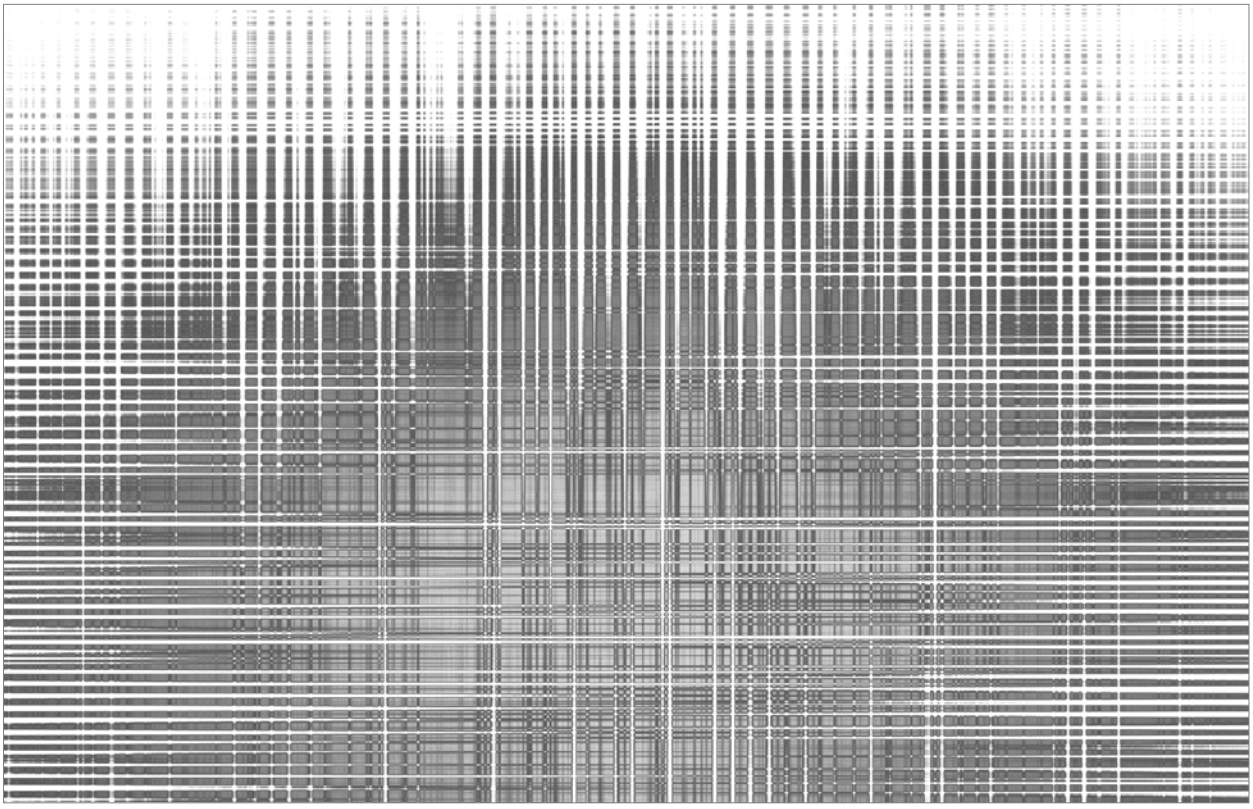
Australian National University

CENTRE FOR SOCIAL RESEARCH & METHODS

Social Research Centre

A SUBSIDIARY OF THE AUSTRALIAN NATIONAL UNVERSITY



The Online Panels Benchmarking Study: a Total Survey Error comparison of findings from probability-based surveys and nonprobability online panel surveys in Australia

D.W. Pennay, D. Neiger, P.J. Lavrakas and K. Borg

CSRM & SRC METHODS PAPER

NO. 2/2018

Series note

The ANU Centre for Social Research & Methods (CSRM) was established in 2015 to provide national leadership in the study of Australian society. CSRM has a strategic focus on:

- development of social research methods
- analysis of social issues and policy
- training in social science methods
- providing access to social scientific data.

CSRM publications are produced to enable widespread discussion and comment, and are available for free download from the CSRM website (http://csrm.cass.anu.edu.au/research/publications).

CSRM is located within the Research School of Social Sciences in the College of Arts & Social Sciences at the Australian National University (ANU). The centre is a joint initiative between the Social Research Centre and the ANU. Its expertise includes quantitative, qualitative and experimental research methodologies; public opinion and behaviour measurement; survey design; data collection and analysis; data archiving and management; and professional education in social research methods.

As with all CSRM publications, the views expressed in this Methods Paper are those of the authors and do not reflect any official CSRM position.

Professor Matthew Gray
Director, ANU Centre for Social Research & Methods
Research School of Social Sciences
College of Arts & Social Sciences
The Australian National University
May 2018

Methods Paper No. 2/2018

ISSN 2209-184X
ISBN 978-1-925715-04-0

An electronic publication downloaded from http://csrm.cass.anu.edu.au/research/publications/methods-research-papers.

For a complete list of CSRM working papers, see http://csrm.cass.anu.edu.au/research/publications/methods-research-papers.

ANU Centre for Social Research and Methods
Research School of Social Sciences
The Australian National University

The Online Panels Benchmarking Study: a Total Survey Error comparison of findings from probability-based surveys and nonprobability online panel surveys in Australia

D.W. Pennay, D. Neiger, P.J. Lavrakas and K. Borg

Darren Pennay is Chief Executive Officer of the Social Research Centre; a Campus Visitor at the Centre for Social Research & Methods, Australian National University; and an Adjunct Professor with the Institute for Social Science Research, University of Queensland.

Dina Neiger is Chief Statistician at the Social Research Centre and a Campus Visitor at the Centre for Social Research & Methods, Australian National University.

Paul J. Lavrakas is Senior Methodological Adviser to the Social Research Centre, Senior Fellow with NORC at the University of Chicago, Senior Research Fellow with the Office of Survey Research at Michigan State University, and Adjunct Professor at the University of Illinois, Chicago.

Kim Borg was a Research Consultant at the Social Research Centre at the time this study was undertaken. Kim is now a Research Officer at BehaviourWorks Australia, Monash Sustainable Development Institute.

Abstract

The pervasiveness of the internet has led online research, and particularly online research undertaken via nonprobability online panels, to become the dominant mode of sampling and data collection used by the Australian market and social research industry.

There are broad-based concerns that the rapid increase in the use of nonprobability online panels in Australia has not been accompanied by an informed debate about the advantages and disadvantages of probability and nonprobability surveys.

The 2015 Australian Online Panels Benchmarking Study was undertaken to inform this debate, and report on the findings from a single national questionnaire administered across three different probability samples and five different nonprobability online panels.

This study enables us to investigate whether Australian surveys using probability sampling methods produce results different from Australian online surveys relying on nonprobability sampling methods, where accuracy is measured relative to independent population benchmarks. In doing so, we build on similar international research in this area, and discuss our findings as they relate to coverage error, nonresponse error, adjustment error and measurement error.

Acknowledgments

The Social Research Centre was the instigator of this research and the project leader. In bringing this project to fruition, the Social Research Centre would also like to acknowledge the research support and financial contribution of the Centre for Social Research & Methods located within the Research School of Social Sciences at the Australian National University.

The Social Research Centre is a wholly owned subsidiary of the Australian National University.

Acronyms

AAPOR	American Association for Public Opinion Research
ABS	Australian Bureau of Statistics
A-BS	address-based sampling
ANU	Australian National University
CATI	computer-assisted telephone interviewing
ISO	International Organization for Standardization
OPBS	Online Panels Benchmarking Study
RDD	random-digit dialling

Contents

1 Introduction

1.1 Background

Official statistics from the Australian Bureau of Statistics (ABS) show that, in 2014–15, 86% of Australian households could access the internet at home (ABS 2016), up from 67% in 2007–08 (ABS 2008). The volume of survey research undertaken via the internet has also increased. Since 2010, opt-in sampling[1] and online research have been the dominant modes of data collection in the Australian market and social research industry, supplanting computer-assisted telephone interviewing (CATI). In Australia in 2015, expenditure on opt-in online research was estimated at A$302 million, accounting for 41% of the revenue generated by the industry, up from 31% 2 years earlier (Research Industry Council of Australia 2016).

Around the world, online panels are now routinely used for gathering survey data for many purposes, including economic, political, public policy, marketing and health research. The ESOMAR 2015 Global Market Research Industry Report (ESOMAR 2015) estimated that, globally in 2014, US$10 billion was spent on online research, most of it on nonprobability online panels.

Web questionnaires, most of which are conducted via online panels, are a relatively recent development in the history of survey research. First used in the United States and Europe in the mid-1990s, they have since spread around the world.

From the mid-1990s to the mid-2000s, there was very rapid growth in the creation of online panels and increases in the sizes of the membership of such panels. This led to a proliferation of panel vendors. Since 2005, the developing need for panels with extremely large numbers of panellists has led to a consolidation of panel vendors through corporate acquisitions (Callegaro et al. 2014).

Most online panels are now established, and most of the people who participate in them are recruited, via nonprobability sampling methods (Baker et al. 2010).

In the United States and parts of Europe, the increased use of computer-assisted web interviewing (CAWI) resulted in establishment of probability-based online research panels from the late 1990s onwards[2] to enable rigorous and representative sampling of the population. In Australia, the first probability online panel, the Life in Australia panel, was not established until 2016.

Fine (2016) suggests that there are up to 50 panels purporting to be ‘research panels’ in Australia, but only seven have been accredited to the International Organization for Standardization (ISO) 26362 standard (‘Access panels in market, opinion and social research’). The oldest of the online panels now operating in Australia were established in the late 1990s.

1.2 Aims and objectives of the study

There are broad-based concerns that the rapid increase in the use of nonprobability online panels in Australia has not been accompanied by an informed debate about the advantages and disadvantages of probability and nonprobability surveys.

To inform this debate, we undertook the 2015 Australian Online Panels Benchmarking Study (OPBS). This paper reports on some of the initial findings of that study and discusses them from a Total Survey Error perspective.

The basic OPBS design involved administering the same questionnaire to eight samples, aiming to achieve approximately 600 completed questionnaires/interviews from each sample. The questionnaire was administered to three probability samples and five nonprobability samples drawn from the online panels operated by five independent nonprobability online panel providers. A dual-frame telephone sampling methodology was used for two of the probability surveys, and the third used an address-based sampling (A-BS) frame.

We explored whether the findings of Yeager et al. (2011) from the United States are replicated in Australia. The conclusions from their 2011 study were that:

- probability-based sample surveys recruited via telephone with telephone or internet data collection were consistently highly accurate across a set of demographics and nondemographics, especially after post-stratification with primary demographics
- nonprobability sample surveys done via the internet were less accurate, on average, than probability sample surveys
- there was considerable variation in accuracy among the findings of nonprobability samples, and much more so than among probability samples
- post-stratification with primary demographics sometimes improved the accuracy of nonprobability sample surveys and sometimes reduced their accuracy, so this method should not be relied upon to repair deficiencies in such samples.

Yeager et al. concluded that their results 'suggest caution before asserting or presuming that non-probability samples yield data that are as accurate or more accurate than data obtained from probability samples'.

2 Research methods

2.1 Overview

The in-scope population for the OPBS was people aged 18 years and older. The questionnaire administered to the eight samples used for this study – the Health, Wellbeing and Technology Questionnaire – was designed by researchers at the Social Research Centre, and included a wide range of demographic measures and questions about health, wellbeing and the use of technology.

We invited eight Australian online panel providers to submit a quotation to administer the questionnaire to members of their Australian nonprobability panels. The eight online providers were not randomly selected from all panel operators in Australia but were approached because they were known to the Social Research Centre either directly or by repute. Two did not submit a quotation by the given deadline and were excluded from further consideration. The six remaining companies were assessed against the following criteria:

- timeliness of correspondence
- sampling strategy
- survey optimisation across devices
- ability to deliver a clean SPSS data file with accompanying metadata and paradata (e.g. the number and date of invitations, bounce-backs, screen-outs, breakoffs, completion date, interview length)
- data cleaning rules that were to be applied
- ISO and ESOMAR accreditation.

The nonrandom selection of panel companies may limit the generalisability of the findings.

Of the five companies selected, four complied with all of the ESOMAR '28 questions to help buyers of online panels', and the other company partially complied with the ESOMAR requirements. Three of the five panels were ISO 26362 accredited. The price differential between the lowest and highest quotes was 24%.

Data collection for all eight iterations of the survey (three probability and five nonprobability) was undertaken between October and December 2015, with varying fieldwork periods designed to accommodate the requirements of each survey design or mode.

The Australian National University Human Research Ethics Committee oversaw ethical clearance for the conduct of this research (application number 2015/621).

The data file and accompanying documentation from this study are lodged with the Australian Data Archive.[3]

The names of the participating panel companies are not being publicly released, to preserve their anonymity.

2.2 The questionnaire

As far as possible, a standardised questionnaire was used across all samples. The questionnaire covered four broad topic areas:

1. *Primary demographics* – sex, age, location, educational attainment, country of birth and telephone status. The Social Research Centre typically uses these variables for post-stratification weighting.
2. *Secondary demographics* – Indigenous status, citizenship, enrolment to vote, geographic mobility, employment status, language spoken at home, home ownership status, volunteerism, household composition, wage and salary income, access to the internet at home, and socioeconomic disadvantage.

3. *Substantive measures* – general health status, psychological distress, life satisfaction, private health insurance coverage, daily smoking status and alcohol consumption in the past 12 months.
4. *Calibration variables* – early adopter questions, volunteerism[4] and use of information technology (accessing the internet, internet use and online survey participation).[5]

All the questions used to measure primary and secondary demographic characteristics, and the substantive items were adapted from high-quality Australian Government surveys. These items were chosen because high-quality population benchmarks are available for these measures. This is a critical part of the overall research design, because it enables the accuracy of the estimates derived from the various probability and nonprobability surveys to be compared with each other and with official population benchmarks.

The calibration measures were included to support further analyses (not included in this paper) of whether the biases associated with nonprobability online panels can be reduced by adjusting for known differences between probability surveys and nonprobability online panels. Results of this research will be released in a future paper in this series.

On average, the questionnaire took 6–11 minutes to complete for all modes other than the hard-copy mode, for which we did not capture completion time. Although questions were presented in as consistent a manner as possible, there were some minor differences in presentation to accommodate the various modes.

2.3 Probability-based surveys: sample design and recruitment

The three probability surveys all used different sampling designs.

Design 1 – random-digit dialling (RDD). This was a standalone dual-frame RDD telephone survey. The sampling frames were randomly generated landline and mobile phone numbers, with 50% of interviews to be completed via the landline frame and 50% via the mobile phone frame. For the landline frame, 15 probability-proportional-to-size geographic strata were established, based on the distribution of adults between capital cities and other cities, with the Australian Capital Territory being treated as a single stratum. For the landline sample, when there were two or more in-scope people in a household, the person with either the 'next birthday' or the 'most recent birthday' was randomly selected (Gaziano 2005). A single national stratum was used for the mobile frame because, in Australia, mobile phone numbers do not contain geographic markers. For the mobile phone sample, the respondent was the person invited to participate in the survey, provided they were in-scope.

Design 2 – address-based sampling. The sampling frame used for this survey was the Geocoded National Address File (G-NAF). G-NAF is maintained by the Public Sector Mapping Authority and is the authoritative national address index for Australia. G-NAF is compiled from existing and recognised address sources from state and territory government land records, as well as address data from Australia Post and the Australian Electoral Commission.[6]

The sample was selected from the G-NAF database using a stratified sample design in accordance with the distribution of the Australian residential population aged 18 years and over across the 15 geographic strata described above.

To accommodate situations in which more than one person in a household was in-scope, the printed instructions on the front of the questionnaire asked for the person aged 18 years or over with either 'next birthday' or 'most recent birthday' (alternating)

to complete the questionnaire. This wording was randomly assigned to sample records and printed questionnaires.

Design 3 – ANU Poll. Participants in this survey were recruited at the end of the ANU Poll, an established dual-frame RDD survey. Respondents who completed the October 2015 ANU Poll, which explored attitudes to ageing and money, were invited to take part in 'a future study about health and wellbeing'. Those who agreed to participate in the subsequent survey provided contact details. Depending on their preferences, these sample members were either emailed a link to complete the survey online or sent a hard-copy questionnaire to return via the mail.

The October ANU Poll used a dual-frame RDD sample design with a 60:40 split between landline and mobile phone interviews. The landline frame used the 15 geographic strata described above. For the mobile frame, a single national stratum was used, for the reason noted above. For the landline sample, when there were two or more in-scope people in a household, the 'next birthday' method was used to select the person to be invited to participate.

Of the 1200 respondents who completed the October 2015 ANU Poll and were invited to participate in a 'future study about health and wellbeing', 693 (58%) agreed and provided an email address and/or a physical address for distribution of the subsequent questionnaire.

2.4 Nonprobability surveys: sample design and recruitment

Recruitment to opt-in online panels is undertaken via several means, including a combination of online and offline nonprobability methods. Methods cited by the panel providers include banner advertising on websites, invitations and messaging, partnerships, print media, online marketing, direct mail, social media, referral programs and piggy-backing off CATI/CAPI (computer-assisted personal interviewing) surveys (Callegaro et al. 2014).

We asked the nonprobability panel providers approached for this study to conduct a 'nationally representative' survey of 600 respondents from their respective panels. We did not provide instructions about how this task should be carried out.

Four of the panel providers moved the age, sex and place of residence questions to the beginning of the questionnaire and used these as screening questions, even though these data (along with a great deal of other demographic, nondemographic, psychographic and other data) had already been collected as part of the initial recruitment and profiling activities undertaken to construct the panels. These screening questions allowed imposition of age, sex and geographic quotas, so that the sample reflected the distribution of these characteristics in the Australian adult population.

The remaining panel provider designed its sample to be an 'Australian Bureau of Statistics representative' sample, and applied quotas to the online survey, allowing for ±5% variation in the number of respondents per quota group. To determine how much sample to draw, this panel provider assumed a within-panel 20% response rate (based on average response rates for similar surveys).

Online panel providers invited their panel members to participate in a survey in several ways (see Table 1). All online panel providers used in this study approached panel members via an email to their personal email address. The common features of this invitation included a direct link to the questionnaire, a description of the length of the questionnaire, an incentive for completing the questionnaire, and the survey closing date. Two of the five panels also provided the survey title/topic. One of the providers recruited for multiple surveys at once, inviting panel members to respond to a variety of screening questions and directing them to one of the relevant questionnaires. Other methods of invitation included SMS, emails to panel members' panel accounts and social media.

Table 1 **Online panel recruitment methods**

Method	Panel 1	Panel 2	Panel 3	Panel 5
Banners	X			
Email invitations and messaging	X	X		
Partnerships	X	X	X	X
Print media		X	X	
Online marketing		X		
Direct mail		X	X	
Social media		X	X	
Other ad hoc initiatives		X		
Other survey methods (e.g. CATI, CAPI)			X	X
Referral programs			X	X

CAPI = computer-assisted personal interviewing; CATI = computer-assisted telephone interviewing
Note: This information was not available for panel 4.

2.5 Survey administration

Data collection for the various surveys occurred between October and December 2015. Standard response maximisation techniques were used for the probability surveys, including advance letters, incentives (contingent and noncontingent), several call/contact attempts, reminder mailings, choice of mode and refusal conversion. For the nonprobability panels, email invitations containing a direct link to the questionnaire were sent out by panel providers using their own software. Following their usual practices, panel companies offered panel members contingent incentives for completing the questionnaire.

Appendix A gives more detail about the administration of the various surveys.

2.6 Measures and analysis

2.6.1 Benchmarks

One of the key objectives of the OPBS was to determine the accuracy of the respective survey estimates relative to independent population benchmarks. The benchmarks used for these purposes are summarised in Table 2. Although not without error, all benchmarks are sourced from reputable national statistical collections and are of high quality.

Table 2 **Sources used for independent benchmark measures**

Demographics	Benchmark source
Primary demographics	
Sex	ABS, estimated resident population June 2015, cat. no. 3101.0
Age	ABS, estimated resident population June 2015, cat. no. 3101.0
Region and state	ABS, Census 2011, TableBuilder (2011); ABS, estimated resident population June 2015, cat. no. 3101.0
Educational attainment	ABS, Census 2011, TableBuilder (2011), highest level of school completed by QALLP (nonschool qualification): level of education by age, people aged 18 years and over, place of usual residence
Country of birth	ABS, TableBuilder (2011), country of birth – 4-digit level by age, people aged 18 years and over, place of usual residence

Table 2 Continued

Demographics	Benchmark source
Telephone status	Australian Communications and Media Authority (2015), *Communications report 2014–15* (ACMA 2015)
Secondary demographics	
Australian citizenship	ABS, Census 2011, TableBuilder (2011), Australian citizenship status by age, people aged 18 years and over, place of usual residence
Enrolled to vote	Australian Electoral Commission, 31 December 2015 (www.aec.gov.au/Enrolling_to_vote/Enrolment_stats/index.htm)
Indigenous status	ABS, Census 2011, TableBuilder (2011), Indigenous status by age, people aged 18 years and over, place of usual residence
Language other than English at home	ABS, Census 2011, TableBuilder (2011), language spoken at home – 2-digit level by age, people aged 18 years and over, place of usual residence
Geographic mobility	ABS, Census 2011, TableBuilder (2011), UAI5P by age, people aged 18 years and over, place of usual residence
Remoteness	ABS, Census 2011, TableBuilder (2011), remoteness area by age, people aged 18 years and over, place of usual residence
Employment status	ABS, Census 2011, TableBuilder (2011), employment type by age, people aged 18 years and over, place of usual residence
Wage and salary income	ABS, National Health Survey, 2014–15, people aged 18 years and over, employed income groups (ABS 2015b)
Household tenure	ABS, Census 2011, TableBuilder (2011), tenure type, dwellings: location on census night
Household composition	Australian Institute of Health and Welfare, National Drug Strategy Household Survey, 2013 (AIHW 2013)
Socioeconomic status	ABS, socioeconomic indexes for areas, 2011 (ABS 2013)
Access to internet at home	ABS, *Household use of information technology, Australia, 2014–15* (ABS 2016)
Volunteerism	ABS, Census 2011, TableBuilder (2011), voluntary work for an organisation or group by age, people aged 18 years and over, place of usual residence
Substantive measures	
Life satisfaction	ABS, General Social Survey, summary results Australia, 2014 (ABS 2014)
Psychological distress (Kessler 6)	ABS, National Health Survey, 2014–15, people aged 18 years and over, psychological distress, Australia (ABS 2015b)
General health	ABS, National Health Survey, 2014–15, people aged 18 years and over, self-assessed health status, Australia (ABS 2015b)
Private health insurance	ABS, National Health Survey, 2014–15, people aged 18 years and over, private health insurance, Australia (ABS 2015b)
Daily smoker	Australian Institute of Health and Welfare, National Drug Strategy Household Survey, 2013 (AIHW 2013)
Alcoholic drink of any kind in the past 12 months	Australian Institute of Health and Welfare, National Drug Strategy Household Survey, 2013 (AIHW 2013)

ABS = Australian Bureau of Statistics; UAI5P = address 5 years ago indicator

2.7 Weighting

For general population probability surveys, the usual approach to weighting consists of two steps:

1. Calculate design weights to allow for respondents having different chances of being selected to take part in the survey, such as the landline and mobile phone participants in RDD samples.
2. Adjust (calibrate) the design weights so that the resultant sample estimates align with known distributions of key sociodemographic characteristics. This accounts for different rates of response among sociodemographic subgroups. Benchmarks for the distribution of characteristics are generally derived from official statistics published by the ABS.

This approach to weighting is outlined in greater detail below.

2.7.1 Design weights

In probability sample surveys, the first step in weighting is to compute design weights as the inverse of the individual selection probabilities. Within a sample frame, an individual's probability of being selected is given as the number of people selected divided by the total number of people in the frame.

For the dual-frame surveys (the RDD survey and the ANU Poll), the chance of selection (p) is calculated by the following formula:

$$p = \frac{S_{LL} LL}{U_{LL} AD_{LL}} + \frac{S_{MP} MP}{U_{MP}}$$

where:

- S_{LL} is the number of survey respondents contacted by landline
- U_{LL} is the number of residential landline telephone numbers in Australia (estimated as 6 888 151[7])
- LL indicates the number of landlines in the respondent's household
- AD_{LL} is the number of in-scope adults in the respondent's household
- S_{MP} is the number of survey respondents contacted by mobile phone
- U_{MP} is the number of allocated mobile phone numbers in Australia (estimated as 19 590 306[8])
- MP indicates the presence of a mobile phone (0 for no, 1 for yes).

The $\frac{S}{U}$ terms can be thought of as the probabilities that the respondents' telephone numbers will be used. LL adjusts for the chance of selection for respondents living in households with multiple landlines, while AD adjusts for the possibility that the respondent will not be the one selected by the screening process. For the mobile phone frame, each respondent is assumed to own a single mobile device for their exclusive use (Lavrakas et al. 2017).[9]

The LL, AD_{LL} and MP data were obtained from respondents' answers to survey questions.

For the A-BS sample, a single-frame design weight was calculated. Because there is no need to adjust for overlapping sample frames and each household has an equal chance of selection into the survey (hence an address weight was not required), only the 'within household' chance of selection is accounted for in the weighting solution. Therefore, the design weight is equal to the number of adults in the household.

Because the probability of selection of the opt-in online nonprobability panels is unknowable, a design weight is not calculable. Therefore, a design weight of 1 was assigned to each nonprobability record.

2.7.2 Post-stratification weight

After the design weight was calculated, it was adjusted to try to reduce possible nonresponse and noncoverage error, giving a final weight (often called a post-stratification weight). For surveys where cross-classification of benchmarks is not considered necessary, the preferred approach to post-stratification weighting is iterative proportional fitting (sometimes called *rim weighting* or *raking*). This approach is preferred because it enables weights to be adjusted so that the weighted sample aligns with external population distributions for several categorical variables at once. This makes the weighted estimates reflect the population, not only with respect to those attributes commonly adjusted for, such as age, sex and geography, but also with respect to additional parameters such as educational attainment, birthplace and telephone status.

Cases with missing values on the weighting variables were treated as missing in rim weighting, so, for example, the weighted mobile-only proportion will be 29.0% of nonmissing cases for this variable (see Table B1 in Appendix B). Appendix B provides the population benchmarks used and their sources.

3 Results and analysis

3.1 Response analysis

3.1.1 Completion rate

Callegaro and DiSogra (2008) noted that a simple response metric for online surveys, including opt-in nonprobability panels, is the completion rate, defined as 'the proportion of those who completed the survey among all eligible panel members who were invited to take part'. This should not be taken to be the equivalent of a response rate.

Table 3 shows the completion rate for each of the probability surveys and four of the five nonprobability surveys. This statistic cannot be calculated for panel 1.

The completion rate for the nonprobability panels ranged from 2.6% to 15.4%. For the probability surveys, the completion rates ranged from 14.7% for the RDD survey to 80.8% for the ANU Poll (unadjusted). When adjusted to take into account the completion rate for the initial ANU Poll survey from which this sample was recruited (12.1%), the completion rate is 9.8% (80.8% × 12.1%). It is not possible to calculate a similar adjustment for the nonprobability panels because these panels do not recruit their members from a knowable sampling frame and thus have no true (or knowable) denominator to use in such calculations.

When comparing the completion rates for probability and nonprobability surveys, it is important to recall that, whereas the completion rate for probability-based surveys is calculated on the premise that all units in the frame have a known non-zero chance of being randomly included in the designated sample, the same cannot be said for the sampling methods used when creating nonprobability online panels. The nonresponse and noncoverage errors present at the sampling stages of such panels are unknowable but are likely to be considerable.

A comparison of the completion rates achieved in this study across the probability and nonprobability surveys does not support one of the main reasons often given for moving away from probability surveys in favour of nonprobability online panels: the supposedly relatively low sample yield achieved by probability surveys in comparison with nonprobability online panels.

Table 3 Completion rate by survey

	Probability samples			Nonprobability samples				
Measure	RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Number of invitations	4 097	2 050	693	na	7 097	4 097	6 132	23 527
Number of invitations opened	na	na	na	na	2 315	1 241	684	1 314
Number of completes	601	538	560	601	600	626	630	601
Completion rate (%)	14.7	26.2	9.8[a]	na	8.5	15.4	10.3	2.6
Time in field (days)	19	48	54	8	8	7	6	4

A-BS = address-based sampling; na = not applicable; RDD = random-digit dialling

a The initial completion rate for the ANU Poll was 12.1% (data not shown). The completion rate for the follow-up study was 80.8% (560 interviews from a designated sample of 693 respondents who agreed to take part in the subsequent survey). Therefore, the overall completion rate for the ANU Poll is 80.8% × 12.1% = 9.8%.

In addition to the sample yield, the American Association for Public Opinion Research (AAPOR) Response Rate 3 (RR3) was calculated for the probability surveys to enable comparisons with similar international studies. The respective response rates for each of the probability surveys are shown in Table 4.

Table 4 AAPOR response rates

AAPOR categories	RDD	A-BS	ANU Poll
Interview	601	538	560
Not eligible	98	5	0
Unusable[a]	381	0	3
Eligible, non-interview	311	130	79
Unknown eligibility, non-interview	2706	1424	0
Response Rate 3 (RR3) (%)	**17.9**	**26.5**	**12.4**[b]
Total records used	4097	2050	693

AAPOR = American Association for Public Opinion Research; A-BS= address-based sampling; RDD = random-digit dialling

a 'Unusable' call outcomes for members of the A-BS sample frame were treated as 'unknown eligibility non-interviews' for calculating RR3.

b The ANU Poll follow-up survey achieved an RR3 of 81.2%; however, this rate does not take into consideration the RR3 from the original iteration of the ANU Poll (15.3%). Therefore, the true RR3 for the Health, Wellbeing and Technology Survey recruited via the original ANU Poll is 81.2% × 15.3% = 12.4%.

3.2 Respondent profile

3.2.1 Respondent profile by survey

Table 5 provides an overview of selected respondent characteristics (unweighted) compared with population benchmarks for the three probability surveys and the five nonprobability surveys.

In comparing these data, it is important to note a fundamental difference in the sampling strategies used for the probability and nonprobability surveys. The approach taken by the nonprobability panels was to set age, sex and geographic quotas so that the completed interviews obtained from these panels reflected the distribution of the population for these characteristics. For the probability surveys, no age or sex quotas were set, and the geographic distribution of the achieved interviews broadly reflected the geographic stratification of these samples.

The Social Research Centre generally prefers to avoid the imposition of age and sex quotas, as adopting this approach introduces a nonprobability element into the sampling process, and does not allow the bias and variance properties of the sample to be known (Groves 1989).

Given these differences in sampling approaches, surveys showed various qualities:

- *Sex, age and location* – as expected given the quota controls imposed, the nonprobability surveys all produced estimates that very closely reflected the population distribution based on these characteristics. As is common, the unweighted estimates from the probability surveys overrepresent females, and people aged 55 years and over, and underrepresent those aged 18–34. The geographic composition of the probability surveys tended to align fairly closely with the nonprobability surveys and the population benchmarks. This is due to the stratified sample design adopted for the A-BS survey, and the imposition of proportionate geographic quotas for the landline component of the RDD survey and the original ANU Poll. The exception to the above is panel 2, which did not provide a postcode for 16% of records.
- *Educational attainment* – all the surveys overrepresented university graduates, the nonprobability panels to a lesser extent. All the samples underrepresented those who did not complete Year 12, the nonprobability panels to a much greater extent.
- *Birthplace* – the Australian born tended to be slightly overrepresented in all the surveys.
- *Telephone status* – all the surveys underrepresented the mobile-only population relative to benchmarks.

Table 5 **Respondent profile by survey, compared with benchmarks**

Demographic	Benchmark	Probability samples (%)			Nonprobability samples (%)				
		RDD	A-BS	ANU Poll	P1	P2	P3	P4	P5
		(*n* 601)	(*n* 538)	(*n* 560)	(*n* 601)	(*n* 600)	(*n* 626)	(*n* 630)	(*n* 601)
Sex									
Male	49.3	46	39	42	49	47	49	49	50
Female	50.7	54	61	58	51	53	51	51	50
Age (years)									
18–24	12.3	7	4	6	12	10	12	9	12
25–34	19.0	9	10	9	18	18	18	18	18
35–44	17.5	15	13	15	19	19	20	20	18
45–54	16.9	15	16	19	18	20	18	19	18
55–64	14.8	20	22	21	14	18	15	16	16
65–74	10.9	18	22	21	9	12	13	13	10
75+	8.4	14	13	9	8	2	5	5	8
Geography									
Major cities	70.2	69	73	69	76	62	68	78	76
Inner regional	18.4	19	19	21	18	17	21	17	18
Outer regional	9.0	8	7	9	5	5	10	5	6
Remote	1.4	1	1	1	*	*	*	*	*
Very remote	0.8	*	0	*	*	*	0	*	*
Missing	na	3	2	1	0	16	0	0	0
Education[a]									
Did not complete Year 12	45.9	35	35	36	24	27	27	23	27
Year 12, no bachelor's degree or higher	33.9	24	22	24	41	43	42	35	39
Bachelor's degree or higher	20.2	41	44	40	35	31	32	42	34
Birthplace									
Australia	68.6	75	73	75	76	76	75	69	75
Other	31.4	25	27	25	24	24	25	31	25
Telephone status									
Mobile only	29.0	18	17	12	17	19	17	19	16
Landline only	7.8	7	9	7	10	9	8	8	8
Dual user	63.2	75	73	81	71	71	73	70	76
No phone	na	0	1	0	2	2	2	2	1

* = <1%; A-BS = address-based sampling; na = not applicable; P = panel; RDD = random-digit dialling

a Benchmark figures for education have been re-based to exclude missing values.

3.3 Mode of data collection

Two of the eight surveys – the A-BS survey and the ANU Poll – allowed mixed modes of completion.

For the A-BS survey, which involved initially approaching sample members by mail, 39% of respondents completed the questionnaire online, and a similar proportion (38%) mailed the completed questionnaire back in the reply-paid envelope provided. Almost a quarter (24%) completed the questionnaire in response to outbound telephone reminders.

Whereas hard copy and online were equally popular for the A-BS survey, online was the preferred mode of completion for sample members recruited via the ANU Poll (52%), followed by telephone (41%) and hard copy (7%).

It is important to note that survey design features most likely had more influence on the mode of response chosen by sample members than any underlying questionnaire completion preference. For both multimode surveys, telephone was only available if sample members had not responded online or via hard copy. Also, only 94 of the 693 ANU Poll sample members who agreed to participate in the survey requested that they be sent a hard-copy questionnaire (sent with a reply-paid envelope).

Appendix C profiles respondents to the multimode surveys by mode of completion.

3.4 Size and profile of the offline population

As noted, in Australia in 2014–15, 86% of households reportedly had access to the internet at home (ABS 2016). One of the criticisms of nonprobability online panels is that they systematically exclude the offline population, leaving open the possibility of noncoverage error.

In this context, it is interesting to compare the size and profile of the offline population as reflected in the composition of the three probability surveys undertaken as part of this study. To do this, we combined the three probability samples and looked at the online/offline distribution of the combined samples. In this study, the offline population is defined as people who were not able to access the internet at home via a broadband connection, a dial-up connection or in some other way, including through mobile phones or some other mobile device.[10] The fraction of the population so classified in the probability surveys is 9% unweighted and 8% weighted. This suggests that the offline population is underrepresented in these surveys relative to the ABS benchmarks (14%).

Table 6 shows that a higher proportion of females are offline (11%) compared with males (7%). The distribution of the combined samples by age group shows some marked differences, with only 1% of those aged less than 45 years offline compared with 14% of those aged 65–74 years and 33% of those aged 75 years and older. In terms of educational attainment, 18% of those without Year 12 education were offline compared with just 2% of university graduates. A large fraction (40%) of landline-only respondents were offline.

Table 6 **Online/offline distribution of the probability samples (unweighted)**

		Probability surveys	
Demographic	Base (*n*)	Online (%)	Offline (%)
Total	1699	91	9
Sex			
Male	717	93	7
Female	982	89	11
Age (years)			
18–24	97	99	1
25–34	158	99	1
35–44	239	99	1
45–54	282	95	5
55–64	360	95	5
65–74	344	86	14
75+	207	67	33
Geography[a]			
Major cities	1193	92	8
Inner regional	331	89	11
Outer regional	139	88	12
Remote	15	100	0
Very remote	3	100	0
Education[a]			
Did not obtain Year 12	588	82	18
Obtained Year 12	392	92	8
Bachelor's degree or higher	695	98	2
Birthplace			
Australia	1265	91	9
Other	434	92	8
Telephone status			
Mobile only	267	94	6
Landline only	126	60	40
Dual user	1298	93	7
No phone	8	88	12

a Bases exclude missing values.

3.5 Survey results

As discussed earlier, a major aim of this study was to learn whether the findings of Yeager et al. (2011) apply in Australia. Accordingly, the analytical method we adopted closely follows the method used by Yeager et al. (2011). The following comparisons were made to identify the between-survey differences in estimates and benchmarks for secondary demographics and substantive measures:

- *Secondary demographics* – unweighted and weighted survey estimates of the modal response category compared with the corresponding benchmark (Table 7).
- *Substantive measures* – unweighted and weighted survey estimates of the modal response category compared with the corresponding benchmark (Table 8).
- *Average absolute error for secondary demographics* – defined as the percentage point deviation from the benchmark between unweighted and weighted survey estimates of the modal response category and the corresponding benchmark averaged across secondary demographics (Table 9).
- *Average absolute error for substantive measures* – defined as the percentage point deviation from the benchmark between unweighted and weighted survey estimates of the modal response category and the corresponding benchmark averaged across substantive measures (Table 10).

Following Yeager et al. (2011), standard errors of survey estimates and standard errors of average absolute errors were calculated using a bootstrapping procedure. The bootstrapping procedure is an accepted method for estimation of the sampling distribution of any statistics, including sampling errors of probability samples (Baker et al. 2013), and was implemented in the R package 'boot' (Davison & Hinkley 1997, Canty & Ripley 2015). Probability methods should not be used to estimate sampling errors of nonprobability samples because nonprobability samples violate key assumptions of sampling theory. Although there is no universally agreed method to estimate sampling errors of nonprobability samples, the AAPOR Taskforce Report on Non-probability

Sampling (Baker et al. 2013) cites bootstrapping (or resampling) as one of the acceptable methodologies for reporting the precision of nonprobability-based estimates.

The *boot* function in the R package replicates a statistic of interest (e.g. proportion of respondents in the sample in a response category) by randomly selecting observations, with replacement from the dataset to produce a sample of equal size to the original sample. When calculating weighted estimates using this procedure, a unique set of post-stratification weights is calculated for each random sample (based on the observations selected in the sample) to force each bootstrap sample to match the population in terms of primary demographics used in weighting. This is repeated 150 times, and 150 estimates of the statistic of interest are then used to calculate an estimate of the standard error for each of the measures.[11]

Benchmarks sourced from the Australian 2011 Census and the Australian Electoral Commission do not have sampling errors associated with them because these are not sample surveys. Standard errors of the remaining benchmarks were acquired directly from the government agency that conducted the survey or calculated from the survey data using weights and information about sample design provided by the government agency. Standard errors for survey-based benchmarks are relatively small (standard errors for all but two measures were less than 1% of the benchmark estimate, and the remaining two were less than 5% of the benchmark estimate).

Independent-samples *t*-tests were conducted to test the null hypotheses of:

- no differences between survey estimates and benchmarks (Tables 7 and 8)
- no differences between the average of the absolute errors for each pair of surveys included in the study (Tables 9 and 10).

In addition, following Yeager et al. (2011), the following summary measures were calculated to illustrate overall accuracy of the eight surveys considered (Table 11):

- ranking of average absolute errors across all eight surveys, with the smallest average absolute error ranked as top
- number of significant differences from benchmarks for each survey
- largest percentage point absolute error for each survey.

We summarise results in the next section.

Table 7 **Survey estimates of modal response category and the corresponding benchmark for secondary demographics**

Secondary demographics benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Indigenous status (non-Indigenous)	98.10								
Unweighted									
Estimate		98.84	97.96	98.39	97.50	96.50*	98.40	98.41	98.84
Percentage point error		0.74	−0.14	0.29	−0.60	−1.60	0.30	0.31	0.74
Weighted									
Estimate		98.76	98.40	98.42	97.90	96.49	98.09	98.27	98.83
Percentage point error		0.66	0.30	0.32	−0.20	−1.61	0.00	0.17	0.73
Australian citizen	83.93								
Unweighted									
Estimate		91.01***	94.42***	92.32***	93.01***	90.50***	93.13***	90.63***	94.68***
Percentage point error		7.08	10.49	8.39	9.08	6.57	9.20	6.70	10.75
Weighted									
Estimate		86.60	92.00***	86.56	91.81***	88.05*	91.04***	90.76***	92.95***
Percentage point error		2.67	8.07	2.63	7.88	4.12	7.11	6.83	9.02
Enrolled to vote	78.47								
Unweighted									
Estimate		88.19***	92.57***	90.36***	86.86***	86.00***	88.50***	86.83***	91.51***
Percentage point error		9.71	14.09	11.88	8.38	7.53	10.02	8.35	13.04
Weighted									
Estimate		83.06*	88.68***	83.02	84.95***	80.18	85.59***	84.75***	89.21***
Percentage point error		4.59	10.21	4.55	6.47	1.70	7.11	6.28	10.74

Continued next page

Table 7 **Continued**

Secondary demographics benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Living at current address 5 years ago	54.80								
Unweighted									
Estimate		69.55***	69.14***	67.50***	61.56***	61.00**	64.22***	63.81***	68.55***
Percentage point error		14.76	14.35	12.70	6.77	6.20	9.42	9.01	13.76
Weighted									
Estimate		62.10**	54.68	58.44	59.79*	58.12	61.80***	63.66***	65.89***
Percentage point error		7.30	–0.11	3.65	4.99	3.32	7.00	8.86	11.10
Currently employed	59.39								
Unweighted									
Estimate		58.24	57.43	60.54	51.08***	54.33*	53.99**	55.71	50.25***
Percentage point error		–1.16	–1.96	1.14	–8.31	–5.06	–5.40	–3.68	–9.14
Weighted									
Estimate		69.34***	64.60	66.43**	49.14***	53.33*	53.11**	51.00***	49.04***
Percentage point error		9.95	5.20	7.04	–10.25	–6.06	–6.28	–8.39	–10.35
Voluntary work (no)	74.22								
Unweighted									
Estimate		58.24***	60.78***	60.18***	72.55	73.83	71.09	68.89**	71.05
Percentage point error		–15.98	–13.44	–14.04	–1.68	–0.39	–3.13	–5.33	–3.17
Weighted									
Estimate		62.65***	62.99***	62.56***	74.51	77.14	71.46	69.86	70.65
Percentage point error		–11.57	–11.24	–11.66	0.29	2.92	–2.76	–4.36	–3.57

Continued next page

Table 7 **Continued**

Secondary demographics benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Language other than English (no)	75.72								
Unweighted									
Estimate		84.19***	81.23**	86.96***	82.70***	84.17***	85.62***	80.00**	84.03***
Percentage point error		8.48	5.51	11.25	6.98	8.45	9.91	4.28	8.31
Weighted									
Estimate		85.45***	80.38	84.47***	85.09***	85.37***	87.53***	84.75***	85.30***
Percentage point error		9.73	4.66	8.75	9.38	9.66	11.81	9.03	9.58
Most disadvantaged area–based socioeconomic status (quintile)	20.00								
Unweighted									
Estimate		14.98***	14.50***	11.79***	16.81*	14.67***	14.38***	13.49***	13.81***
Percentage point error		–5.02	–5.50	–8.21	–3.19	–5.33	–5.62	–6.51	–6.19
Weighted									
Estimate		13.76**	15.08*	10.27***	16.97	14.75**	14.85***	14.52***	14.14***
Percentage point error		–6.24	–4.92	–9.73	–3.03	–5.25	–5.15	–5.48	–5.86
Resident of a major city	70.22								
Unweighted									
Estimate		69.05	72.68	69.11	76.04***	61.83***	68.05	77.30***	75.37**
Percentage point error		–1.17	2.45	–1.12	5.82	–8.39	–2.17	7.08	5.15
Weighted									
Estimate		69.03	72.92	69.76	73.15*	69.88	68.31	72.63	71.81
Percentage point error		–1.19	2.69	–0.46	2.93	–0.34	–1.91	2.41	1.59

Continued next page

Table 7 **Continued**

Secondary demographics benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Access the internet from home	85.90								
Unweighted									
Estimate		86.86	87.92	92.50***	99.00***	98.67***	98.72***	99.21***	99.67***
Percentage point error		0.96	2.02	6.60	13.10	12.77	12.82	13.31	13.77
Weighted									
Estimate		89.57*	91.81***	93.07***	98.76***	98.24***	97.91***	99.05***	99.60***
Percentage point error		3.67	5.91	7.17	12.86	12.34	12.01	13.15	13.70
Home ownership with a mortgage	29.61								
Unweighted									
Estimate		30.95	32.34	33.57	31.78	30.17	33.87*	33.81*	31.61
Percentage point error		1.34	2.73	3.96	2.17	0.56	4.26	4.20	2.01
Weighted									
Estimate		33.75	39.96***	37.40**	28.56	30.92	30.41	29.91	28.00
Percentage point error		4.14	10.35	7.79	–1.05	1.31	0.80	0.30	–1.61
Couple with dependent children	38.35								
Unweighted									
Estimate		22.80***	21.00***	23.39***	26.79***	25.83***	27.00***	29.21***	29.95***
Percentage point error		–15.55	–17.34	–14.95	–11.56	–12.51	–11.35	–9.14	–8.40
Weighted									
Estimate		27.90***	28.19***	26.97***	23.12***	23.97***	24.62***	25.46***	28.11***
Percentage point error		–10.45	–10.15	–11.38	–15.22	–14.38	–13.73	–12.89	–10.23

Continued next page

Table 7 **Continued**

Secondary demographics benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Wage and salary income $1000–1249 per week	13.80								
Unweighted									
Estimate		9.97	14.14	14.33	9.76	12.06	13.17	14.54	11.76
Percentage point error		–3.83	0.34	0.53	–4.04	–1.74	–0.63	0.74	–2.04
Weighted									
Estimate		11.78	12.81	15.04	9.73	12.77	12.96	15.93	12.90
Percentage point error		–2.02	–0.99	1.24	–4.07	–1.03	–0.84	2.13	–0.90

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; A-BS = address-based sampling; RDD = random-digit dialling

Note: All errors are deviations from the benchmark.

Table 8 **Survey estimates of modal response category and the corresponding benchmark for substantive measures**

Substantive measures benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Life satisfaction (8 out of 10)	32.60								
Unweighted									
Estimate		34.61	30.11	31.25	21.80***	20.17***	27.48*	23.81***	24.63**
Percentage point error		2.01	–2.49	–1.35	–10.80	–12.43	–5.12	–8.79	–7.97
Weighted									
Estimate		34.50	30.58	30.60	20.67***	21.03***	28.11	23.38***	24.72**
Percentage point error		1.90	–2.02	–2.00	–11.93	–11.57	–4.49	–9.22	–7.88
Psychological distress – Kessler 6 (low)	82.20								
Unweighted									
Estimate		73.97***	76.05**	75.59**	54.50***	56.88***	59.27***	59.21***	58.72***
Percentage point error		–8.23	–6.15	–6.61	–27.70	–25.32	–22.93	–22.99	–23.48
Weighted									
Estimate		74.12***	71.61***	70.63***	56.34***	58.68***	60.00***	57.24***	59.00***
Percentage point error		–8.08	–10.59	–11.57	–25.86	–23.52	–22.20	–24.96	–23.20
General health status (SF1) (very good)	36.20								
Unweighted									
Estimate		30.62	34.39	33.75	33.28	32.67	32.59	32.38	36.94
Percentage point error		–5.58	–1.81	–2.45	–2.92	–3.53	–3.61	–3.82	0.74
Weighted									
Estimate		33.55	36.55	34.20	32.06	30.36*	30.89	31.24	37.73
Percentage point error		–2.65	0.35	–2.00	–4.14	–5.84	–5.31	–4.96	1.53

Continued next page

Table 8 **Continued**

Substantive measures benchmark comparison	Value	Probability sample surveys			Nonprobability sample internet surveys				
		RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Private health insurance	57.10								
Unweighted									
Estimate		65.56**	67.29**	65.18*	53.08	49.00**	53.35	59.52	58.57
Percentage point error		8.46	10.19	8.08	−4.02	−8.10	−3.75	2.42	1.47
Weighted									
Estimate		60.35	60.48	59.05	48.22**	44.59***	53.42	56.46	54.54
Percentage point error		3.25	3.38	1.95	−8.88	−12.51	−3.68	−0.64	−2.56
Daily smoker	13.52								
Unweighted									
Estimate		10.32*	9.11**	12.50	21.80***	20.17***	17.25**	14.76	15.64
Percentage point error		−3.21	−4.41	−1.02	8.28	6.64	3.73	1.24	2.12
Weighted									
Estimate		15.12	9.37**	17.03*	23.33***	20.21***	17.41**	16.19	17.84**
Percentage point error		1.60	−4.16	3.51	9.81	6.69	3.89	2.67	4.32
Consumed alcohol in the past 12 months	81.87								
Unweighted									
Estimate		82.20	82.53	84.46	79.53	75.83***	77.32**	77.94*	79.20
Percentage point error		0.32	0.65	2.59	−2.34	−6.04	−4.56	−3.94	−2.67
Weighted									
Estimate		85.87*	85.48*	84.75	79.49	76.61**	77.99*	77.66*	80.38
Percentage point error		4.00	3.60	2.87	−2.39	−5.26	−3.89	−4.21	−1.50

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; A-BS = address-based sampling; RDD = random-digit dialling
Note: All errors are deviations from the benchmark.

Table 9 **Pairwise *t*-tests comparing average absolute errors on secondary demographics using bootstrapped standard errors**

Secondary demographics survey	Probability sample surveys			Nonprobability sample internet surveys				
	RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Unweighted								
Average absolute error	6.5983	6.9525	7.3141	6.2824	5.9303	6.4800	6.0502	7.4193
Pairwise differences								
RDD	na	−0.35	−0.72	0.32	0.67	0.12	0.55	−0.82
A-BS	0.35	na	−0.36	0.67	1.02	0.47	0.9	−0.47
ANU Poll	0.72	0.36	na	1.03	1.38*	0.83	1.26	−0.11
Panel 1	−0.32	−0.67	−1.03	na	0.35	−0.2	0.23	−1.14
Panel 2	−0.67	−1.02	−1.38*	−0.35	na	−0.55	−0.12	−1.49*
Panel 3	−0.12	−0.47	−0.83	0.20	0.55	na	0.43	−0.94
Panel 4	−0.55	−0.90	−1.26	−0.23	0.12	−0.43	na	−1.37*
Panel 5	0.82	0.47	0.11	1.14	1.49*	0.94	1.37*	na
Weighted								
Average absolute error	5.7060	5.7536	5.8744	6.0485	4.9264	5.8871	6.1754	6.8443
Pairwise differences								
RDD	na	−0.05	−0.17	−0.34	0.78	−0.18	−0.47	−1.14
A-BS	0.05	na	−0.12	−0.29	0.83	−0.13	−0.42	−1.09
ANU Poll	0.17	0.12	na	−0.17	0.95	−0.01	−0.30	−0.97
Panel 1	0.34	0.29	0.17	na	1.12	0.16	−0.13	−0.80
Panel 2	−0.78	−0.83	−0.95	−1.12	na	−0.96	−1.25	−1.92**
Panel 3	0.18	0.13	0.01	−0.16	0.96	na	−0.29	−0.96
Panel 4	0.47	0.42	0.30	0.13	1.25	0.29	na	−0.67
Panel 5	1.14	1.09	0.97	0.80	1.92**	0.96	0.67	na

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; A-BS = address-based sampling; na = not applicable; RDD = random-digit dialling

Table 10 **Pairwise *t*-tests comparing average absolute errors on substantive measures using bootstrapped standard errors**

	Probability sample surveys			Nonprobability sample internet surveys				
Substantive measures survey	**RDD**	**A-BS**	**ANU Poll**	**Panel 1**	**Panel 2**	**Panel 3**	**Panel 4**	**Panel 5**
Unweighted								
Average absolute error	4.6345	4.2839	3.6833	9.3435	10.3453	7.2839	7.2006	6.4079
Pairwise differences								
RDD	na	0.35	0.95	−4.71***	−5.71***	−2.65*	−2.57**	−1.77*
A-BS	−0.35	na	0.60	−5.06***	−6.06***	−3.00**	−2.92**	−2.12*
ANU Poll	−0.95	−0.60	na	−5.66***	−6.66***	−3.60**	−3.52***	−2.72**
Panel 1	4.71***	5.06***	5.66***	na	−1	2.06	2.14	2.94**
Panel 2	5.71***	6.06***	6.66***	1	na	3.06*	3.14**	3.94***
Panel 3	2.65*	3.00**	3.60**	−2.06	−3.06*	na	0.08	0.88
Panel 4	2.57**	2.92**	3.52***	−2.14	−3.14**	−0.08	na	0.79
Panel 5	1.77*	2.12*	2.72**	−2.94**	−3.94***	−0.88	−0.79	na
Weighted								
Average absolute error	3.5775	4.0163	3.9817	10.5021	10.8984	7.2430	7.7786	6.8304
Pairwise differences								
RDD	na	−0.44	−0.40	−6.92***	−7.32***	−3.67**	−4.20**	−3.25**
A-BS	0.44	na	0.03	−6.49***	−6.88***	−3.23*	−3.76**	−2.81**
ANU Poll	0.40	−0.03	na	−6.52***	−6.92***	−3.26*	−3.80**	−2.85*
Panel 1	6.92***	6.49***	6.52***	na	−0.40	3.26*	2.72*	3.67***
Panel 2	7.32***	6.88***	6.92***	0.40	na	3.66*	3.12*	4.07**
Panel 3	3.67**	3.23*	3.26*	−3.26*	−3.66*	na	−0.54	0.41
Panel 4	4.20**	3.76**	3.80**	−2.72*	−3.12*	0.54	na	0.95
Panel 5	3.25**	2.81**	2.85*	−3.67***	−4.07**	−0.41	−0.95	na

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; A-BS = address-based sampling; na = not applicable; RDD = random-digit dialling

Table 11 **Overall accuracy metrics**

	Probability sample surveys			Nonprobability sample internet surveys				
Summary metric	RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Average absolute error								
Secondary demographics								
Unweighted	6.60	6.95	7.31	6.28	5.93	6.48	6.05	7.42
Weighted	5.71	5.75	5.87	6.05	4.93	5.89	6.18	6.84
Substantive measures								
Unweighted	4.63	4.28	3.68	9.34	10.35	7.28	7.20	6.41
Weighted	3.58	4.02	3.98	10.50	10.90	7.24	7.78	6.83
Rank: average absolute error								
Secondary demographics								
Unweighted	5	6	7	3	1	4	2	8
Weighted	2	3	4	6	1	5	7	8
Substantive measures								
Unweighted	3	2	1	7	8	6	5	4
Weighted	1	3	2	7	8	5	6	4
Number of significant differences from benchmarks at 0.05[a]								
Secondary demographics (out of 13)								
Unweighted	6	6	5	4	3	4	3	4
Weighted	8	7	7	8	6	8	8	8

Continued next page

Table 11 **Continued**

Summary metric	Probability sample surveys			Nonprobability sample internet surveys				
	RDD	A-BS	ANU Poll	Panel 1	Panel 2	Panel 3	Panel 4	Panel 5
Substantive measures (out of 6)								
Unweighted	3	3	2	3	5	4	3	2
Weighted	2	3	2	4	6	3	3	3
Largest percentage point absolute error								
Secondary demographics								
Unweighted	15.98	17.34	14.95	13.10	12.77	12.82	13.31	13.77
Weighted	11.24	11.66	11.57	15.22	14.38	13.73	13.15	13.70
Substantive measures								
Unweighted	8.46	10.19	8.08	27.70	25.32	22.93	22.99	23.48
Weighted	8.08	10.59	11.57	25.86	23.52	22.20	24.96	23.20

A-BS = address-based sampling; RDD = random-digit dialling

a To provide comparability with earlier results reported by Yeager et al. (2011), differences significant at the 0.05 level are counted for this summary measure.

3.6 Detailed analysis

3.6.1 Primary demographics

As part of weighting, probability and nonprobability samples were matched to the population distributions of primary demographics. Accordingly, weighted comparisons are not meaningful for primary demographics.

Comparison of unweighted primary demographics is also of limited interest because nonprobability sample providers use these demographics to set quotas for the sample, effectively forcing the sample distribution to mirror the population.

As will be seen from the comparison of secondary demographics and substantive measures, imposing population distributions of primary demographics on a sample does not guarantee a representative sample or accuracy of the estimates.

3.6.2 Secondary demographics

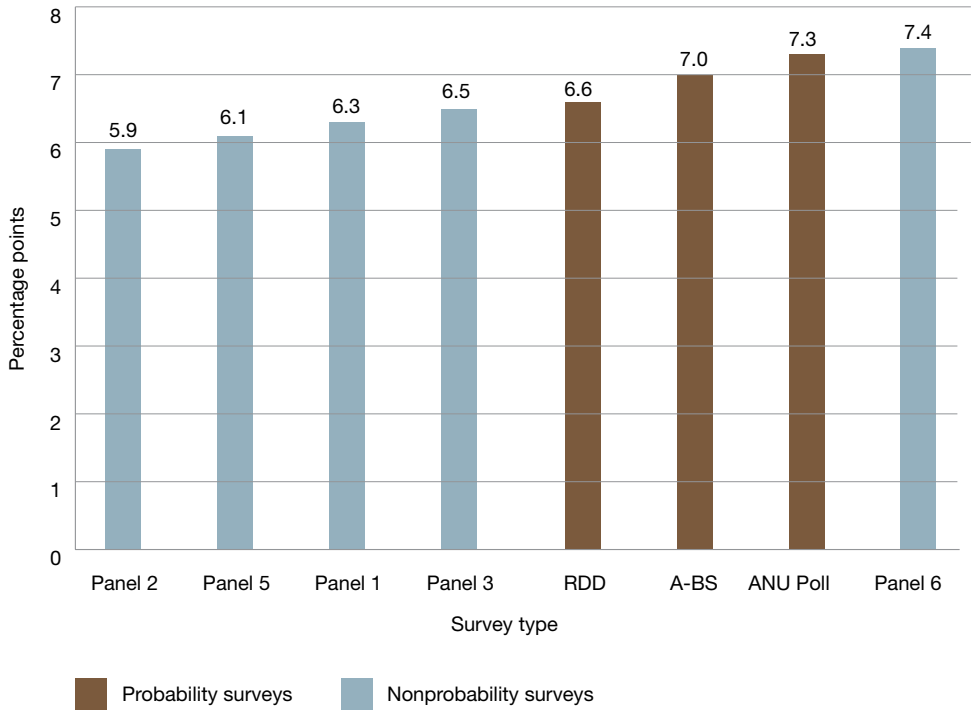
Unweighted

As shown in Figure 1, all but one of the nonprobability panels' unweighted estimates of secondary demographics were closer to the benchmarks than the probability surveys. The probability surveys rank fifth, sixth and seventh out of eight in terms of absolute average error. Reference to Table 11 shows that the probability surveys also reported the largest percentage point absolute error. This also shows that, in terms of unweighted estimates of secondary demographics, the probability surveys differ significantly from the benchmarks on five or six measures (out of 13), whereas the nonprobability panels differ significantly from the benchmarks on three or four measures.

However, when comparing average absolute errors across surveys (Table 10), there was only one significant difference between probability and nonprobability surveys. Only the ANU Poll's unweighted absolute average error was significantly different ($p < 0.05$) from the best nonprobability survey (panel 2).

As quotas were not applied to probability surveys, and the weighting is necessary to account for chances of selection and to align with known distributions of key geodemographic characteristics, these results in relation to the secondary benchmarks are not unexpected.

Figure 1 **Average absolute errors: secondary demographics (unweighted)**



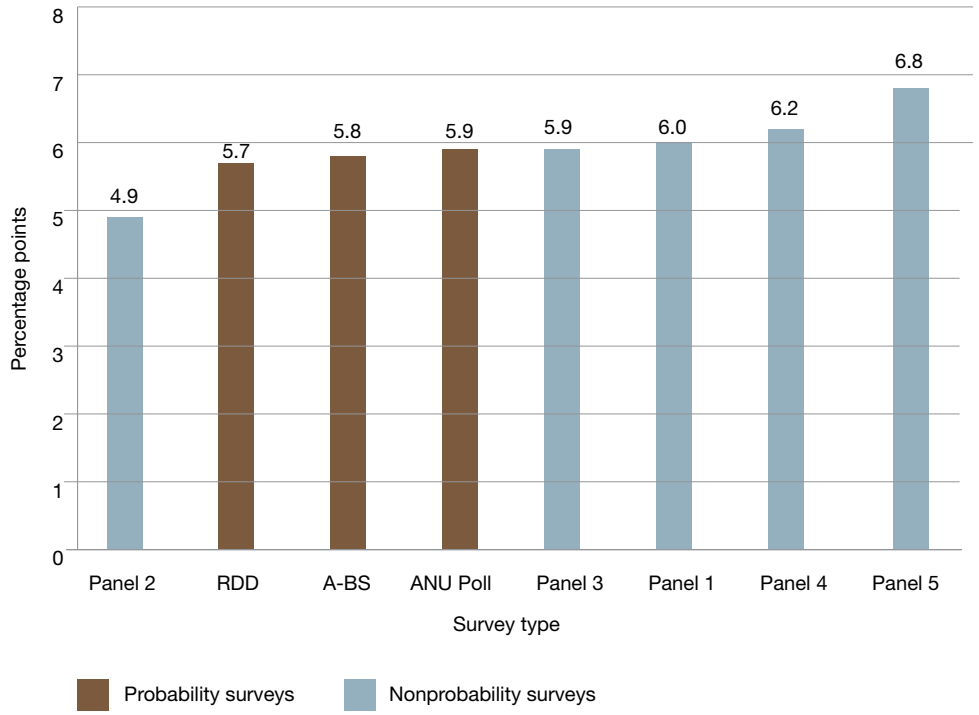
A-BS = address-based sampling; RDD = random-digit dialling

Weighted

As shown in Figure 2 and, as expected, weighting improves the accuracy of the probability survey estimates relative to secondary benchmarks and brings the accuracy of the probability surveys more in line with that of the nonprobability surveys. The average absolute error for the weighted probability survey estimates of secondary demographics ranges from 5.7 to 5.9 percentage points; for the nonprobability panels, it ranges from 4.9 to 6.8 percentage points.

There are no significant differences between weighted probability and nonprobability surveys' average absolute errors with respect to secondary demographics. Although weighting improves the accuracy across the board, panel 2 remains the most accurate of all surveys.

Figure 2 **Average absolute errors: secondary demographics (weighted)**



A-BS = address-based sampling; RDD = random-digit dialling

3.6.3 Substantive measures

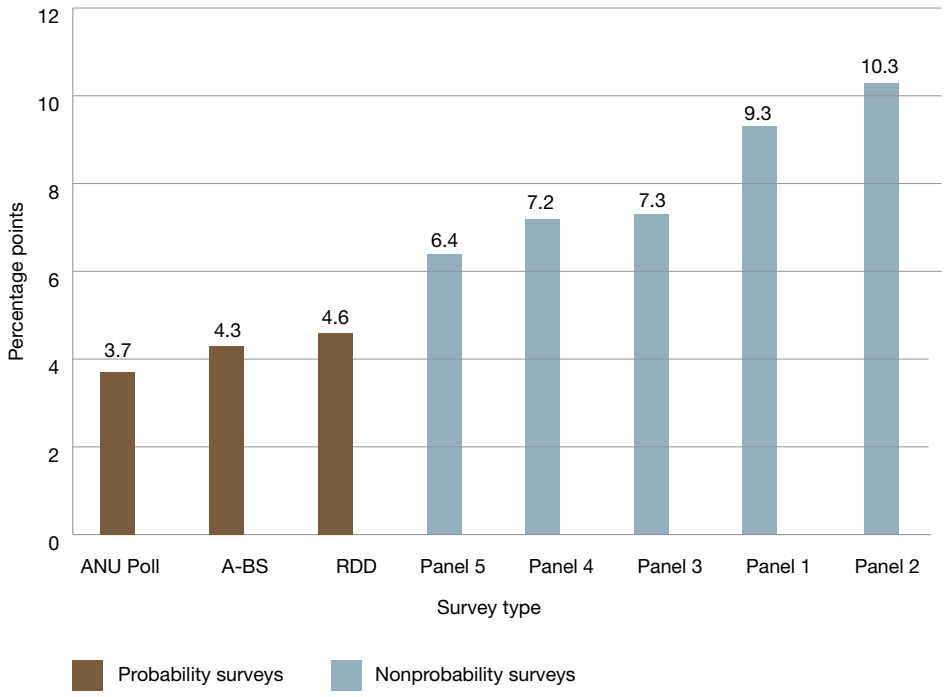
For the substantive measures, the probability surveys are consistently more accurate when comparing both unweighted and weighted data. This supports the findings of Kennedy et al. (2016), which show that balancing the sample on demographic variables (as is the case for the nonprobability panels) is no guarantee of accurate measurement of the outcome variables of interest.

Unweighted

There were no significant differences in average absolute errors across unweighted estimates of substantive measures among the probability surveys. The ANU Poll showed the smallest average absolute error (3.7 percentage points) and RDD the largest (4.6 percentage points; see Figure 3).

The average absolute errors of unweighted nonprobability surveys were almost double those of the probability surveys (ranging from 6.4 percentage points for panel 5 to 10.3 percentage points for panel 2) and significantly different from all probability surveys. Importantly, panel 2 was the best performing sample on demographics variables and the worst performing sample on the substantive variables for both weighted and unweighted data (see Figures 3 and 4). This again highlights the danger of relying on a 'demographically balanced' nonprobability sample to provide accurate measurement on substantive measures of interest.

Figure 3 Average absolute errors: substantive measures (unweighted)



A-BS = address-based sampling; RDD = random-digit dialling

Weighted

Like unweighted data, there were no significant differences in average absolute errors across probability surveys for the weighted estimates of the substantive measures (Figure 4). The standalone RDD survey, when weighted, had the lowest average absolute error at 3.6 percentage points. The biggest improvement as a result of weighting was achieved for the RDD estimates. ANU Poll estimates deteriorated slightly – from 3.7 unweighted average absolute error to 4.0 weighted average absolute error (in percentage points).

Weighting caused the average absolute error to increase for all nonprobability surveys except panel 3, which was largely unchanged (from 7.3 to 7.2 percentage points). Panel 1 recorded the largest increase in average absolute average error attributable to weighting, increasing from 9.3 percentage points (unweighted) to 10.5 percentage points (weighted).

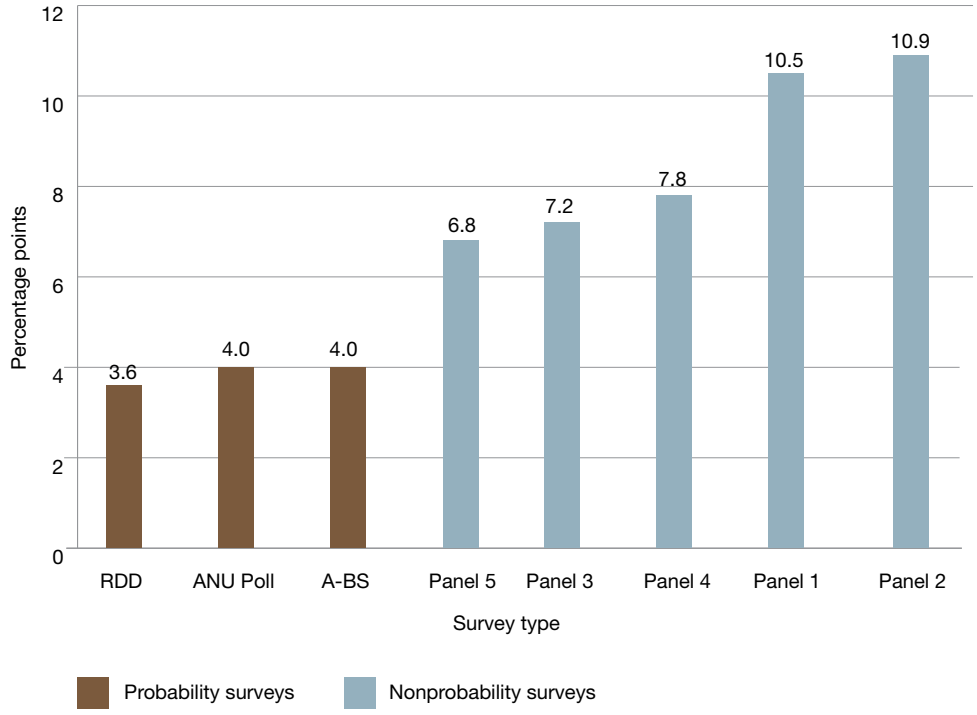
Weighting brought the probability samples closer together, reducing the difference in average absolute error between them. Weighting had the opposite effect on nonprobability samples, slightly increasing the range of average absolute errors. Weighting also increased the largest absolute error for A-BS, the ANU Poll and panel 4 (Table 11).

3.6.4 Variance/inconsistency of absolute errors across surveys

In line with the findings of Yeager et al. (2011) and others (e.g. Chang & Krosnick 2009, Walker et al. 2009), probability surveys were more consistent in their measurement of both secondary demographics and substantive measures.

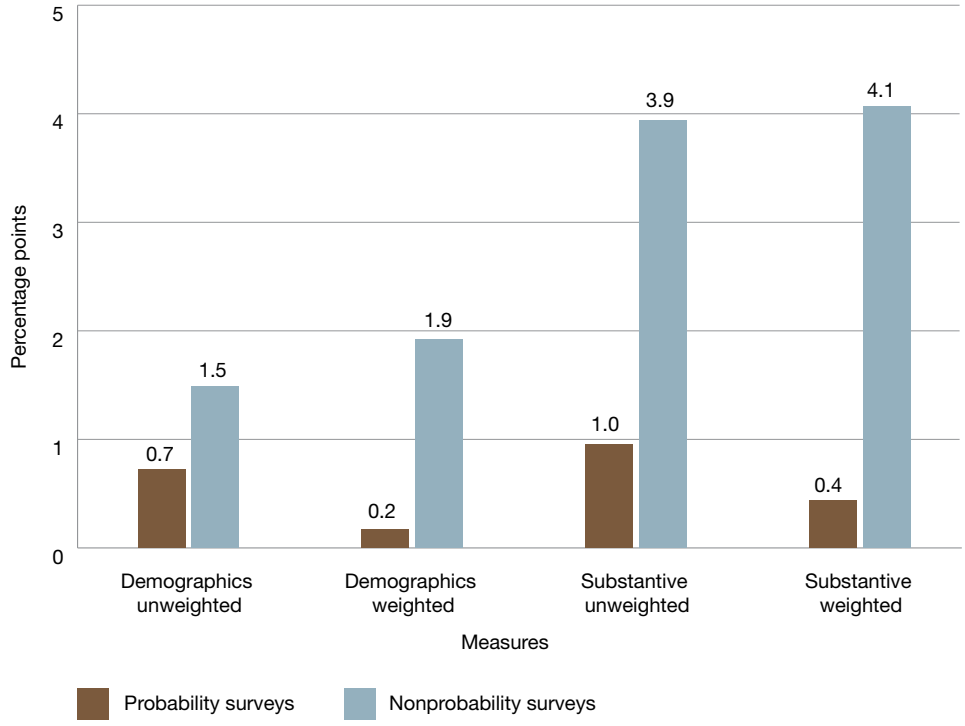
As shown in Figure 5, without weighting, the average absolute error for the three probability surveys had a range of 0.7 percentage points among secondary demographics and 1.0 percentage points among substantive measures, whereas the five nonprobability panels had corresponding unweighted ranges that were more than twice as wide at 1.5 and 3.9 percentage points, respectively. Corresponding ranges for weighted data were 0.2 and 0.4 percentage points for probability survey demographics and substantive measures, respectively, and much wider at 1.9 and 4.1 percentage points for nonprobability survey demographics and substantive measures, respectively.

Figure 4 **Average absolute errors: substantive measures (weighted)**



A-BS = address-based sampling; RDD = random-digit dialling

Figure 5 **Range of average absolute errors across surveys**



These findings are also supported by the results of pairwise *t*-test comparisons (Table 11) that show few significant differences among probability surveys but persistent significant differences among nonprobability surveys.

4 Discussion

In this section, we discuss our findings from a Total Survey Error perspective, paying special attention to the errors associated with the measurement of the substantive variables that were gathered in the eight surveys.

4.1 Coverage and coverage errors

4.1.1 Probability samples

Studies that use probability samples typically take care in choosing their sampling frame(s) to minimise noncoverage and the possibility of non-ignorable coverage error. The three surveys conducted for this study used frames with extremely high coverage of the Australian residential population. There is no reason to expect that any non-ignorable coverage error was present in these three surveys.

4.1.2 Nonprobability panels

The five surveys that used nonprobability-based panels had no known frame from which they selected a 'sample'. They did not cover the Australian residential population well because anyone not using the internet when these panels were undertaking online recruitment would not have been able to consider joining. The noncoverage inherent in these panels is undoubtedly very large and is differential (nonrandom) in nature. It is differential because people exposed to an invitation to join nonprobability panels are different in many non-ignorable ways from those not exposed to an invitation. These differences are often correlated with what is being measured in surveys, such as the substantive measures gathered in this study.

4.1.3 Comparison

Undoubtedly, uncorrectable coverage error in the nonprobability panels contributed to the observed inferiority of the nonprobability panels in how well they generated data for the substantive variables measured in the study. It was also likely to be a reason why the nonprobability panels showed considerably more variation in the accuracy of their substantive measures than did the probability panels.

4.2 Sampling and sampling errors

4.2.1 Probability samples

Surveys that use probability samples generally take considerable care in deciding how they draw their initially designated sample from their sampling frame(s). This was the case for the three probability-based surveys conducted for this study. Using probability sampling gives the users of such surveys a known degree of statistical confidence (associated with sampling error) in those findings. Probability sampling allows confidence intervals to be computed. These can be used to explore the reliability of findings, including point estimates and differences between sampled subgroups. The extent of the error (variance) associated with sampling error can be stated precisely with probability sampling, and its meaning is readily understood. Each of the substantive findings generated from the three probability samples in this study can be assigned confidence intervals.

4.2.2 Nonprobability panels

There is contention as to whether sampling error can or should be calculated for the part of the 'surveying' that is used to recruit opt-in online panels. In contrast, if inferences are limited to the population that is made up of a current nonprobability panel's members, we can calculate sampling error for a survey that uses a sample from that panel. In such an instance, the panel becomes the target population that is being studied. This assumes that a probability sample of the members

is drawn from the panel membership, and it seems unlikely that any nonprobability panel vendor or its clients will want to limit their conclusions only to the members of the panel. So, in reality, most nonprobability panels do not do this, and thus there is no well-accepted, reliable way to calculate the size of the 'sampling error' for a survey of opt-in panel members. Those who conduct surveys based on nonprobability samples sometimes report a *credibility interval* as though it is the equivalent of a confidence interval. However, there are many assumptions that should be met before a credibility interval can be interpreted with any confidence, and most times it is unclear whether nonprobability samples meet these assumptions. Other methods, such as bootstrapping (used in the analysis in this paper) and model-based design weights, can be used to provide measures of precision and reliability of nonprobability surveys. However, because of the complexity and assumptions underlying these methods, they are rarely used in practice.

4.2.3 Comparison

The findings for the substantive measures from the set of probability samples are much more consistent (reliable) than the findings from the five nonprobability samples. Estimates of the substantive measures generated by each survey in the study are much more variable for the five nonprobability panels than for the probability samples. Whether this is due to greater 'sampling error' in the nonprobability samples is unknowable.

4.3 Nonresponse and nonresponse errors

4.3.1 Probability samples

The size of the nonresponse that occurs can be readily calculated in a probability sample survey. Even when a survey is of members of a probability-based panel, such calculations are easy to make, and are essentially the product of the response achieved when building the panel and the completion rate for the survey for which panel members were sampled. For probability panels, several approaches can be pursued to estimate nonresponse bias. This is also a function of the nature of nonresponse that occurred when building the panel and the nature of the nonresponse that occurred within the panel for a particular survey. The three probability-based surveys that were conducted as part of this study suffered from considerable nonresponse, but their response rates were credible by current standards. If the nonresponse that occurred was of a differential nature, as it likely was, post-stratification weighting will likely have reduced the size of any nonresponse bias that was present in the probability samples.

4.3.2 Nonprobability panels

For nonprobability panels, it is impossible to compute a response rate for the time when the panel was established. This is because it is impossible to know how many people were exposed to invitations to join the panel. It is commonly understood that far less than 1% of all people who are exposed to invitations to join a nonprobability panel actually join (Tourangeau et al. 2013:42). Although a completion rate can be calculated for within-panel surveys, this rate does not account for the 'response rate' when the panel was established. The completion rates for the surveys from the four nonprobability sample panels for which this statistic can be calculated were typical for these types of panels.

For opt-in nonprobability panels, there is no well-accepted scientific approach to account for the nonresponse bias. One reason is that those who self-select into such panels are not representative of the cohorts to which they belong. For example, people of Asian background who opt in to nonprobability panels are unlikely to be representative of people of Asian background in the population as a whole. And those who did not complete high school but choose to join a nonprobability panel are unlikely to be representative of those who did not complete high school in the population as a whole. Thus, trying to adjust nonprobability panels solely on demographic characteristics (such as sex, age and education) is unlikely to address non-ignorable biases in the substantive data that the panels generate.

4.3.3 Comparison

It is likely that nonresponse bias stemming from the effort to build the nonprobability panels used in this study explains a good portion of the error that these panels showed for the substantive measures. The same is likely for the error that the probability samples showed on the substantive measures compared with the population benchmarks, but errors for the probability samples are much smaller than the errors for the nonprobability samples.

4.4 Weighting and adjustment errors

4.4.1 Probability samples

Probability samples were designed to be adjusted for selection probability and to conform to population distribution via weighting. Not surprisingly, overall, weighted estimates for these surveys are more accurate than unweighted estimates. Weighting for the ANU Poll did not make any adjustment for the two-stage selection process for these respondents. This process might introduce additional nonresponse and noncoverage errors that are not as effectively corrected for by the post-stratification weights as the single-stage probability samples. Among the three probability samples, the RDD sample has the best weighting efficiency and A-BS the worst.

4.4.2 Nonprobability panels

In nonprobability panels, it is not possible to adjust for the selection probability, so all units are given a design weight of 1. Whether weighting adjustments to enforce population distribution should be applied is debatable, given the enforcement of quotas on these variables and a variety of proprietary mechanisms used by nonprobability panel providers to ensure that their samples resemble the population. Weighting the nonprobability panel estimates reduced the average sampling error for secondary demographic benchmark estimates for all panels. For substantive measures, weighting reduced accuracy for all but one panel (panel 3).

4.4.3 Comparison

Weighting generally reduces survey error for probability samples. Average absolute deviations from the benchmark are lower for weighted probability samples than for unweighted probability samples, and weighted and unweighted nonprobability panels. Without understanding the sample selection processes in the nonprobability panels, it is complex to satisfactorily determine with confidence whether 'probability-style weighting' of nonprobability samples would result in a decrease rather than an increase in total survey error.

4.5 Measurement and measurement errors

The questionnaires that were used in the eight surveys were the same, so there is no reason to expect any differential questionnaire-related measurement error associated with whether a probability sample or an online nonprobability sample was used to gather the data. As such, this error will not be addressed below.

4.5.1 Probability samples

Usually, great care is given to data quality when using probability samples. For interviewer-administered data collection, this includes attention to interviewer training and monitoring. It also includes attention to the way respondents may create error in the form of bias or variance. The data in the ANU Poll and the A-BS survey (both probability samples) are likely affected by the different modes of data collection used in these surveys (online, hard copy, telephone). This is a disadvantage of mixed-mode data collection surveys. It is difficult to identify and correct for such data collection mode effects.

4.5.2 Nonprobability panels

The five nonprobability surveys all used self-administered online data collection. However, it is known that some members of online panels speed through questionnaires (Callegaro et al. 2009, Vannette 2016). Our panel providers took steps to exclude poor-quality responses from the final data

provided. These steps included removing 'straight-liners', removing 'junk'/poor-quality responses to open-ended questions, and removing speeders. Speeders were variously defined as completing the questionnaire in less than 3 minutes, completing the questionnaire within an unspecified departure from the average completion times, or completing the questionnaire in a time one-third or more below the median response time. One panel provider scrutinised respondents who were flagged in their database as having previously provided poor-quality responses.

4.5.3 Comparison

Of the various modes of survey administration, normally it would be assumed that overall data quality would be relatively high in self-administered online surveys. However, data quality within nonprobability panels is often lower than ideal, suggesting that respondent-related measurement error may be an issue for such panels. The probability sample surveys used interviewers to gather at least some of the data. Thus, interviewer-related error may be present in these surveys, and would not be present at all in the nonprobability panel surveys since no interviewers were involved in data collection. Whether the quality of the data was higher in the probability sample surveys or in the nonprobability panel surveys is impossible to know. However, the probability sample surveys had less error (bias and variance), and higher data quality may be part of the explanation.

5 Conclusions

5.1 Strengths and limitations of probability samples

The three probability sample surveys in this study likely had little coverage error, a known amount of sampling error, a non-ignorable amount of nonresponse error, little adjustment error, and a small to modest amount of measurement error.

Overall, the three probability samples as a group were less biased on the substantive measures and had less variance from the benchmark values than the nonprobability surveys.

5.2 Strengths and limitations of nonprobability panels

The five nonprobability surveys in this study likely had a non-ignorable amount of coverage error, an unknowable amount of sampling error, a non-ignorable amount of nonresponse error, unknown adjustment error, and a small to modest amount of measurement error.

Overall, the five nonprobability panel surveys as a group were more biased on the substantive measures and had more variance from the benchmark values than the probability surveys.

5.3 Caveats and declaration of interests

5.3.1 Caveats

As with most research, this study had limitations that need to be taken into account when considering these findings and weighing up how generally applicable they are:

- A limited set of benchmarks was used for this study – 13 secondary demographic benchmarks and 6 substantive health characteristics.
- Benchmarks may have errors. These can arise from various sources, including sampling errors (in the case of surveys), nonresponse errors and measurement (response) errors. This should be considered when thinking about the findings reported in our study. Because errors in the benchmarks should not differentially affect any of the comparisons made between the statistics from nonprobability versus probability samples, benchmark errors would not change our conclusions.
- The nonprobability panel companies that took part in this study were not selected at random, although they were a diverse group of Australian online panels.
- Data collection mode effects may account for some of the observed differences between probability surveys and nonprobability panels.

Nonetheless, the findings from this study accord very well with similar studies undertaken in the United States and Europe.

5.3.2 Declaration of interest

The OPBS was undertaken by the Social Research Centre to learn whether the findings of Yeager et al. (2011) would be replicated in Australia. Since this study did replicate those findings (and others), it has been used to support the establishment of a probability-based online panel in Australia. This panel, the Life in Australia™ panel, is owned and operated by the ANU Social Research Centre.

Appendix A Survey administration

RDD survey

Fieldwork for the RDD survey was conducted over 3 weeks between 30 November and 18 December 2015. Telephone contact was initiated with 3897 members of the original sample pool of 6230 landline and mobile phone records.

Advance letters were distributed to those members of the landline sample frame for whom it was possible to obtain an address via a commercial list provider. Of the 2933 landline numbers selected for the study, 1048 (36%) were matched to an address and sent an advance letter. An advance message (SMS) was sent to the 3297 members of the mobile frame.

The following call procedures were employed:

- Repeated call attempts were spread over different times of day and days of the week, with a view to maximising the sample yield.
- Appointments were set for any time that the call centre was open (weekdays 9 am to 8:30 pm; weekends 11 am to 5 pm).
- A Freecall 1800 number and dedicated email address were established to provide communications channels for sample members to use.
- A survey webpage was established on the Social Research Centre's website.
- The maximum number of unanswered call attempts was capped at three for mobile phone sample records and six for landline records.
- The time zone differences across Australia (up to 3 hours in summer) were taken into account when placing calls to mobile phone numbers, as there is no way of knowing the location (and hence time zone) of the sample member based on their mobile phone number.
- No interviews were conducted in languages other than English, despite offering to conduct interviews in-language via the Translating and Interpreting Service.
- A short message was left the first time an answering machine/voicemail was encountered.
- In addition to the above, 374 records (comprising both landline and mobile phone numbers) were identified as being suitable for refusal conversion activity. Of these, 48 resulted in a completed interview.
- No incentives were offered to sample members approached as part of the RDD survey.

A-BS survey

Fieldwork was conducted over 7 weeks, between 6 November and 23 December 2015.

The A-BS survey was primarily available via a hard-copy booklet (12 pages in length, including the covering letter). This was mailed to 2050 randomly selected sample members on 6 November 2015.

While the invitation method for the A-BS survey was via hard copy, the invitation letter also included instructions for completing the survey online. Online completion was encouraged and incentivised, because it was the preferred method of data collection for cost and timeliness reasons. The questionnaire could also be completed via telephone by contacting the Social Research Centre's Freecall 1800 number or by replying to outbound telephone reminder activity.

Because it is prohibited by Australian law to send cash via the postal service, unless via registered post, included with the hard-copy questionnaire was a $5 prepaid Visa card. The advance letter also informed sample members that, if they participated in the survey, they would get an additional $10 gratuity, by way of a prepaid Visa card or an eGift card.

Regarding the incentive options, 253 respondents (47% of all completes) selected the prepaid Visa card and 180 respondents (33%) opted for the eGift card. Of respondents, 20% either refused the gift card (*n* = 83) or did not respond (*n* = 22).

In addition, a prize draw was offered to A-BS sample members if they completed the survey online. This was to encourage online completion, and keep survey administration and data processing costs to a minimum. The prize was a $250 prepaid Visa gift card.

Response maximisation also entailed reminder activity. A reminder postcard was sent to all A-BS sample members on 20 November 2015. This was to thank those who had already completed the survey and remind others that the survey was still open.

A final reminder was sent on 27 November 2015 to the 1704 remaining nonrespondents. The final reminder included another copy of the survey booklet for those who had misplaced or discarded the original.

Telephone follow-up for sample members for whom we were able to obtain a matched telephone number was the final method of response maximisation adopted for the A-BS survey. This took place between 7 and 18 December 2015. Sample members we were able to contact were offered the opportunity to complete the survey at that time or to arrange an appointment for a more convenient time. Telephone contact was attempted with 743 sample members and yielded an additional 129 completed questionnaires.

ANU Poll

Fieldwork was conducted over 8 weeks, between 19 October and 11 December 2015.

As previously mentioned, sample members for this survey were recruited at the end of the October ANU Poll, undertaken on 12–25 October 2015. At the conclusion of the ANU Poll, respondents were asked to provide an email address for distribution of the Health, Wellbeing and Technology Survey. If no email was available or provided, sample members were asked to give a physical address so that a hard copy of the survey booklet could be sent (with the accompanying cover letter).

The email invitation and hard-copy letter gave participants information about the nature of the study, including how to participate, their right to opt out or withdraw from the study, relevant incentive details and contact details for the ANU Ethics Manager. Sample members could also choose to complete the questionnaire via telephone by contacting the Social Research Centre's Freecall 1800 number and by replying to outbound telephone reminder activity.

During the recruitment process at the end of the October ANU Poll, participants were told that, if they participated in the Health, Wellbeing and Technology Survey, they would receive $10 as a thank you for their time. After completing the survey, respondents had the opportunity to opt for an electronic payment (an eGift card from Coles–Myer) or a prepaid Visa gift card. Respondents were also given the opportunity to decline the incentive.

Almost half of ANU Poll respondents (45%) opted for the prepaid Visa gift card (*n* = 250). The eGift card was selected by 211 respondents (30%), and the remaining 18% of respondents either refused the incentive (*n* = 92) or did not provide a response (*n* = 7).

In addition to the $10 incentive sent to all respondents who did not decline it, a prize draw was offered to ANU Poll respondents if they completed the survey online within the first 4 weeks of fieldwork. The prize was a $250 Visa prepaid gift card.

A multipronged response maximisation strategy was adopted. For nonresponders who had supplied an email address, an initial reminder email was sent one week after the initial invitation (*n* = 410). A second reminder email was also sent on 16 November 2015 to all remaining noncontacts (*n* = 282) to further encourage participation.

A reminder postcard was used for initial follow-up for those sample members who provided only a physical mailing address. The purpose of the postcard was to thank sample members who had

already completed the survey and remind others that the survey was still open.

Following on from these activities, any remaining nonresponders were contacted via telephone and further encouraged to participate. At this point, participants were offered the option to complete an interview at that time or arrange an appointment for a more convenient time. This telephone contact was attempted with 364 sample members and yielded a further 230 interviews.

Online panels

Online data collection took place in late November and early December 2015.

For the online panel surveys, panel providers sent out invitations to panel members. The invitations included a direct link to the questionnaire, which was programmed in-house by the panel providers using their own software. The fact that panel providers used differing software to program the survey questionnaire resulted in variations in the presentation of some questions, while also providing respondents with the look and feel they were accustomed to when participating in research via a particular panel.

The number of invitations sent out by the panel providers and the number of invitations that were opened by panellists varied considerably:

- Panel 1 – data not available
- Panel 2 – 7097 invitations sent, 2315 (33%) opened
- Panel 3 – 4060 invitations sent, 1241 (31%) opened
- Panel 4 – 6132 invitations sent, 684 (11%) opened
- Panel 5 – 23 527 invitations sent, 1314 (6%) opened.

Online panel respondents were provided with a contingent incentive in accordance with the usual practice of the panel companies. Incentive strategies used by the panel providers included:

- point incentives that can be ‘cashed’ in
- prizes or sweepstakes
- donations to charity
- gift vouchers
- offers to receive information (usually in the form of a results summary).

Incentive methods are often tailored to individual panel members based on their motivation to take part. Further, some panel members are happy to take part without being incentivised. Incentives are linked to survey length and the effort required by the panel member.

None of the online panel providers undertook response maximisation or reminder activity beyond their normal incentive strategies.

Appendix B Population benchmarks used for post-stratification weighting

Table B1 Benchmarks used for post-stratification weighting

Benchmarks	Percentage
Telephone status[a]	
Mobile only	29.0
Dual user	63.2
Landline only	7.8
Education by age[b]	
25–34 years, with university degree	6.5
35–44 years, with university degree	5.0
45–54 years, with university degree	3.8
55–64 years, with university degree	2.8
65–74 years, with university degree	1.4
75+ years, with university degree	0.6
18–24 years	12.3
25–34 years, without university degree	12.5
35–44 years, without university degree	12.5
45–54 years, without university degree	13.1
55–64 years, without university degree	12.0
65–74 years, without university degree	9.5
75+ years, without university degree	7.8
Region[b]	
Sydney	20.5
Rest of New South Wales	11.6
Melbourne	18.9
Rest of Victoria	6.3
Brisbane	9.5
Rest of Queensland	10.3
Adelaide	5.6
Rest of South Australia	1.7
Perth	8.4
Rest of Western Australia	2.4
Tasmania	2.2

Benchmarks	Percentage
Northern Territory	1.0
Australian Capital Territory	1.6
Sex[c]	
Male	49.3
Female	50.7
Country of birth[d]	
Australia	68.6
Other English-speaking country	11.2
Non-English-speaking country	20.2
Age group[c]	
18–24 years	12.3
25–34 years	19.0
35–44 years	17.5
45–54 years	16.9
55–64 years	14.8
65–74 years	10.9
75+ years	8.4
State[c]	
New South Wales	32.1
Victoria	25.2
Queensland	19.8
South Australia	7.3
Western Australia	10.8
Tasmania	2.2
Northern Territory	1.0
Australian Capital Territory	1.7

a The post-stratification benchmark for telephone status is constructed as follows: 29.0% of Australians have mobile phone only (ACMA 2015); therefore the remaining 71.0% must own a landline (assuming the proportion of people owning neither is negligible); 89% of people in households with a landline also own a mobile (ACMA 2011), so 63.2% (0.89 × 0.71 = 0.632) of people are dual users, leaving 7.8% of Australians as landline phone only.

b Data from ABS Census 2011 TableBuilder and ABS estimated resident population June 2015, cat no. 3101.0.

c Data from ABS estimated resident population June 2015, cat. no. 3101.0.

d Data from ABS Census 2011 TableBuilder.

Appendix C Multimode surveys: respondent profile by mode of completion

When looking at the mode of completion by selected respondent characteristics for the A-BS survey (Table C1), it is evident that a lower proportion of completes were obtained from males (34%) than females (66%) as a result of telephone follow-up. A higher proportion of younger respondents (aged 18–44) responded online (41%) than via hard copy (13%) or telephone (22%). Hard copy was more popular outside the major cities and among those who did not complete Year 12 at secondary school.

For those respondents recruited via the ANU Poll, telephone was the primary mode of completion for those aged less than 35 (perhaps reflecting that this group was more likely to respond when followed up than when first invited). There was little geographic variation in the mode of completion, and Australian-born people were more likely to respond online than those born overseas.

Given that the online and hard-copy modes of completion are similar in that they are both self-completion modes of data collection, there is an opportunity to undertake further analysis of the ABS and ANU Poll findings to understand the impact that the inclusion of responses obtained from the telephone surveys, undertaken in response to telephone reminder activity, has on the accuracy of the survey estimates.

Table C1 **Respondent profile by mode of completion for the multimode surveys**

Demographic	A BS (%)			ANU Poll (%)		
	Hard copy (*n* = 202)	Online (*n* = 208)	Telephone (*n* = 128)	Hard copy (*n* = 40)	Online (*n* = 292)	Telephone (*n* = 228)
Total	38	39	24	7	52	41
Sex						
Male	41	39	34	28	40	46
Female	59	61	66	72	60	54
Age (years)						
18–24	2	5	5	–	5	8
25–34	4	17	6	2	6	13
35–44	7	19	11	2	14	18
45–54	12	18	20	8	20	19
55–64	24	23	19	10	23	21
65–74	28	12	27	28	25	14
75+	20	6	12	48	6	7
Geography						
Major cities	65	82	71	75	67	71
Inner regional	26	12	20	18	23	18
Outer regional	8	5	9	8	9	9
Remote	2	*	0	0	1	1
Very remote	0	0	0	0	*	*
Education						
Did not obtain Year 12	49	19	37	67	33	35
Obtained Year 12	20	24	21	23	26	22
Bachelor's degree or higher	30	57	42	10	41	43
Birthplace						
Australia	73	70	76	70	79	71
Other	27	30	24	30	21	29
Telephone status						
Mobile only	20	20	7	15	7	19
Landline only	11	5	10	32	4	6
Dual user	66	74	83	52	89	75
No phone	2	1	0	0	0	0

* = <0.1%; A-BS = address-based sampling

Notes

1. The AAPOR Opt-In Task Force Report (Baker et al. 2010) describes opt-in online panels as follows: '… Most online panels are not constructed using probability-based recruitment. Rather, they use a broad range of methods to place offers to join in front of prospective panellists. Those offers are generally presented as opportunities to earn money but also emphasize the chance to have a voice in new products and services and the fun of taking surveys. People join by going to the panel company's website and providing varying amounts of personal and demographic information that is later used to select panellists for specific surveys.'
2. The Knowledge Panel, established in 1999, and owned by GfK since 2012, is widely acknowledged as the first probability-based online panel.
3. Data and documentation can be found at https://www.ada.edu.au/ada/01329.
4. The item measuring volunteerism – 'Over the last 12 months, did you spend any time doing voluntary work through an organisation or group? Please note this does NOT include anything you do as part of paid employment, to qualify for a government benefit or any work done in a family business' – was treated as both a secondary demographic and a possible calibration variable.
5. The item measuring access to the internet at home was treated as both a secondary demographic and a possible calibration variable.
6. https://www.psma.com.au/products/g-naf
7. Based on the proportion of people who own landlines (71%; ACMA 2015) and the number of households according to ABS projections (9 182 917; ABS 2015a).
8. Based on Australian Communications and Media Authority reports on the percentage of people who are mobile only (29%; ACMA 2015), the percentage of people with a landline who also own a mobile phone (89%; ACMA 2011) and the ABS estimated resident population for people aged 18 years and over at June 2015 (18 437 213; ABS 2017).
9. Whether this is a prudent assumption is not known by survey researchers because there is a conspicuous lack of reliable data about sharing of mobile phones. Some studies in the United States and Europe estimate the rate of sharing to exceed 20%.
10. This is a very conservative definition, because not all people who 'can' access the internet do in fact access it or are willing to access it to complete a questionnaire.
11. In theory, infinitely many replications could be used; however, in practice, the number of replications is driven by the time and computing power available. Yeager et al. (2011) used 100 replications for their estimates. Because of the computationally demanding nature of regenerating post-stratification weights for each replication, 150 was selected for the weighted estimates, 1000 was used for unweighted estimates. Changing the random seed for weighted estimates did not meaningfully change the results.

References

ABS (Australian Bureau of Statistics) (2008). *Household use of information technology, Australia, 2006–07*, cat no. 8146.0, ABS, Canberra.

—— (2013). *Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011*, cat no. 2033.0.55.001, ABS, Canberra.

—— (2014). *General Social Survey, 2014*, cat. no. 4159.0, ABS, Canberra.

—— (2015a). *Household and family projections, Australia, 2011 to 2036*, cat. no. 3236.0, ABS, Canberra.

—— (2015b). *National Health Survey, 2014–15*, cat. no. 4634.0.55.001, ABS, Canberra.

—— (2016). *Household use of information technology, Australia, 2014–15*, cat. no. 8146.0, ABS, Canberra.

—— (2017). *Australian demographic statistics, Jun 2017*, cat. no. 3101.0, ABS, Canberra.

ACMA (Australian Communications and Media Authority) (2011). *Converging communications channels: preferences and behaviours of Australian communications users*, Communications Report Series 2010–11, report 2, ACMA, Canberra.

—— (2015). *Communications report 2014–15*, ACMA, Canberra.

AIHW (Australian Institute of Health and Welfare) (2013). *National Drug Strategy Household Survey, 2013*, AIHW, Canberra.

Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, Dillman DA, Frankel MR, Garland P & Groves RM (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly* 74(4):711–781.

——, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, Gile KJ & Tourangeau R (2013). *Report of the AAPOR Task Force on non-probability sampling*, American Association for Public Opinion Research, Lenexa, Kansas.

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

——, Yang Y, Bhola DS, Dillman DA & Chin TY (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 103(1):5–25.

——, Baker M, Bethlehem J, Göritz A, Krosnick J & Lavrakas P (eds) (2014). *Online panel research: a data quality perspective*, Wiley, New York.

Canty A & Ripley B (2015). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-17. https://cran.r-project.org/web/packages/boot.

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

Davison AC & Hinkley DV (1997). *Bootstrap methods and their applications*, Cambridge University Press, Cambridge, UK.

ESOMAR (2015). *Global market research: an ESOMAR industry report*, ESOMAR, Amsterdam, The Netherlands.

Fine B (2016). *Online research panels around the world: the situation in Australia*, presentation to Current State and Future of Online Research in Australia Workshop, Canberra.

Gaziano C (2005). Comparative analysis of within-household respondent selection techniques. *Public Opinion Quarterly* 69(1):124–157.

Groves RM (1989). *Survey errors and survey costs*, Wiley, New York.

Kennedy C, Mercer A, Keeter S, Hatley N, McGeeney K & Gimenez A (2016). *Evaluating online non-probability surveys*, Pew Research Center, Washington, DC.

Lavrakas PJ, Benson G, Blumberg S, Buskirk T, Flores Cervantes I, Christian L, Dutwin D, Fahimi M, Fienberg H, Guterbock T, Keeter S, Kelly J, Kennedy C, Peytchev A, Piekarski L & Shuttles C (2017). *The future of US general population telephone survey research*, American Association for Public Opinion Research, Deerfield, USA.

Research Industry Council of Australia (2016). *Australian Market and Social Research Industry Survey*, Research Industry Council of Australia, Glebe.

Tourangeau R, Conrad FG & Couper MP (2013). *The science of web surveys*, Oxford University Press, New York.

Vannette D (2016). *Survey speeding, part 1: is it harmful or harmless?* https://www.qualtrics.com/blog/survey-speeding-harmful-or-harmless

Walker R, Pettit R & Rubinson J (2009). The foundations of quality initiative. *Journal of Advertising Research* 49(4):464–485.

Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, 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):709–747.

CENTRE FOR SOCIAL
RESEARCH & METHODS

+61 2 6125 1279
csrm.comms@anu.edu.au

The Australian National University
Canberra ACT 2601 Australia

www.anu.edu.au