online panel (Vonk et al., 2006).

There are two papers of particular relevance to our project that empirically illustrate the limitations of non-probability panel studies. Pennay, Neiger, and Lavrakas (2016) from the Social Research Centre/Life in Australia panel conducted the Australian online panels benchmarking study. The Australian researchers were inspired by the work of Yeager et al., (2011) on American panels. Many non-probability panels perform fairly well in terms of demographic variables as they quota sample from their wider panel, i.e., only allow the proportion of middle-aged women of European descent in the survey to mirror the proportion in the population, then cut off access to the survey for this group. However, when compared with probability panels and benchmark values (from official surveys) on indicators of wellbeing like life satisfaction, non-specific psychological distress, general health status,

private health insurance coverage, smoking and drinking behaviours, these non-probability panels perform poorly.

That being said, international researchers have successfully created probability panel samples. Indeed, research has shown that these panels produce more accurate results (Chang & Krosnick, 2009; Scherpenzeel & Bethlehem, 2010; Yeager et al., 2011). These samples are typically recruited from Random Digit Dialing (RDD) phone numbers or from Address Based Sampling (ABS); mailing address databases of all of the postal addresses in the country. That way everyone has an equal chance of being selected to participate in the panel and the characteristics of the broader population that the sample were drawn from can be known, i.e., through the census, and reliable weights can be created to ensure reliability.

1.3 What is Currently in Existence?

A number of national probability online panel samples exist worldwide, many of which are reviewed below. However, please note that the amount of information available on the methods behind each panel varies widely, depending on who funds the research (Callegaro et al., 2014). The following section provides information on similar international panels/studies of note.

1.3.1 Australia

The **Life in Australia** (N = ~3,500; http://www.srcentre.com.au/our-research/panel) study is run by the Social Research Centre, Australian National University. The Life in Australia study randomly samples Australian residents aged over 18 through both landline (40%) and mobile phones (60%; dual frame, RDD). The initial sample was taken in September/October 2016, the first survey was scheduled for November 2016 and the first (media) release of the result was February 2017 on political attitudes (*n* = 2,600).

The **Australian Health and Social Science Survey** (https://www.cqu.edu.au/industry-and-partnerships/services/population-research-laboratory/research/ahhs-project) project is a dual frame representative panel study of Australian adults funded by the Institute for Health and Social Science Research at Central Queensland University, Rockhampton, Australia. They regularly survey participants on subjects relating to health, wellbeing, and "contemporary issues". Their sample was recruited using Computer Assisted Telephone Interviewing (CATI) and then provided a website link.

1.3.2 North America

The **GfK Knowledge Panel** (N = ~55,000; http://www.gfk.com/products-a-z/us/knowledgepanel-united-states/) is one of the older panels, starting in 1999. Originally they used Random Digit Dialing (RDD) but they changed approach 2009 "to improve the representativeness of the panel". They use an ABS method where they utilize the delivery sequence file from the US postal service (a computerized file that contains all valid postal addresses serviced by the US postal service). The main selling point being that phone status does not matter, thus they can access young adults and other minority groups which may be hard to reach by phone. They target 4 strata: Hispanic households with at least one 18-24 year old, remaining Hispanic households, remaining households with at least one 18-24 year old, then all remaining households. Hundreds of academic publications have used their data, largely in health research. However, in their technical documents they use a lot of buzzwords and generally lack transparency in comparison to the other panels. The main strength of this

panel is their Hispanic sample, since there is also a Spanish speaking panel called KnowledgePanel Latino to get representative Hispanic samples. The research team draws a subset for representative surveys and also allows researchers to target subpopulations. For example, journal articles on research that investigated smoking in LGBT+ populations.

RAND American Life Panel (https://alpdata.rand.org/) started in 2003 with a five-year grant from the National Institute on Aging to study methodological issues in interviewing older populations (initially N = 800). The panel has expanded considerably in its aims and sample. They now have over 6,000 participants and has run over 460 surveys with 35 different research groups including commercial, academic, and government clients. They have largely recruited participants from other probability-based surveys using multiple modes and both address and RDD samples. They allow participants to join from non-probabilistic method surveys, but they keep everyone separate through distinct sample frame codes. Some of these participants are additional household members, around 17% of the households surveyed contain more than one household member. They have the capacity to run a wide range of projects, including biometric and cellphone based data collection. For example, they ran the 6-wave 2016 RAND Presidential Election Panel Survey (PEPS; around 3,000 US citizens) out of a subsample of the study. PEPS was longitudinal and allowed political scientists to track candidate preferences over the election cycle.

Pew American Trends Panel (N = ~5,000; http://www.pewresearch.org/methodology/u-s-survey-research/american-trends-panel/) started in 2014 with two RDD dual frame surveys conducted in English and Spanish, where participants were asked if they wanted to join the panel at the end. One of the surveys was the 2014 Political Polarization and Typology Survey (5,338 recruited to the panel; a little over half of participants), the second was the 2015 Survey on Government (2,976 recruited to the panel; a little under half of the participants). However, because they recruited participants from surveys on politics that were up to 20 minutes long, their sample is biased. They have found that their sample is more civically minded than other studies.

The gold standard seems to be the NORC (National Opinion Research Center) University of Chicago **AmeriSpeak panel** (http://amerispeak.norc.org/). Originally, their sample was around 15,000, but they expanded to 20,000 households in 2016 and have a goal of 60,000 members. NORC has its own special sample frame which they say covers 97% of American households. This is a two stage process. Additionally, because the NORC sample frame doesn't cover all states, they use the US Postal service list, 0.9% of their sample came from this. They then use weights when combining samples. They also strategically stratify to oversample young (18-30) African American, Hispanic, and Asian households (based on "consumer vendor data"). Potential participants received an oversized postcard, then a recruitment package containing a cover letter, a summary of their privacy policy, frequently asked questions, and a brochure), two follow up postcards and a phone call reminder for those matched to phone. The hard-to-reach populations also receive a second recruitment package and a face-to-face visit to their door. This study is also of note as it has a component called Time-sharing Experiments for the Social Sciences (TESS) which has been funded through the National Science Foundation. TESS allows researchers to approach the study with questions and, if selected, include them free of charge. AmeriSpeak claims to be the top panel in the United States and has the highest AAPOR (American Association of Public Opinion Research) response rate of any panel.

The **Gallup Panel** (http://www.gallup.com/services/172364/gallup-panel.aspx) was started in 2004, and currently claims to have 100,000 participants reachable by phone, about 80,000 of whom can also be emailed. Participants were selected through dual frame RDD and ABS. For the RDD sample, participants are first asked a short series of questions on current events and presidential approval they are then asked if they would like to join the panel. Potential participants are then mailed a welcome pack with a pamphlet, business card of the panel relationship manager, and a survey booklet with a return postage paid envelope. The welcome pack also invites other family members over age 13 to join the panel. An interesting thing about the Gallup Panel is their use of participant-preferred and mixed modes; they even have an app.

1.3.3 Europe

There are four well known national probability internet panels in Europe, all of them focus on academic and social research, make their data publically available through various data archives, and collaborate with one another.

German Internet Panel (GIP; http://reforms.uni-mannheim.de/internet_panel/home/) was started in 2012 (N = ~1,500) at the University of Mannheim and is funded by the German Research Foundation. Their sampling frame was from an "area probability sample with separate listing of households". Individuals (16-75 years of age) were recruited following a face-to-face interview, after sampling that was clustered in households, and stratified by region and urbanicity, years of age. Participants are surveyed bimonthly on their attitudes towards different policies.

GESIS (Leibniz Institute for the Social Sciences http://www.gesis.org/en/institute/) run a representative panel of individuals in Germany (N = ~4,900 aged 18-70) from 2013. This group also runs the International Social Survey Programme. This research is funded by their Government and data collection is free. Their study includes a unit called the Longitudinal Core Study that measures characteristics that are frequently asked for by researchers such as demographics, personality, values, political behaviour and attitudes, well-being and quality of life. In terms of sampling, their pilot study was 50/50 landline/cellphone RDD. However, their main sample was recruited in person with a Computer Aided Personal Interview (CAPI) from municipal population registers (stratified by region and urbanicity).

LISS (Longitudinal Internet Studies for the Social Sciences; https://www.lissdata.nl/lissdata/) was started in the Netherlands in 2007 at Tilburg University (N = ~8,000). The study is a household design, of those aged 16+. Participants were contacted by many modes including mail, phone, and face-to-face. Fortunately, the Netherlands has a national population register that the researchers were able to utilize. They first sent a letter and brochure to households, then they either called the household or if no number was available, visited the household (Scherpenzeel & Das, 2010; Scherpenzeel, 2011). Initially all household members were asked to participate with one household member being the main informant. Similar to the GESIS panel, the LISS has a 'Core Study' that tracks changes across health, politics and values, religion and ethnicity, social integration and leisure, family and household, work and schooling, personality, and three economic areas (assets, income, and housing) over time.

The **ELIPSS Panel** in France was started in 2012 (N = ~1,000; http://quanti.dime-shs.sciences-po.fr/en/) by 7 research institutions as a pilot study. They sampled from something called the rotating census using a stratified two-stage probability sample of

households, randomly selecting one member. Potential participants were first sent a letter (half randomly received a 10 euro gift voucher), then were sent postal and telephone reminders, then were contacted face-to-face if they had not responded. Although the main study started in 2016 with 3,500 panel members (individuals aged 18-75), there is very little information available about it at the time of writing.

Additionally, the UK has a national probability panel at the National Centre for Research Methods, the **NatCen Panel** (N = ~4,000; http://natcen.ac.uk/taking-part/studies-in-field/natcen-panel/). Participants were recruited from the British or Scottish Social Attitudes surveys (were recruited face-to-face). The study is funded from non-profit organizations. They run 6 x 15 minute surveys per year, with each open for 6 weeks and expected sample sizes of either 1,250 or 2,500.

1.4 Online Non-probability Panels in New Zealand

There are online panels currently in existence in New Zealand. However, none of them use a national probability sample. As is the case with some of the larger American panels, sometimes there is very little publically available information on these panels. Indeed, some companies mention that they have an online panel or access to online panels (e.g. People for Information, MM Research, Nexus Research, and Public Voice) but there are no details at all on the panel anywhere on their website or elsewhere on google.

1.4.1 Market research panels

Consumerlink (Colmar Brunton is their "parent company"; https://consumerlink.co.nz/) claims to have over 250,000 active panel members through their Fly Buys Panel (participants are rewarded with Fly Buys points, must be older than 15, current Fly Buys cardholders and NZ residents) and have almost half a million surveys competed per year (if you calculate response rates they are using a very liberal definition of "active panel member"). They also advertise a business panel of 30,000 members. They claim to have a nationally representative sample but provide very little information on this other than that they draw panelists from Fly Buys members (which has "74% national household penetration") Their website says they have been doing this for over 10 years. An ad on their Facebook page claims a cost of $999 per question for a nationally representative sample of New Zealanders (as part of their omnibus surveys). **Colmar Brunton** also have an online panel that appears to be linked to Consumerlink, however, there is very little online information about it.

SAYit (run by UMR research; http://www.umr.co.nz/) rewards participants (online sign up) with Prezzy cards, fuel vouchers and iPad prize draws. They seem to draw quota samples (later weighted) of 1,000 and send mostly omnibus style surveys (http://sayit.co.nz/blog). Clients can purchase up to 10 questions at once and share the costs of demographic questions. The website for this panel says that they have a "track record for accuracy has been proven in real-world tests". They also have panels of hard-to-reach and sought after groups, including: dairy, beef, and lamb farmers, shareholders, "opinion leaders", and Māori and Pasifika.

The Reid Research panel (http://www.reidresearch.co.nz/) conducts surveys on their panel members (online sign up) in person, online, by mail and phone on tasting and trialing new products, looking at new concepts, and other social and market research. Participants are rewarded with prize draws. While their website seems to have a lot of content about survey

quality, the importance of being conscientious, and bemoaning the lack of transparency of most research, there is very little technical information about their panel available.

Nielsen Digital Voice (https://nz.digitalvoice.nielsen.com/pnl/nz/home) is based on secure software that tracks panel member's internet behaviour and ask panelists to complete the odd "pop up" survey.

Horizon Research (https://www.horizonpoll.co.nz/) is the most transparent of all of the market research panel providers. Horizon claims to have 2 "nationally representative" panels; both the HorizonPoll panel and the Horizon Research Māori panel (claims to represent 75 iwi), they say that they can deliver results for about 2,000 panelists in 24-48 hours. As an incentive, participants are entered into quarterly cash and prize draws. Horizon says that prize draws ensure that people are doing research for the right reasons and so the sample is less biased. Each participant is surveyed up to 2 times per month. They weight samples by six factors: age, gender, personal income, education, ethnicity, region, employment status, and party vote. Their surveys are also enabled for easy use on smartphones.

Perceptive Research Panel (https://www.perceptive.co.nz/) have an online panel of NZ residents aged 16 or older (14-15 year olds can join if their parent/guardian consents). Their aim is to send panelists at least 3 surveys per month but possibly up to 5-6 per month. Panelists earn approximately $2 per survey, once they reach $10.05 credit they can exchange this for vouchers or a donation.

Buzzchanel (http://www.buzzchannel.co.nz/) has a panel called "buzzthepeople". While they have a blog post about how high quality online panels are important, there is very little information about this panel, except that they claim sub-24 hour turnaround time and have given away $300,000 to charity through the incentives from the panel.

Opinionworld (https://www.opinionworld.co.nz/en-nz) is an online panel of NZ residents (aged 15+) that uses reward points and prize draws as incentives. Most surveys are said to range from 10-40 minutes and their website says that their clients include "businesses, governments, public bodies and similar organisations." Although not immediately obvious, the **YourVoice** panel is owned by the same company (Survey Sampling International LLC; a large company with 40 offices and over 4,000 employees worldwide) and when reading the website it is exactly the same material.

Marketpulse International (https://www.marketpulse.co.nz/) say that they have an online panel in New Zealand, but the links on their website do not work. They seem to have one NZ staff member.

1.4.2 Other

Additionally, **Bauer Media group** claim to have two online panels of thousands of New Zealanders (the "All Woman Talk" and "His-call" panels; http://www.bauermedia.co.nz/research). Camorra research market themselves on their ability to create custom online panels (www.camorraresearch.co.nz). Infield International maintained a farmers panel for three years through online and CATI surveys with a 90% response rate each quarter (http://www.infield-international.com/).

1.4.3 Council panels

Various district councils around NZ have created "people's panels", including Auckland (25,000 people, with between about 1,000 and 4,000 people completing each survey; www.aucklandcouncil.govt.nz/peoplespanel), Dunedin (up to a few hundred responses for each survey; http://www.dunedin.govt.nz/your-council/peoples-panel), and the newly-formed Hutt City Views panel. The stated purpose of these panels is to survey residents on issues, policies, and decisions

1.5 Summary

While there are a range of commercially available nonprobability online panels currently conducting research in New Zealand, they are of varying quality. Many nations worldwide have started probability online panels, including Australia, the USA, Canada, and a growing number of European nations. However, there is currently no online national probability sample panel in New Zealand.

2. Practical Considerations

2.1 Pilot Study

A strong theme in the literature is the need to conduct a pilot study. Although many panels have published reports and articles on the usefulness of incentives, enrolment packages including the use of advance letters, and sampling strategies, the effectiveness of different methods varies across nations and contexts (Blom et al., 2016).

2.1.1 Past pilot studies

The pilot study would need to be built into a broader timeline for the project. Many of the overseas panels, e.g. the Life in Australia and the GESIS panels, have conducted pilot testing. The GESIS panel conducted an extensive 8-month online panel where they experimented with different incentives and survey methodology. The Life in Australia panel had the following plan for their time line: field trial (n=400) in August 2016, the first survey of the trial participants in September; then in mid-September to mid-October they recruited their panel. Their first survey was mid-October until mid-November and their first results were obtained mid-December.

In the US, the RAND American Life panel has a non-probability sample frame that they use for pilot testing surveys (N=700; Gutsche, Kapteyn, Meijer, & Weerman, 2014). The GfK Knowledge panel conducted recruitment experiments (N=10,000) and found that while there were no effects for advance postcards (postal sample), they recommended sending a 1 week reminder postcard. They also found that they gained a 7.3% response rate for $0 incentive, 9.1% for $1, and 12.3% for $5. From this process they wished they had tested a wider range of incentive amounts by including a $2 incentive condition (DiSogra, Callegaro, & Hendarwan, 2009).

The LISS panel in the Netherlands tested a wide range of conditions on a sample of 410 households, including: CATI or CAPI; special or standard advance letter; comparing no incentive to a promised or prepaid incentive; incentive amounts 0, 10, 20 or 50 Euros; and introducing the panel in the advance letter or after the interview. The main lessons from their study was that the 10 Euro reward was the most efficient way to recruit as all monetary incentives increased the response rate by a similar amount. There were no effects for letter type, but focus group type research they had conducted indicated that the letter should be short and more informal. Additionally, people were two times more likely to respond if the incentive was prepaid (Scherpenzeel & Toepoel, 2012).

Similarly, the ELIPSS pilot study (N=1,026) tested incentives by giving half of their sample 10 Euros up front and half nothing. The participants who received the incentive were 1.4 times more likely to respond than those who did not (Blom et al., 2016). Furthermore, the GIP also conducted an incentive experiment, by giving half of the sample a 5 Euro bill upfront and half of the sample 10 Euros once they had finished the survey. The upfront 5 Euro incentive garnered a 8.9% higher response rate. They then tested an incentivized reminder by sending a reminder with 5 Euros to half of those who had not responded and a reminder with no incentive to the rest of the sample. The incentivized reminder led to a 30% response rate, whereas the non-incentivized one lead to a 13.7% response rate (Blom, Gathmann, & Krieger, 2015).

2.1.2 Pilot study considerations

Arising from a review of the literature there are several potential conditions to test in the pilot study.

Prenotification letter or postcard: Researchers send a postcard or letter to potential participants that alerts them to the fact that they will soon be surveyed/asked to join a panel. However, this has typically been done in phone surveys or face to face surveys as a warning of sorts and to add legitimacy to the phone call/face to face visit. Even in those situations there are mixed findings. There is no reason to believe it will help us, as a mail sample (see section below on sampling) in any way, this will likely just be as waste of money.

Varying letter/information/recruitment package designs: Some studies have varied the length, style, and use of persuasive techniques in their materials. We will not test this as very few effects have been found, except some suggest to make it glossy/fancy and keep it short and informal. This may be best to test through feedback from colleagues and friends or through formal focus group/qualitative research.

Non-contingent versus Contingent incentives: Researchers have found great results sending money or vouchers with the initial recruitment package, compared to promising to send a voucher etc. later to the participant. However, this is prohibitively expensive and the extra participants that this recruits may not be committed/provide quality data (see section on incentives).

Incentive amount: Researchers have found that incentives increase participation rates. Beyond tests of incentive vs. no incentive, researchers have systematically varied incentive amounts. At the pilot study phase it is good to vary incentive amounts (likely $5, $10, $15, and $20) to find the best balance between costs and response rates. We would do this, however, this is not easy to do under ethics regulations.

Initial survey order: One factor that could be varied is whether the initial survey begins with demographic questions (potentially boring, but useful for the researchers), or substantive/interesting questions. We considered doing this because this comes up in the literature a lot as a potential influencer but no one has thought to test it at these early phases - instead they have mentioned it as something that potentially may influence response rates (the stage between where people think "yeah I'll do the survey" and "screw it, this seems boring" and never finish the initial survey/fully join the panel. However, this would likely not work given our mode.

Reminder/Number of reminders: Researchers send a reminder and may vary the number and type of reminders (postcards or the full pack again?). We will send a reminder to everyone as we have found this to be a good way to boost response rates in our surveys. Our typical order is: initial mailing, postcard, then a repeat of the initial mailing with a shorter information letter. This will not be varied or manipulated

Ask first or ask last: Researchers either ask people to join the panel up front or after an initial survey (typically previously done in face to face and phone). This is mode dependent and would likely also relate to ethics. We will be telling participants up front that it is a panel.

2.1.3 Summary and recommendations

Most existing panel studies started with a pilot study where they could test sampling methods, software, and other considerations before the costly, main sample for their study. It is wise to conduct a pilot study, especially given that every context is different.

2.2 Sampling

The way that past studies have conducted sampling has varied depending what was available in the particular country of interest. The two viable options for national sampling for the panel in New Zealand would be through phone (RDD) and through the electoral roll (ABS). An additional option would be a face-to-face method, however, this is likely quite expensive and not widely done in New Zealand.

2.2.1 Electoral Roll

In New Zealand we are fortunate to have access to the electoral roll as researchers. A common approach that researchers use to get a nationally representative sample has been to take a random sample of people from the electoral roll. We could take a random sample of people for this study and recruit them via posted materials.

One of the advantages of using the electoral roll is that we have their address, which is an alternative means of contacting participants. We would also (instantly, without having to explicitly request it) get the associated electoral roll information like occupation, address based information like NZ deprivation index scores, occupation, Māori descent and so on.

Other panel studies have used address based sample frames. For example, the GfK panel uses all existing US postal service addresses. Rao, Kaminska and McCutcheon (2010) found that ABS recruited significantly more ethnic minorities than RDD. They also found that mail surveys were more cost effective given their response rate, the argument being that having attractive recruitment materials added professionality that a phone call could not, it also gave participants more time to choose to respond or not. Thus, participants did not feel pressured, say they would join the panel to the phone operator, and later decide not to.

One issue with using the electoral roll is that it does not include everyone in New Zealand. The electoral roll only includes permanent residents and citizens. Additionally, not everyone who is eligible to vote is enrolled to vote (although it is compulsory). The New Zealand Electoral Commission (2016) report that only 89.88% of eligible New Zealanders are enrolled to vote. Enrolment rates vary drastically over age group, with only 65.74% of 18-24 year olds and 78.70% of 25-29 year olds enrolled to vote, yet 98.39% of 55-59 year olds and 99.97% of those aged 70+ enrolled. It also varies by electorate, with the lowest by far being Auckland central, with only 51.15% enrolled to vote (likely reflecting the young age of the electorate), yet 98.19% of eligible voters and enrolled in Rodney and 98.13% in Ohariu.

Additionally, there is evidence to suggest that once recruited, participants from phone samples tend to stay in panel studies for longer (McCutcheon, Rao, & Raminska, 2014). This may be because the study was given a human voice or face and may be able to be cancelled out using other features in the recruitment process (glossy pamphlets, embedded video clips and so on).

2.2.2 Telephone sampling

The other viable option would be to recruit the sample from Random Digit Dialing. We would need to spend some time investigating and possibly contacting NZ market research companies to investigate how they do this and the costs associated with this option.

One of the key decisions would be choosing the right cellphone to landline balance, especially given the potential increase in cellphone only households in NZ. Data from the 2013 census shows that 85.5% of New Zealanders had access to a landline, and 83.7% to a cellphone (Statistics New Zealand, 2013). Internationally, Pew (USA) uses 75 cellphone/25 landline, GESIS in Germany uses 50 cellphone/50 landline, as does the Life in Australia panel.

Calling cellphones presents extra challenges. With a landline sample it is common for the researcher to request to talk to e.g. the person with the most recent birthday to prevent biases based on who answers the landline phone. The problem with cellphones it is hard to say "please give the phone to the male with the next birthday" etc. because they are individually owned and also you sometimes get younger people. For example, only 2% of cellphones in Germany are co-owned (Gabler at al., 2012). This can create bias, plus dialing cellphones adds extra time-related costs. Pew Research (2015) says that cellphone surveys in the US are one and a half to two times more expensive than landline surveys. Cellphone calls are not picked up because people can screen their calls and are often busy or working/driving/have privacy concerns based on their location/or have connection issues. In short, their willingness to answer or participate depends on their environment. As such, the number of calls that are made varies considerably, e.g. the GESIS pilot called 15 times (Gabler at al., 2012). This obviously affects the response rate but also the financial costs. However, research is yet to find data quality differences between landline and cellphone interview surveys.

2.2.3 Age

Additionally, and this depends on our sampling method (if using the electoral roll then potential participants will only be 18 and older, however, with RDD recruitment we would have access to younger people), we need to consider if there will be a lower and upper age limit to the sample (Blom et al., 2016). Younger samples require different ethical procedures and older panelists may not be able to use the internet, may have physical barriers to survey completion, and may have a higher dropout rate due to illness and mortality. Many panels have an upper age limit, e.g. 70 (GESIS), or 75 (GIP, ELIPSS) that is enforced at the time of recruitment. Although, participants stay in the panel once they pass these ages i.e. the studies do not remove them from the sample.

2.2.4 Summary and recommendation

There are two realistic recruitment options: via post from the electoral roll and via phone through RDD. The electoral roll seems to be a good option for New Zealand, and by all international estimates is typically cheaper. Both have non-coverage issues. In sum, the best option is to use a random sample from the electoral roll for the initial sample frame. It is also important to consider having an upper limit on the age range for the sample (typically 70-75).

2.3 Response Rates and Sample Size

Response rates vary depending on the phase of the study. These will determine the initial sample size and, ultimately, the size of the panel.

2.3.1 Sample Size

Past studies have varied considerably in size (see review of existing studies above). One consideration for size is the ability to provide custom panels of hard to survey populations. For example, Amerispeak advertises its ability to survey hard-to-reach populations like low-income households, smokers, and veterans. We would need to figure out how many participants to sample through the pilot study process and how many we would need overall in the final panel to be representative of the NZ population.

2.3.2 Past response rates for panel registration

The quality of a survey is affected by its response rate and the potential differences between respondents and non-respondents. This is a particular problem for panel studies though, as participants are not simply filling in one survey, they are signing up for a time commitment and effort. In online panel recruitment, there are many stages when participants need to respond. First is the introductory stage where panelists need to say they want to be on the panel, then they need to fill out the initial survey and consent form, then they need to actually fill out the regular surveys.

Across Europe, initial response rates have been between 31-73% for the recruitment survey and then turnout at 18-48% for panel registration (Blom et al., 2016). LISS performed best in Europe, and got 48% of the initially sampled households to join the panel. Struminskaya, Kaczmirek, Schaurer, and Bandilla (2014) report that the GESIS pilot study had an AAPOR RR3 response rate of 17.8% (N = 4,840). From there, the surveyor asked whether someone was a non-work internet user (N = 3,514), of those respondents 1,665 agreed to be on the panel and an email was sent to them. Finally, 1010 participants started the first survey and 934 finished it.

There are often biases introduced at initial response. In the US, the NORC Amerispeak panel has an AAPOR R3 rate of 34.3% (weighted to account for selection probabilities). When exploring the attitudes of those recruited through the NORC Amerispeak Nonresponse follow-up campaign they are more politically conservative, less knowledgeable about science, less interested in current events/the news, and are less likely to read a print newspaper. Kennedy and colleagues (2016) also found that black, Hispanic, younger people, and men were all less likely to join panels, making any estimates using these groups less accurate. In the NORC Amerispeak panel, rural people, those with lower incomes, cellphone only people, those aged 18-34, African Americans, Hispanics and people without education beyond high school are less likely to respond. To compensate, the survey has a different recruitment procedure called the Nonresponse follow-up campaign that more aggressively targets members from these groups.

International evidence also suggests that online panel members are more civically minded, i.e. they say that they vote and volunteer more often than non-panel members (Callegaro & DiSogra, 2008) and are more educated (Gronlund & Standberg, 2014). There may also be personality differences between those who opt in and those who do not. Bosnjak and colleagues (2013) found that participants who were successfully recruited to a German national probability panel were higher in openness to experience and extraversion, but lower in conscientiousness. However, this was largely attributed to the differences between internet users and nonusers and their younger age.

2.3.3 Initial Recruitment methods

If postal sampling is chosen, through our pilot study we may want to test the efficacy of pre notification letters and printed information. Evidence suggests that a postcard or pre-invitation letter boosts initial response rates (for summaries see de Leeuw et al., 2007; Scherpenzeel & Toepoel, 2014), although the authors conclude that pilot studies are crucial as this may depend on the topic, letter style, and the population of interest. Additionally, Link and Mokdad (2005) found that an advance letter meant that older, white, higher SES people were more likely to respond.

Unfortunately, it would likely be unrealistic for us to phone those sampled from the electoral roll due to ethics regulations. The GfK panel uses a series of mailings: an initial invitation letter, a reminder postcard, and then a follow up letter. Any addresses that can be matched to landlines are and 5 weeks after the initial mailing are called. People opt in via completing a paper form and mailing it back in a postage paid envelope, calling a toll free number, or going to a website and filling out a form there.

The LISS panel, GIP, ELIPSS pilot study, and GESIS panel all sent advance letters and various materials including brochures. Some of this advance notification was done in person (GIP, ELIPSS, GESIS), which is likely too costly for our project. After the first recruitment attempt, these studies used several follow ups, all dependent on how they were contacting participants (up to 3 reminders).

2.3.4 Obtaining basic data on participants

An advantage of panel studies is that you do not need to ask participants for their demographics on every survey. Demographics are often collected up front and on a regular basis, for example, respondents to GfK Knowledge Panel completed a “Core Profile” survey including all basic demographics and information about their households. Existing panel members are asked to complete these every year to keep information up to date. There is debate as to whether to start with core demographics or not since demographic surveys tend to be boring to participants (Struminskaya et al., 2014). Creating participant burden near the beginning of the panel study may not be the best idea for retention. A key consideration for the project will be deciding if there is a longitudinal component, including these demographics, any other contents and their frequency.

2.3.5 Past response rates for survey completion

Once people have opted into the panel, part of the challenge is over. Although past research has reported lurkers (i.e. those who do not complete every survey) and sleepers (i.e. those who stop responding; see section on Panel Attrition below) can provide considerable problems, most participants tend to respond to each survey. Key considerations are that participants may not be available during the period that the survey is available due to work, illness etc., or participants may not be interested in the survey topic. If the panel wishes to field longitudinal surveys with the same questions over time additional considerations must be made around participant boredom and potential overall panel attrition.

Some international panels provide estimates as to what we might expect in terms of response rates to each survey. The RAND American life panel has 70-80% of their participants respond to each survey. The response rates for the European panels are at around 90% per survey. However, response rates and retention vary a lot from study to study as different panels have

different policies for removing inactive members (Callegaro et al., 2014). Older and more educated participants in panels are more likely to consistently complete surveys (Bosnjak et al., 2013).

2.3.6 Weighting information

Although the surveys are national probability samples, they still weight based on various demographics. There are two key areas where probability panels may underperform: under-coverage and non-response (Bethlehem & Callegaro, 2014). Researchers need to decide the extent to which they are willing to spend money recruiting hard to reach participants and maintaining their presence on the panel (and perhaps their lower data quality) and weighting (which is lower cost; Roberts, Allum, & Sturgis, 2014).

Under-coverage. One area where weighting is relied on heavily is if a panel does not survey the offline population. They then attempt to weight to make up for the demographic biases (Callegaro & DiSogra, 2008). We are fortunate in NZ that we can access comprehensive official statistics easily, as this has been a problem in other countries (Struminskaya et al., 2014). This will be particularly good for benchmarking the eventual panel results against.

Non-response. This can occur through refusal, non-contact, or inability. Non-response can also happen due to the topic of a particular survey. Non-response may be systematic, thus weighting may be needed.

Because there are two phases where these biases can occur, at initial recruitment and in response to each survey, two steps of weighting are preferred (Bethlehem & Callegaro, 2014). The first step of weighting should be computed for all panel members to adjust any biases so the panel is nationally representative. The second step should make each survey representative of the panel.

2.3.7 How existing studies weight

The American Life Panel weights against the Current Population Survey. They initially uses three methods; cell-based post stratification, logistic regression, and raking. However, they now use raking (https://alpdata.rand.org/index.php?page=weights).

GfK uses a "patented methodology" using equal probability selection method – weighting the panel members to various demographics from the most recent US Census Bureau Current Population Survey. It actually sounds pretty standard when you read the details. They weight on all of their studies/reports in the end. Final weights use raking.

The Pew American Trends Panel has a multi-step weighting process. They calculate a base weight that includes the respondent's original survey selection probability, then adjust for the fact that different participants of the two surveys panelists are drawn from had different chances of joining the panel. The final step matches gender, age, education, race (including specifically Hispanic origin), and region to the US Census Bureau American Community Survey. They also match and weight on population density (2010 US Census), telephone service (2016 projections from the National Health Interview Survey), volunteerism (2013 Current Population Survey volunteer supplement), party affiliation (average of 3 most recent Pew polls), and internet use frequency (2013 Current Population Survey).

2.3.8 Summary and recommendations

Given the sample size of many past online national probability panel surveys, including the nearby Life in Australia panel, a sample size of around 3,000 is desirable. Through the pilot study the research team would be able to calculate a likely response rate at each phase for the full sample. We may need to use weighting at the end of some of the surveys due to non-response and attrition.

2.4 Panel Attrition

2.4.1 Panel maintenance

Maintaining a representative panel is a major issue for data quality as over time those who stop responding to surveys or drop out are different from those who stay in (e.g. Lugtig, 2012; Satherley et al., 2015). Researchers are also more likely to lose the contact details of groups who move house more (particularly present in groups who are less likely to own their own home). For example, we know that young people and Māori move house more, therefore they are harder to maintain contact with. Additionally, research commonly finds higher attrition among males, ethnic minority group members, singles, those without children, those not in the labour force, non-home owners, and the urban (Watson & Wooden, 2009; cf. Kruse et al., 2010 which was attributed to higher incentives for certain groups). However, while it is fairly easy to compare participants based on demographic information, and benchmark this against official statistics, this becomes a lot more difficult with psychological variables and those not measured officially. Research has found that retention is higher among the more introverted, agreeable, and conscientious, plus those with higher honesty-humility, need for cognition and need to evaluate (e.g. Lugtig, 2012; Satherley et al., 2015).

2.4.2 Participant to Researcher Contact

It is important for retention purposes for participants to be able to easily contact the study to update contact information. The Life in Australia has a toll free number that has the opening hours of 9am to 8pm AEST on weekdays and from 11am until 4.30pm on weekends.

Many of the studies have a phone hotline, some are staffed 24/7 (GESIS). Many have the capacity for direct messaging, and of course, email.

2.4.3 The types of attrition and how to combat them

Callegaro and DiSogra (2008) have summarized panel attrition into four categories: voluntary, passive, mortality, and panel-induced.

Voluntary Attrition is where people take the initiative to contact the study and explicitly drop out because they no longer want to be involved. Participants often do this by calling a toll free number or writing/typing in a comments box on their surveys (McCutcheon, Rao, & Kaminska, 2014). Panelists are unlikely to explicitly drop out and thus engage in **Passive Attrition**, that is, they do not explicitly withdraw, they just stop responding to the surveys. This increases costs as the research team repeatedly try to contact the participant. Therefore, any panel study needs to have a strategy on how to deal with people who are no longer responding to surveys. Evidence from the LISS panel suggests that some panelists may return if offered incentives (Scherpenzeel & Das, 2010). Scherpenzeel (2013) tried to find out how to bring lurkers (inconsistent responders) and sleepers (non-respondents) back to the LISS panel but gifts nor newsletters worked. Their eventual best practice became calling the

participants after 2 months of nonparticipation and offering an incentive. However, those phoning need to be trained, quality controlled, and trained in refusal avoidance (Kroh, 2011; Uhrig, 2008; Voorpostal & Lipps, 2011).

There is now research on the predictors of passive attrition, Kruse and colleagues (2010) showed that late responding to a previous wave, lack of income and health survey responses, and breaking off during the questionnaire predict non response to future surveys. Those who became "sleepers" in the LISS panel were more likely to be younger, in paid work, less-educated, and have children (Scherpenzeel & Toepoel, 2014).

Additionally, **Mortality Attrition** is unavoidable and a reason why participants need to be contacted via phone or mail after not responding for some time. Finally, there is **Panel-induced Attrition** which has also been called "forced turnover" or "retirement". This is where there are term limits on participation to prevent fatigue or where some panels drop participants e.g. in the Gallup panel if they do not answer five consecutive surveys (McCutcheon, Rao, & Kaminska, 2014). Annual attrition in the RAND American Life Panel has been between 6-15% (based off of 'active' panel members, 'active' meaning at least one survey per year). They say that panel members generally do not give notice, they instead stop completing surveys. RAND tries to contact these people personally and removes those who have not participated for one year automatically.

A good way to avoid attrition is the have a reasonable level of frequency of contact, as if a routine is established, i.e., a monthly survey to the panel, then monthly contact is made. It is also good to get multiple contact details for participants.

2.4.4 Topping up the panel

A key challenge in the success of online panel studies is maintaining a representative sample after attrition. This depends in part on the research questions that the researchers seek to answer. If they are longitudinal in nature then topping up the panel is less important, however, if they are cross-sectional they probably want to keep as representative of a sample as possible. As reviewed above, attrition is likely to inevitably happen, and be systematic, which may lead to biases in the data because the sample becomes no longer representative.

Extant studies have provided a few different approaches to this challenge. The Life in Australia study is intending to top up the sample using a single-frame mobile phone survey. The LISS study has conducted "refreshment" samples in 2009 (stratified to improve representativeness of those who are hard to survey), 2011 (random), and 2013 (again, stratified to oversample the hard to survey). The GIP conducted a refreshment sample in 2014, following the sample procedure as initial recruitment (random household sample).

2.4.5 Length of Panel Service

Any prospective panel needs to make a decision on how long that panelists can stay on board (see **Panel-induced Attrition** above). We would need to consider this in the planning phases as it would affect the budget (resampling costs money, and has been said to cost more than most initiatives used to keep participants in the panel).

2.4.6 Summary and recommendations

Attrition can greatly bias a panel over time, thus, all efforts should be made through the process to retain participants (much of this can be based on past research by other studies).

The literature seems to suggest that it is cheaper and easier to maintain panelists rather than resample, however, if we use a stratified sample from the electoral roll, resampling may be easier than for those who use RDD methods. Practical and consistent procedural decisions would need to be made on the length of panel service and what to do with those who consistently do not respond.

2.5 Survey Administration

There are a number of additional, "miscellaneous" type of administration considerations for an online panel study.

2.5.1 Ethics and opt in

According to Callegaro et al. (2014) a common approach now is to use a double opt-in process. They describe the process as follows: (1) people indicate their willingness to join by providing basic contact information e.g. their name and email (2) they are sent an email with a unique link (3) they click the link and fill in a questionnaire/initial enrolment survey with a full set of demographic questions this also typically includes privacy information (including how cookies are used), and information on rewards and membership (sometimes step 3 is split into a survey and an information step; see also ISO 26362).

We would need ethics for the pilot study, the initial sampling and survey, then ethics for each survey. This needs to be taken into consideration when deciding on the frequency of surveys sent to panelists.

2.5.2 Website and software

Panel studies need two websites. Firstly, one with general information, including contact information and frequently asked questions. This website also serves as a point of contact for journalists, researchers, and other outsiders that wish to learn about the panel. Secondly, many panels have a website that panelists can log into. This allows them to see available surveys, update their contact information and so on. Sometimes studies do not have this second website and instead contact participants via email.

Macer (2014) conducted a review of online panel software. There are a number of paid software providers that market researchers use and a few open-source developers. This would likely be something that would need to be weighed up in terms of cost compared to researcher hours spent doing these tasks. Online panel survey is particularly useful for ease of accessibility for panelists, it makes it easier to track responses and send reminders, and also it makes it easier to keep track of their rewards.

Another consideration is the use of apps. The Gallup and ELIPSS panels both have apps. Apps can be useful for concerns about mode effects, as they make everything consistent across devices and browsers (Blom et al., 2016).

2.5.3 In house versus contracting out

A decision that must be made by the research team (budget dependent of course), is what will be done in house and what will not (Blom et al., 2016). The LISS and ELIPSS panels manage their entire operations within their respective research institutes. The GESIS panel manages most things in house, although they use a mailing agency for their paper surveys and their phone hotline. The GIP works closely with a data collection agency for a lot of their process, except questionnaire development and testing.

2.5.4 Frequency and Length of surveys

It is important to balance participant burden (asking the participants to do too much, too often) with contact and opportunity (being in regular contact means you are less likely to lose participants and you are providing participants more of a chance to have their say/earn rewards).

Past studies have taken a variety of responses. The Life in Australia surveys take between 10 and 15 minutes to complete and participants are surveyed approximately once a month. The surveys stay live for a couple of weeks (data collection is undertaken over 2-3 weeks), three reminders are sent via email, text message and phone. Collaborators can request standalone surveys or add questions to another survey, they can also run longitudinal studies. The study provides "standard outputs" within 2 working days.

The European surveys seem to be longer and either monthly or bi-monthly. In Europe, LISS runs a 30 minutes survey on the first Monday of every month, GIP runs a 20-25 minute survey on the first day in uneven months (bi-monthly), ELIPSS does a 30 minute survey on the first Thursday of every month, and GESIS sends a 20-25 minute survey on the 15th in even months (bimonthly). These European panels tend to remind people to complete the survey after one week and again after two weeks. The NatCen panel in the UK surveys participants 6 times per year.

American panels tend to run more surveys, perhaps reflecting differences in funding structures (they tend to have commercial clients). The GfK panel aims to assign no more than one survey per week, however, on some weeks panel members receive more invitations. Typically, panel members first get an email with a custom link, an email reminder after 3 days, and a phone call four days later. Surveys are typically 10-15 minutes in length. The Gallup panel surveys members on average 3 times per month and NORC Amerispeak does 2-3 per month.

2.5.5 Choosing when to close off the survey

One decision that needs to be made is how long the surveys are available to complete. If this time is too short the study may lose out on those who are sick or busy. We would also need to decide if participants can return to surveys they have missed (this could be on a survey-by-survey basis). There are also additional considerations around the length of this time period when dealing with the offline population. If the offline population is surveyed through the post then this time period needs to be longer than if it is via phone.

2.5.6 Costs

There are a number of costs associated with setting up and maintaining an online probability panel. The main ones are summarized below.

Sampling costs: either access to the electoral roll, printing, and postage, or the costs for RDD sampling, printing, and postage. These will vary based on what is in the information packages, the number of reminders, and if a pre-notification package is sent or not.

Incentive costs: the costs associated with rewarding the participants. This will vary depending on whether they are contingent on response or not. There are also ongoing costs associated with rewarding each participant for each survey.

Website costs: the costs of creating a website for both interested researchers/collaborators (that would contain e.g. technical documents) and a website for participants.

Software costs: the costs associated with the panel maintenance software.

Branding costs: creating a logo and cohesive study image for all materials.

Panel Relationship Management/Staff costs: some studies have a panel relationship manager who answers emails and messages. There are also costs associated with having a 0800 number and costs associated with other staff.

Offline Sample costs: there may be extra costs associated with contacting the offline panel. Either phone related costs or mailing costs.

2.5.7 Costs for clients

There is very little publically available information online about the price of including questions or running surveys on the probability panels. The Life in Australia survey say they are somewhere between the costs of a CATI and non-probability panel. Amerispeak will provide an omnibus representative study of 1,000 Americans for US$1,000 per question for the first group of three questions, US$800 for each subsequent question, and US$1,200 for each non-coded open-ended question. The client also gets standard demographic information (age, gender, education, race/ethnicity, and 4 categories of household income). The RAND American Life Panel provides the following price guide in USD: "$3.00 per interviewee minute for the first 500 respondents, $2.50 per interviewee minute for the next 500 respondents, $2.00 per interviewee minute for any respondent beyond 1000, $2,000 per survey handling costs. For example, 1,500 respondents taking a 10 minute interview would cost $39,500."

2.6 Incentives

Another decision that we need to make is the frequency and rate of incentives. Financial incentives are used in research to create reciprocity norms, also called the "Leverage Saliency" theory (i.e. I give you something, you give me something; Lavrakas et al., 2012; Scherpenzeel & Toepoel, 2014). Incentives have been shown to particularly help boost response rates in groups who typically do not respond to surveys (Singer, 2002). Two kinds of incentives are typically used: prepaid (paid before survey completion) and postpaid (paid after survey completion; Callegaro et al., 2014). While prepaid incentives are logistically challenging and may allow participants to cheat the system, they do create higher response rates as they build trust and the need for reciprocation (Lavrakas et al., 2012; Toepoel, 2012). However, this initial zeal to participate that was created by the incentive wears off. Research shows that initial incentives create higher response rates but have no effect on attrition (McCutcheon, Rao, & Kaminska, 2014). Researchers have to be careful to not create a situation where participants are only responding to the survey to get the incentive, and thus, race through the questions and provide rubbish data or respond in a nonsensical way due to resentment (Groves & Couper, 1998; Malhotra, Miller, & Wedeking, 2014; Roberts, Allum, & Sturgis, 2014).

Overall, incentives are a good idea (McCutcheon, Rao, & Kaminska, 2014). Callegaro and colleagues discuss that this may be because, especially early on, they create trust between the participant and the researcher. Charity donations also provide a chance for participants who

are uncomfortable with personal rewards, although by themselves they have not been shown to be efficacious (Goritz & Neumann, 2013).

The other possibility is that participants go into a draw/lottery/raffle for a larger prize or prizes. A prize draw is a cheaper option (and easier to administrate) but may be less appealing to participants. However, it is better over no incentive. In a meta-analysis, Goritz (2006) found that participants were 19% more likely to start a study that offered a prize lottery and 26% more likely to finish one. This may not be the case for panels though because of their differential time investment, although this may depend on the topic of the survey (i.e., commercial or not; Goritz & Luthe, 2012). Some panels have used a points based system, where points can be exchanged for vouchers or prizes.

2.6.1 How other studies deal with incentives

In the Life in Australia survey participants receive $5-$10 for each survey that they can donate to charity, receive as money through PayPal, or as a Coles (grocery) or Myer (department store) gift card. Participants cannot use their charity donations for tax assessment purposes. The study selected four large, registered charities that they say may change over time, and has donated the following amounts as of March 2017: the Australian Wildlife Conservancy ($8,240), CanTeen ($18,970), UNHCR Australia ($8,910), and the White Ribbon Foundation ($10,230).

In Europe, some panels offered an unconditional prepaid incentive to join of €5 (GIP, GESIS), €10 (LISS), or in the case of ELIPSS a tablet for personal use as well as survey use (see section on the offline population; notably 60% of participants said this was their motivation to participate in the study; Blom et al., 2016). These panels then offer between €5 and €10 for participation in each survey, LISS operates off of the guideline of €15 for every hour spent answering questionnaires. Additionally, the GIP pays a €10 yearly bonus for participation in all waves of the survey and €5 if someone participated in all but one survey that year. The LISS survey does bank transfers every 3 months, GIP sends bank transfers, Amazon vouchers or charity donations every 6 months. The frequency of payment can be a pain.

GfK panel enters respondents into raffles, but sometimes offers cash rewards. Pew American Trends Panel mentions “a small monetary incentive” for each survey and offered one upon joining. However, research suggests that the most effective incentives for response rates may be a flat contingent reward that participants get if they continue to complete surveys (i.e., $25 per month Lavrakas et al., 2012).

2.6.2 Other methods that encourage participation

Beyond financial incentives and gifts, researchers have used a wide range of tactics to keep participants interested in participating (as summarized in Scherpenzeel & Toepoel, 2014). Part of this is about creating a culture or brand for participants, as though they are an influencer or doing a service to society. Framing the study as in the public interest is a common strategy. Other tactics include: sending newsletters (LISS), seasonal/birthday greetings (relying on positive regard type concepts from social psychology, i.e., if you like someone you will do things for them), fridge magnets, pens, appreciation of participation cards, and embedded YouTube videos from academics involved in the study (Blom et al., 2016; Laurie & Lynn, 2009; Scherpenzeel & Toepoel, 2014). Surveys show that participants most like postcards out of these methods (61% of LISS study participants liked them or very much liked them; Scherpenzeel & Toepoel, 2014).

However, different methods work differently for subgroups, e.g. newsletters and results from the study have been shown to keep particularly those with less education involved in panels, men are more likely to respond if shown graphs, and the highly educated were less likely to respond after a YouTube video (in the LISS panel; Scherpenzeel & Toepoel, 2014). Interaction between the researcher and the researched or putting a human face or faces on the study helps too. Tailoring the materials as much as possible to the participant has been shown to help (Dillman, 2007). Although also useful to the research team, obtaining as much contact information as possible helps to encourage participation as it indicates investment (Sue & Ritter, 2007). Personalized reports have also been used as incentives, so people can see how they score in relation to others. While participants say that a personalized results summary would increase their likelihood of responding, the actual results are mixed (Edwards et al., 2009; Goritz, 2006).

2.6.3 Summary and recommendations

Researchers have found that incentives, especially prepaid or unconditional incentives, help to boost response and retention rates. Additionally, other methods such as building a panel brand, providing results summaries, and sending greeting cards may also help. Given the vast number of findings that incentives help to improve response rates, incentives should be a part of the panel study. A cost effective way to incentivize is by postpaid or contingent incentives, rather than the prepaid kind which creates ethical and practical issues. Other methods should be trialed to help improve response rates throughout the panel study process.

2.7 Accessibility

2.7.1 Language

Most studies in NZ only send surveys in English, but this misses out on some parts of the population. Data from the 2013 census (Statistics New Zealand, 2014) indicates that 87,534 people surveyed did not speak English (2.2% of people "were able to have a conversation about everyday things in at least one language" and did not include English). The most common languages these people spoke were as follows: Sintic not further defined (including Chinese) 13.7%, Yue (including Cantonese) 12.1%, Northern Chinese (including Mandarin) 11.7%, Samoan (11.2%), and te reo Māori (10.2%). The majority of those who said they could not speak English (63.8%) identified with an Asian ethnicity. Most of the non-English speakers lived in the Auckland region (65.3%). If we were to randomly sample from the electoral roll this number (2.2%) of non-English speakers may be lower as theoretically, people who are at least permanent residents of NZ may have spent more time in NZ and had more time to learn English than those who are surveyed in the census.

2.7.2 How other studies have dealt with other language speakers

Most panels worldwide tend to survey in one language. However, in the United States, it is particularly difficult to survey the Spanish speaking population. Amerispeak does surveys in Spanish. GfK KnowledgePanel Latino was created in 2008 to get representative Hispanic panels by targeting households where Spanish is spoken at least half of the time. They recruited this panel using a dual-frame RDD method targeting telephone exchanges with a higher proportion of Hispanics. The panel is designed to supplement the main panel (which can do surveys in English or Spanish).

An issue found in the US with Asian and Hispanic participants in market research panels is that these participant's identities were often faked (Baker et al., 2014). Thus, in non-probability samples there are people that join panels and lie about their characteristics to be surveyed more and earn more rewards. The implication for probability samples, however, may be that those who speak English in the household fill out the survey under the name of the sampled participant who may not speak English in order to gain the incentive.

2.7.3 Households without internet or computer access

How many New Zealanders do not have internet access and how do they differ from those who do? The NZ version of the World Internet Project survey (Crothers, Smith, Urale, & Bell, 2015) says that 91% of New Zealanders are active internet users. COMPASS found that 89% of the 2016 ISSP sample said that they had internet access at home. These are the most up-to-date figures available, however, the 2013 census found that 76.8% of people had internet access in their home (Statistics New Zealand, 2013).

Researchers refer to this as the "digital divide". International research shows that these households are more likely to be older, of lower socioeconomic status, lower education, ethnic minorities, live in certain regions of the country (i.e., non-urban), as well as country specific biases (Callegaro, 2013; Pennay et al., 2015; Ragnedda & Muschert, 2013). Indeed, benchmarking research has shown that online panelists are heavier internet users than participants in postal panel surveys (Gronlund & Standberg, 2014).

Another issue are those that have internet access but do not really know how to use a computer/the internet or do not like to. These have been called "Net Averse" by the Amerispeak panel. This group of participants is hard to define because on the one hand you do not want a lot of participants in the non-online sample due to the increased costs, yet you do not want to miss out on potential participants.

2.7.4 How other studies have dealt with the offline population

There are a number of ways that extant studies have dealt with internet non-coverage or the computer illiterate/net averse (Blom et al., 2016; DiSogra & Callegaro, 2010). One way this has been dealt with in the past is if someone does not access the internet for personal use more than twice a week they are counted as an offline participant (McCutcheon, Rao, & Kaminska, 2014). Other studies simply allow people to choose. The Amerispeak panel find that 79% of participants are happy complete the study online, but phone the other 21%. However, in later recruitment efforts (2016) they restricted the sample that they would phone to those without internet access, those only with smart phones and no other internet access, and those unwilling to share their email address.

Researchers have been conscious that there may be mode effects, e.g. that online respondents might be taking less care than the pen and paper respondents, or that phone respondents may have a lower rate of "don't knows" as they feel accountable to the interviewer (Greszki, Meyer, & Schoen, 2014). Online respondents are also probably less vulnerable to social desirability bias when compared to those answering in other mode (for a summary of mode effects see Chang & Krosnick, 2009).

The strategies used by online panels to deal with the offline or net averse population can be summarized into four categories:

1. **Give all panel recruits the same device and a connection to the internet.**

Some studies have very large budgets and are able to provide every participant with equal access to the online surveys. For example, the ELIPSS pilot provides everyone with a tablet and 3G internet so that participants can access their panel app. This strategy eliminates mode effects but does not account for those who do not know how to use the internet/computers or do not want to.

2. **Give panel recruits with no internet a device/an internet connection.**

Other studies allow for the offline population by taking them online. In the case of the LISS, and the original sample in the GIP panels, the study installed a computer that which is equipped for the visually impaired and features a special button to push that takes the participant to the survey page. The GIP moved to tablets from their 2014 sample onwards. The RAND American Life Panel provide Chromebooks (laptops) and Wi-Fi hotspots for those without access to the internet or a computer (around 10% of the sample). This is a cheaper solution than to give everyone online survey access, however, it still does not allow for those who are not confident internet users or who prefer to complete the survey offline.

However, the first two solutions are also logistically and ethically challenging at the beginning of the study.

3. **Survey those without a device/internet connection through another mode.**

For example, the GESIS panel mails a paper copy of the survey to those with no internet at home or those who prefer a paper copy. The Gallup panel lets people do phone interviews if they prefer. The NatCen Panel phone surveys the 14% of British participants that do not have internet. In Australia for 2014-15, the Australian Bureau of Statistics reported that 86% of households had internet access at home. As such, the Life in Australia study allows participants to complete surveys on the phone with an interviewer if they either: a) do not have internet access, or b) prefer to complete the surveys that way. The Pew American Trends Panel simply posts a paper copy of the survey to participants who do not have the internet or who did not provide an email address.

4. **Ignore the non-internet population and correct with sample weights.**

This is the most cost effective way of dealing with the issue. See the section on weighting later in this report.

2.7.5 Summary and recommendations

Survey accessibility for non-English speakers proves to be a substantial practical issue. However, this affects a very small (2.2% in the last census) number of New Zealanders, many of which may not be on the electoral roll. It would likely be quite costly to have these few participants in the panel. A larger issue is the offline or net averse population. There are real differences between the online and offline population and efforts must be made to recruit these participants. Providing internet and devices in impractical for our study, however postal surveys seems to be a good option given the expertise of the research team.

3. Final Summary and Recommendations

The purpose of this report was to investigate the feasibility of creating a national probability online panel study. The report reviewed existing studies and from them, covered key considerations for setting up such a study. Overall, it is certainly feasible to run a probabilistic online panel study in New Zealand. However, the report recommends the following factors be taken into account.

3.1 Pilot Study

It is recommended by most sources that researchers conduct a pilot study to test sampling, methods, and procedures. This report concludes that the online panel study should first conduct a smaller sized (in comparison to the main study) pilot study. The main thing that should be tested is the efficacy of the sampling method i.e. the initial response rate and whether this is demographically biased in any way, then the response rates for the subsequent surveys and the ways that the sample may be biased on social/psychological/health indicators. The pilot study also provides researchers with the opportunity to 'fine tune' procedures before going on to the main study where errors may be more costly.

3.2 Sampling

The report weighed up both postal (electoral roll) and phone (random digit dialing) sampling methods. Both have non-coverage biases. However, postal sampling is more cost effective and the research team have extensive experience conducting postal samples. In New Zealand, we are fortunate enough to have access to the electoral roll and demographic details associated with each enrolled voter. Hence, it is recommended that the study be conducted through an electoral roll sample. Another consideration is the age range to be sampled. Since the minimum age on the electoral roll is around 18, the sample would have to be those aged over 18. However, it is recommended that the maximum age be limited to around 75, as internet access has been shown to be lower for those older than 75. Additionally, these participants may be reluctant to opt in to long term studies and have a higher risk of mortality or drop out due to ill health.

3.3 Sample Size, Response Rates, Survey Length, and Survey Frequency

The report covers the sample sizes of other studies, the length of their surveys, how frequently they survey participants, and what they have found in terms of response rates. Looking at the sizes of comparable studies, such as the Life in Australia panel, it is recommended that the study aim to have a final sample of around 3,000 participants signed up the panel. The number of people we need to sample to get such a response rate will be established in the pilot study (see recommendations above). Additionally, based on other studies, the surveys should be up to half an hour in length (see also incentives section below for an additional consideration) and participants will be surveyed at least a few times per year (to maintain minimum contact levels), but no more than once a month. This would also be dependent on the level of interest from clients to run surveys, such as academics and government.

3.4 Incentives

A common finding in the literature is that studies which incentivise participation received higher response rates. Thus, the panel study should incentivise respondents. An incentive of approximately $10 per half hour/$20 per hour survey seems appropriate given the rate paid in

comparative international studies. The study would also need to explore the possibility of making charitable donations for those who do not wish to receive rewards themselves. It would also be worth trying different non-monetary rewards for participants, such as greeting cards, summaries of findings, and other features that help to create the idea that the panellist is part of a larger "influencer" group/brand.

3.5 Accessibility

This report covered two issues people may have in accessing the online panel surveys. Firstly, there may be language barriers in that some New Zealanders do not speak English. This issue only affects a small proportion of potential participants (a maximum of 2.2%). Furthermore, the costs involved in remedying this issue would be consuming of disproportionate time and money relative to the benefit of including this small number of participants. Secondly, some potential participants do not have internet access or do not wish to complete surveys online (they may not opt into the sample because while they have a computer, they do not use it). It is recommended that the study send mail copies of the survey to these participants. Typically such participants are phoned as the timeline is shorter, however, for the purposes of the panel study the amount of time given to participants will be longer than some of its international comparisons.

References

Baim, J., Galin, M., Frankel, M. R., Becker, R., & Agresti, J. (2009). *Sample Surveys Based on Internet Panels: 8 Years of Learning.* In Worldwide Readership Symposium, Valencia, Spain.

Baker, R., Miller, C., Kachhi, D., Lange, K., Wilding-Brown, L., & Tucker, J., (2014). Validating respondents' identity in online samples: The impact of efforts to eliminate fraudulent respondents. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 441-451). Hoboken, NJ: Wiley.

Bethlehem, J. & Callegaro, M. (2014). Part IV weighting adjustments: Introduction. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 128-134). Hoboken, NJ: Wiley.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods, 27*(4), 391-408.

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A. S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review, 34*(1), 8-25.

Bosnjak, M., Haas, I., Galesic, M., Kaczmirek, L., Bandilla, W., & Couper, M. P. (2013). Sample composition discrepancies in different stages of a probability-based online panel. *Field Methods, 25*(4), 339-360.

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly, 72*(5), 1008-1032.

Callegaro, M., Villar, A., Yeager, D., & Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 23-54). Hoboken, NJ: Wiley.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*(4), 641-678.

De Leeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys a meta-analysis. *Public Opinion Quarterly, 71*(3), 413-443.

DiSogra, C., & Callegaro, M. (2010). *Computing response rates for probability-based online panels*. In Proceedings of the Joint Statistical Meeting, Survey Research Methods Section (pp. 5309-5320).

DiSogra, C., Callegaro, M., & Hendarwan, E. (2009, August). Recruiting probability-based web panel members using an address-based sample frame: Results from a pilot study

conducted by knowledge networks. In *Joint Statistical Meetings*. Full paper available at http://www.knowledgenetworks.com/ganp/reviewer-info.html.

Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method.* NY: John Wiley & Sons.

Drewes, F. (2014). An empirical test of the impact of smartphones on panel-based online data collection. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 367-386). Hoboken, NJ: Wiley.

Edwards, P. J., Roberts, I., Clarke, M. J., DiGuiseppi, C., Wentz, R., Kwan, I., ... & Pratap, S. (2009). Methods to increase response to postal and electronic questionnaires. *The Cochrane Library.*

Gabler, S., Häder, S., Lehnhoff, I., & Mardian, E. (2011). Weighting for unequal inclusion probabilities and nonresponse in dual frame telephone surveys. In *Telephone Surveys in Europe* (pp. 147-167). Springer Berlin Heidelberg.

Göritz, A. S. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science, 1*(1), 58-70.

Göritz, A. S., & Luthe, S. C. (2013). How do lotteries and study results influence response behavior in online panels?. *Social Science Computer Review, 31*(3), 371-385.

Göritz, A. S., & Neumann, B. P. (2016). The longitudinal effects of incentives on response quantity in online panels. *Translational Issues in Psychological Science, 2*(2), 163.

Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 238-262). Hoboken, NJ: Wiley.

Gronlund, K. & Standberg, K. (2014). Online panels and validity: Representativeness and attrition in the Finnish eOpinion panel. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 86-103). Hoboken, NJ: Wiley.

Groves, R. M. & Couper, M. P. (1998). *Nonresponse in household interview surveys*. NY: John Wiley and Sons Inc.

Gutsche, T. L., Kapteyn, A., Meijer, E., & Weerman, B. (2014). The RAND continuous 2012 presidential election poll. *Public Opinion Quarterly, 78*(S1), 233-254.; https://alpdata.rand.org/?page=panelcomposition

Kroh, M. (2011). *Documentation of sample sizes and panel attrition in the German Socio Economic Panel (SOEP) (1984 until 2013).* Retrieved from http://www.diw.de/documents/publikationen/73/diw_01.c.492113.de/diw_datadoc_2014-075.pdf

Kruse, Y., Callegaro, M., Dennis, J. M., Subias, S., Lawrence, M., DiSogra, C., & Tompson, T. (2009). *Panel conditioning and attrition in the AP-Yahoo! news election panel*

study. In 64th conference of the American Association for Public Opinion Research (AAPOR), Hollywood, FL.

Laurie, H., & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. *Methodology of longitudinal surveys*, 205-233.

Lavrakas, P. J., Dennis, M. J., Peugh, J., Shand-Lubbers, J., Lee, E., & Charlebois, O. (2012). *Experimenting with noncontingent and contingent incentives in a media measurement panel*. Paper presented at the 67th annual conference of the American Association for Public Opinion Research, Orlando, Florida.

Link, M. W., & Mokdad, A. H. (2005). Alternative modes for health surveillance surveys: an experiment with web, mail, and telephone. *Epidemiology, 16*(5), 701-704.

Lugtig, P. (2012). *I think I know what you did last summer: Improving data quality in panel surveys.* PhD dissertation: University of Utrecht, The Netherlands.

Macer, T. (2014). Online panel software. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 413-440). Hoboken, NJ: Wiley.

Malhotra, N., Miller, J. M., & Wedeking, J. (2014). The relationship between nonresponse strategies and measurement error: Comparing online panel surveys to traditional surveys. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 313-336). Hoboken, NJ: Wiley.

McCutcheon, A. L., Rao, K., & Raminska, O. (2014). The untold story of multi-mode (online and mail) consumer panels: From optimal recruitment to retention and attrition. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 104-126). Hoboken, NJ: Wiley.

Pennay, D. W., Neiger, D., & Lavrakas, P. J. (2016). *A TSE comparison of findings from probability-based surveys and non-probability online panel surveys in Australia*. International Total Survey Error Workshop: Australian Bureau of Statistics.

Pew Research. (2015). *Advances in Telephone Sampling*. Retrieved from http://www.pewresearch.org/2015/11/18/advances-in-telephone-survey-sampling/

Postoaca, A. (2006). *The anonymous elect: Market research through online access panels.* Springer Science & Business Media.

Ragnedda, M., & Muschert, G. W. (2013). *The digital divide: The Internet and social inequality in international perspective* (Vol. 73). Routledge.

Rao, K., Kaminska, O., & McCutcheon, A. L. (2010). Recruiting probability samples for a multi-mode research panel with internet and mail components. *Public Opinion Quarterly, 74*(1), 68-84.

Roberts, C., Allum, N., & Sturgis, P. (2014). Nonresponse and measurement error in an online panel: Does additional effort to recruit reluctant respondents result in poorer quality data? In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick,

& P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 337-362). Hoboken, NJ: Wiley.

Satherley, N., Milojev, P., Greaves, L. M., Huang, Y., Osborne, D., Bulbulia, J., & Sibley, C. G. (2015). Demographic and psychological predictors of panel attrition: evidence from the New Zealand Attitudes and Values Study. *PloS one*, *10*(3), e0121950.Scherpenzeel, 2011

Scherpenzeel, A. (2013). *Experiences from a probability based Internet panel: Sample, recruitment, and participation.* Paper presented at the web Surveys for the General Population: How, Why and When? Conference, NRCM Network of Methodological Innovation, London.

Scherpenzeel, A., & Bethlehem, J. (2010). How Representative are Online Panels. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*. Oxford: Taylor and Francis.

Scherpenzeel, A., Das, J. W. M., Ester, P., & Kaczmirek, L. (2010). 'True' longitudinal and probability-based internet panels: Evidence from the Netherlands. *Social and behavioral research and the Internet*, 77-103.

Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a probability sample for an online panel effects of contact mode, incentives, and information. *Public opinion quarterly*, *76*(3), 470-490.

Scherpenzeel, A. & Toepoel, V. (2014). Informing panel members about study results: Effects of traditional and innovative forms of feedback on participation. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 192-207). Hoboken, NJ: Wiley.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse*, (pp. 163-177). Hoboken, NJ: Wiley.

Statistics New Zealand (2014). *2013 Census Quickstats about Culture and Identity.* Retrieved from http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity/languages.aspx

Statistics New Zealand (2013). *2013 Census Quickstats about National Highlights: Phone and Internet Access.* Retrieved from http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-about-national-highlights/phones-internet-access.aspx

Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing representativeness of a probability-based online panel in Germany. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 61-85). Hoboken, NJ: Wiley.

Ritter, L.A. & Sue, V.M. (2007). Selecting a Sample. *New Directions for Evaluation*, *115*, 23-28.

Toepoel, V. (2012). Effects of incentives in surveys. In L. Gideon (ed.), *Handbook of survey methodology for the social sciences* (pp. 209-226). NY: Springer.

Tourangeau, R., Conrad, F. G., & M. P. Couper (2013). *The Science of Web Surveys*. NY: Oxford University Press.

The New Zealand Electoral Commission (2016). *Enrolment Statistics by Electorate*. http://www.elections.org.nz/research-statistics/enrolment-statistics-electorate

Uhrig, S. C. N. (2008). *The nature and causes of attrition in the British Household Panel Study*. Retrieved from https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2008-05.

Vonk, T., van Ossenbruggen, R., & Willems, P. (2006). *The effects of panel recruitment and management on research results: A study across 19 online panels*. Paper presented at the Panel research 2006 ESOMAR, Barcelona.

Voorpostel, M., & Lipps, O. (2011). Attrition in the Swiss Household Panel: Is change associated with drop-out?. *Journal of Official Statistics, 27*(2), 301.

Williams, J. (2012). Survey methods in an age of austerity Driving value in survey design. *International Journal of Market Research, 54*, 35-47.

Crothers, C., Smith, P., Urale, P. W. B., & Bell, A. (2015). *The Internet in New Zealand 2015*. Retrieved from http://www.aut.ac.nz/__data/assets/pdf_file/0008/635669/150416-Online-Version-WIPNZ-2015-April-15.pdf

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, *75*(4), 709-754.