



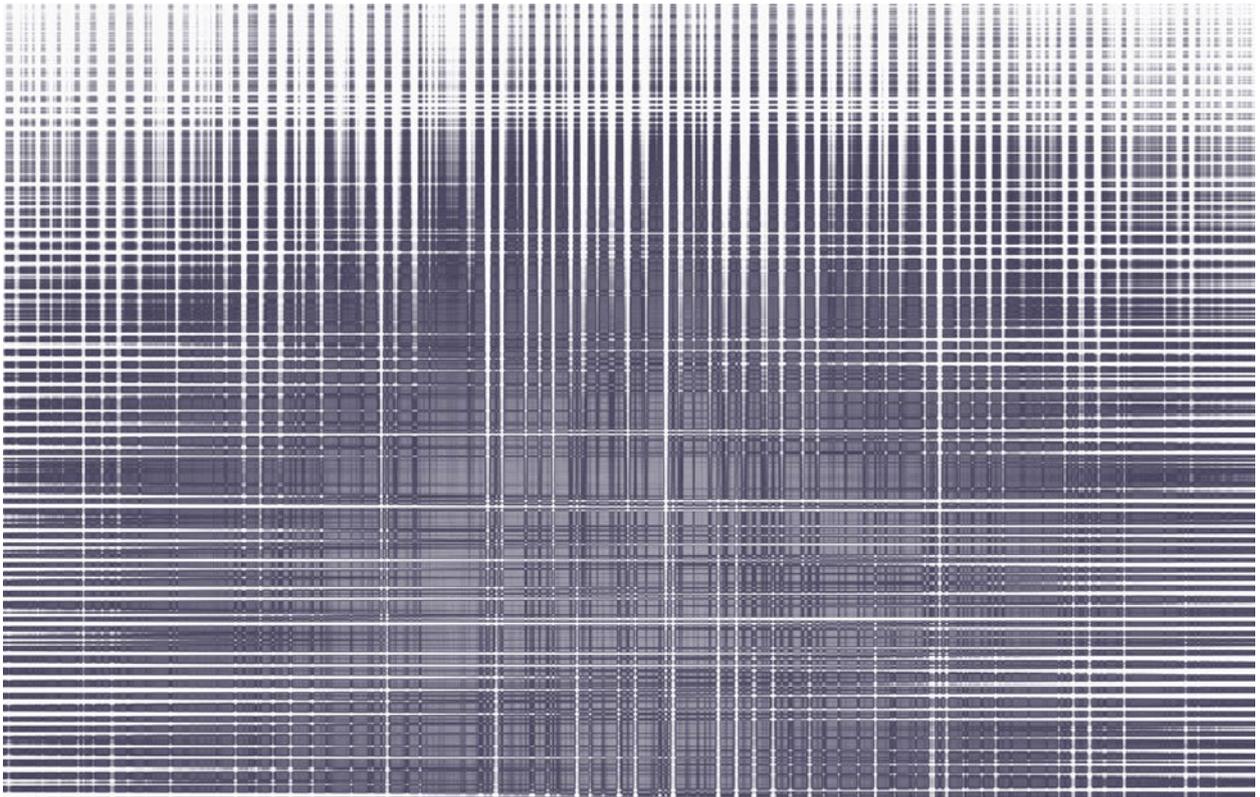
Australian
National
University



Social
Research
Centre

An ANU Enterprise business

CENTRE FOR SOCIAL
RESEARCH & METHODS



Measuring social attitudes with voter advice application data

J Sheppard

CSRM & SRC METHODS PAPER

NO. 3/2018

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://csmr.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. 3/2018

ISSN 2209-184X

ISBN 978-1-925715-08-8

An electronic publication downloaded from
<http://csmr.cass.anu.edu.au/research/publications>.

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

ANU Centre for Social Research & Methods

Research School of Social Sciences
The Australian National University

Measuring social attitudes with voter advice application data

J Sheppard

Jill Sheppard is a Lecturer in the School of Politics and International Relations, Australian National University.

Abstract

This study directly compares survey data on social attitudes collected from an opt-in sample of voter advice application (VAA) users and a randomly recruited, probability-based online panel of respondents. Whereas much research to date has focused on the demographic representativeness of VAA-generated data, less is known about the attitudinal and other representativeness of these data. The findings from these Australian samples contribute to the emerging international literature.

VAA's are proliferating as a source of 'big data' among public opinion and political science researchers, despite concerns about the representativeness of the opt-in samples. During July 2016, VAA developer Election Compass collected email address details for approximately 40 000 Australian users of its application in the weeks before the 2016 Australian federal election. In November 2016, this study surveyed the sample of VAA users on their attitudes towards a range of Australian social issues. In December 2016, the same questionnaire was administered to a probability-based sample,

using an identical mode of administration and similar response maximisation techniques. The questionnaire contains a broad range of questions designed to identify dimensions of sociopolitical attitudes in Australian society. Comparing the composition of dimensions and relationships between variables within the data contributes to our understanding of incidental samples such as VAA users, and the extent to which we can and should make inferences from VAA-generated data.

Comparison of point estimates of the unweighted and weighted datasets, and of estimates of the internal relationships between variables within both datasets suggests that VAA user data should not be considered externally valid sources of social attitude data, even after adjustment. Further, researchers and reviewers should use greater caution when adjusting VAA user data on the basis of observable measures (with well-established population parameter estimates) to explain unobservable measures without robust population parameter estimates (such as social attitudes).

Acknowledgments

The author is grateful for the contributions of data and advice from Professor André Krouwel, Vrije Universiteit Amsterdam and Kieskompas, and Dr Dina Neiger, Social Research Centre, Australian National University.

Acronyms

ANU Australian National University

CSRM Centre for Social Research & Methods

TSE total survey error

VAA voter advice application

Contents

Series note	ii
Abstract	iii
Acknowledgments	iv
Acronyms	iv
1 Introduction	1
1.1 Quality and representation error in social surveys	2
1.2 Voter advice applications as survey data sources	3
2 Data and methods	6
3 Analysis	8
3.1 Demographic distributions	8
3.2 Attitudinal distributions	11
3.3 Relationships between variables	13
4 Conclusion	17
References	18

Tables and Figures

Table 1	Social and political attitude measures _____	7
Figure 1	Unadjusted distribution of age among three samples (local regression smoothed) _____	8
Figure 2	Adjusted distribution of age among three samples (local regression smoothed) _____	9
Figure 3	Unadjusted distribution of female and male respondents, by sample _____	10
Figure 4	Adjusted distribution of female and male respondents, by sample _____	10
Figure 5	Unadjusted and adjusted frequencies (percentage) of respondents with a bachelor degree or higher, by sample _____	10
Figure 6	Unadjusted distribution of attitudes to income equality, by sample _____	11
Figure 7	Adjusted distribution of attitudes to redistribution, by sample _____	12
Figure 8	Unadjusted distribution of attitudes to national pride, by sample _____	12
Figure 9	Adjusted distribution of attitudes to national pride, by sample _____	13
Figure 10	Unadjusted regression coefficients predicting agreement that ‘Everything is changing too often and too fast’ _____	14
Figure 11	Adjusted regression coefficients predicting agreement that ‘Everything is changing too often and too fast’ _____	15
Figure 12	Unadjusted regression coefficients predicting agreement that ‘All of us together have a responsibility to make sure that everyone has enough money to get by’ _____	16
Figure 13	Adjusted regression coefficients predicting agreement that ‘All of us together have a responsibility to make sure that everyone has enough money to get by’ _____	16



1 Introduction

Social survey research – in the political science disciplines, and elsewhere – faces an uncertain future. Declining rates of participation among potential survey respondents threaten the generalisability of social surveys (e.g. see Dillman et al. 2014). When 10% or less of a probabilistically recruited sample complete surveys, can we reasonably extrapolate results to that population? While traditional probabilistic sampling techniques seek to maximise the likelihood that any individual in a given population will be sampled, declining response rates in both telephone- and postal-administered surveys mean that, regardless of the population coverage achieved in any given sample, the eventual respondents can bear little resemblance to the population from which they are drawn. This error – nonresponse bias – is arguably the greatest problem facing contemporary social survey research (Groves 2006).

At the same time, sources of nonprobabilistically sampled data are proliferating, as are statistical methods to adjust these data to population parameters. Social researchers are increasingly turning to cheaper, more convenient data sources, such as opt-in samples of paid survey respondents (e.g. consumer research panels, newer sources such as Amazon’s Mechanical Turk workforce), river sampling, and incidental social media and internet browsing data. The techniques applied to make the data more closely resemble population estimates include design weighting, calibration weighting, propensity score matching, and multilevel regression and post-stratification. Although the combination of convenient, nonprobabilistic data and sophisticated adjustment techniques has yielded successful population estimates in some domains (Wang et al. 2014), probability-based data tend to have greater external validity across a broad range of social behaviours and attitudes (Yeager et al. 2011, Dutwin & Buskirk 2017).

This study examines the external validity of an emerging, but under-interrogated, data source: voter advice application (VAA) respondent data.

VAA’s are web applications that use respondents’ attitudes to relevant policy issues to identify the ideological distance between the respondent and the parties or candidates contesting an election. As suggested by the name, VAA’s originated as a tool to advise voters on how to vote for their most ideologically proximate party or candidate. They have rapidly become a feature of general elections in a growing number of liberal democracies (Rosema et al. 2014). As they have grown in popularity, their use has extended from providing voter advice to forming a source of public opinion and political behaviour data. However, a small number of studies have found systematic selection biases among VAA users, with the implication that VAA-generated data are not typically externally valid (Pianzola 2014, Van de Pol et al. 2014).

In the rush to identify and analyse convenient sources of survey data, VAA user data have not received the degree of scrutiny applied to other emerging data sources, such as Mechanical Turk (e.g. see Berinsky et al. 2012). Moreover, the adjustment techniques commonly applied to VAA user data (such as raking weights) assume that biased samples of VAA users can be weighted to reflect populations of voters on the basis of observable sociodemographic variables (e.g. see Carson et al. 2016). Without sufficient information on the unobservable (e.g. attitudinal or motivational) biases among VAA users, adjusting VAA-generated data against observable population parameters does not necessarily improve external validity (cf. Dutwin & Buskirk 2017).

In querying the extent to which VAA user data can validly measure population-level social attitudes, this study compares survey data generated from a nonprobabilistic sample of Australian VAA users with that generated from respondents belonging to a probability-based online sample. Both surveys were conducted in October–November 2016. The two questionnaires include identical questions on attitudes to a range of policy issues, and comparable measures of sociodemographic

characteristics. In the absence of population parameters for the attitudinal measures, the probability-based data provide proxy population benchmarks.

1.1 Quality and representation error in social surveys

Recent declines in response rates rightfully call into question the external validity of social surveys, but nonresponse bias is not the only potential source of error. This paper uses the ‘total survey error’ (TSE) approach to assessing the respective quality of nonprobabilistically and probabilistically sampled datasets (Weisberg 2009). The TSE framework considers the various sources of error inherent in any survey: measurement error, including construct validity, incorrect or inaccurate responses, and processing or administrative errors; and representation error, including undercoverage or overcoverage of a population, errors in raising a representative sample, nonresponse bias, and adjustment error (Biemer 2010, Groves & Lyberg 2010). The framework synthesises the literature on sources of survey error and comparative survey quality dating back to the 1930s (Groves & Lyberg 2010).

One substantial source of potential survey error is in the identification of the target population. For instance, investigators enumerating most post-election surveys of voter attitudes, such as the American National Election Studies, usually target a national population of eligible voters. However, rarely does a list of all eligible voters exist (e.g. see Stoker & Bowers 2002). Survey investigators need to decide on an optimal combination of population coverage (i.e. the extent to which the target population and available sampling frame overlap), inclusion of hard-to-reach respondents such as ethnic minorities and young people, and cost of accessing the sampling frame. The suitability of any sampling frame for population coverage varies by context: members of a population might be more important to answering one research question than another (Dillman et al. 2014). Coverage error is typically measured as bias in members of the sampling frame with respect to the population, on any given (observable) variable (Groves et al. 2009).

At the second stage of the sampling process, a sample of potential respondents is drawn from the chosen sampling frame. Sampling error (in this context) describes systematic bias in the likelihood that members of a sampling frame will be sampled. Sampling error is mitigated when all members of a sampling frame (or, in the case of stratified and cluster sampling techniques, all members of a subgroup within a sampling frame) have exactly the same likelihood of being randomly sampled. Researcher decisions and external constraints can introduce sampling error. For instance, larger sample sizes mitigate sampling error, but are expensive; money spent ensuring that every member of a population is equally likely to be sampled is money not spent mitigating other errors. Similarly, conducting a survey using several different modes might minimise sampling error, but increase measurement error by introducing cross-modal variance (Dillman et al. 2014). As with coverage error, techniques to minimise sampling error vary by context (Biemer 2010).

Arguably the most urgent problem facing survey researchers is nonresponse error: the bias inherent in survey estimates when subgroups of a population systematically fail to respond (e.g. see Massey & Tourangeau 2012). Declining response rates to surveys conducted by telephone, by mail and in person have attracted substantial attention from researchers. Response maximisation techniques include nonresponse follow-up, respondent incentives, responsive design, interviewer training and pre-recruitment notifications (Dillman et al. 2014, van Veen et al. 2016). However, nonresponse error is not just a factor of nonresponse rates; if nonrespondents are sufficiently random in composition, the number of them should not increase bias in the sample (Groves 2006). Accordingly, techniques designed to maximise overall response rates might also increase nonresponse error through the systematic bias of respondents on either observable or unobservable domains.

Post-enumeration adjustments to survey estimates (such as weighting) can help to mitigate overall error in the final survey statistic by accounting for nonresponse error. For instance, if members of the population aged 18–24 years are underrepresented in the final sample, they

can be weighted at >1 in the survey statistic estimated from the data. That is, each individual respondent aged 18–24 will account for more than one respondent in any estimation. Weighting survey data on observable strata such as age, sex or educational attainment is referred to as post-stratification, reflecting that stratification of the sample to population benchmarks occurs after the survey is enumerated, rather than before (as in the case of stratified random or quota-based sampling). Although post-stratification is commonly used to reduce nonresponse error, it can also introduce a new form of error: adjustment error (Groves et al. 2009).

Post-stratification seeks to adjust a survey sample so that it more closely resembles a population on the basis of observable differences only. The process of creating post-stratification weights at the level of the individual respondent is therefore dependent on both the sample estimate and the population estimate (or marginal distribution); reliance on the sample estimate means that error at previous stages of the survey life cycle can contribute to error at the adjustment stage (Gelman 2007). Accordingly, post hoc adjustment of sample estimates can embed, and even amplify, previously introduced error. Adjustment can also introduce bias into the distribution of variables for which the population distribution is not known (Groves 2006). If a sample estimate is post-stratified on the bases of educational attainment, sex, household composition and age at the individual level, the adjusted estimate on any number of variables of substantive interest (i.e. relevant to a research question) may be further from the (unknown) parameters of those variables within the population.

In contrast, we may consider some samples ‘fit for purpose’, despite the presence of substantial measurement and/or representation error. For instance, research that benchmarks opt-in samples of paid respondents against more robust, probabilistically recruited samples has found that samples with higher rates of absolute error have high rates of population representativeness on one or more specific dimensions (Baker et al. 2010). Nonprobabilistic samples from Australia are shown to overrepresent individuals who engage in risk-taking behaviours such as cigarette and alcohol

use, making them suitable for nonprevalence-related public health research (Pennay et al. 2015). In such cases, the internal validity of the sample estimates – for example, the relationship between educational attainment and smoking incidence – is as important as, or more important than, the external validity of the estimates against population parameters. Likewise, the internal validity of opt-in VAA-generated data can justify analysis that does not seek to extrapolate estimates to a population of all voters; examples are temporal variation, or changes in response patterns due to exogenous political events.

1.2 Voter advice applications as survey data sources

A typical VAA allows web users to opt in to a (generally short) questionnaire measuring political and social attitudes; a small number of demographic questions are included for the purpose of post hoc adjustment and academic research (Gemenis & Rosema 2014, Rosema et al. 2014). As of 2014, millions of individuals internationally have used VAAs to help inform their vote; we can reasonably expect that number to have increased since then (Rosema et al. 2014). Political scientists have sought to explain VAAs’ effects in mobilising voter turnout (Marschall & Schultze 2012, Dinas et al. 2014, Gemenis & Rosema 2014, Pianzola 2014), vote choice (Walgrave et al. 2008, Mendez 2012, Wagner & Ruusuvirta 2012, Wall et al. 2014, Kleinnijenhuis et al. 2017) and users’ political knowledge (Fivaz & Nadig 2010, Fossen & Anderson 2014, Schultze 2014, Kamoen et al. 2015). Although important broadly, the various effects of VAA use and output on each of these dimensions are peripheral to this particular study.

A small number of studies have analysed the demographic and sociopolitical profile of VAA users, finding substantial biases. Published studies unanimously acknowledge the effect of selection biases on samples composed of VAA users. These biases are likely more substantial than the nonresponse bias incurred in a probabilistically sampled survey in which potential respondents are actively recruited to participate (Pianzola 2014). VAA users across cultural, democratic and electoral contexts are,

on average, younger, more highly educated and more politically engaged, and have stronger partisan identity than the voting-age population (Dumont & Kies 2012, Dinas et al. 2014, Van de Pol et al. 2014, Marschall & Schultze 2015). In many countries, VAA users are disproportionately male (Van de Pol et al. 2014). Importantly, they are also likely to be consumers of other politics-related media, and recruitment to the VAA through mainstream media use can skew the composition of the user sample (Çarkoğlu et al. 2012, Van de Pol et al. 2014). Several studies indicate that individuals' decision to use a VAA has more in common with a uses and gratifications approach (from the communications tradition) than with conventional models of survey participation (Schultze 2014, Van de Pol et al. 2014, Kamoen et al. 2015).

Assessing VAA-generated data against TSE criteria for data quality highlights the likely introduction of substantial error throughout the data collection process. On the measurement side, the construct validity of VAA instruments usually depends on investigators' interpretation and operationalisation of party and candidate manifestos. VAA investigators commonly base the survey instruments on some definition of 'relevant' policy issues, as identified by one or more experts (Van Camp et al. 2014). Across VAAs at 26 elections in 10 countries, Van Camp et al. (2014) found a 'troubling' high number of 'double-barrelled' issue statements (with which users are asked to agree or disagree), and that often the need for breadth in operationalising constructs loses out to the imperative to build a short, user-friendly survey. On the other hand, the issue statements included in VAAs tend to be specific, leaving little room for misinterpretation (Van Camp et al. 2014).

The potential for measurement and processing errors throughout the VAA process lies largely with the user, because questionnaires are designed to be self-completed, and processing occurs within the application. Measurement error is mitigated by the use of clear, specific questions or statements, but self-administered web surveys can incur more incorrectly entered responses than 'offline' or interviewer-administered surveys (Andreadis 2014). Programming errors at the processing stage could introduce systematic

error, but are likely to be discovered during testing processes or the fieldwork period and subsequent reporting. Probably the most substantial claim made against VAAs concerns the spatial positioning of parties or candidates relative to voters (Mendez 2012). This remains a valid, and unresolved, criticism with no universal solution.

The more serious potential for systematic error in VAA-generated data is with representation. The initial stage of the representation process – identifying an inferential population – is not addressed in the design of VAAs or in the subsequent use of their data. This is not necessarily a source of error; VAAs are generally designed for voter education, not for scientific research ends (Van Camp et al. 2014). Incidental users from outside any broadly conceived population (e.g. voting-eligible citizens) do not undermine the reliability of VAAs as a voter education tool. Because VAA investigators do not target a specific population for inferential purposes, they are not commonly concerned with reducing coverage error (Van Camp et al. 2014). That is, a VAA is no less valid or reliable if not all individuals within an (undefined) population have access to the VAA – for example, individuals without internet access, who are systematically excluded from VAA use. Such bias would seriously undermine the external validity of any survey sample claiming to represent a voting-eligible population.

Alternatively, if VAAs can be said to target a population of interested individuals with access to the internet, coverage error within the ensuing sample of users is very low. Any individual with sufficient interest and internet access has approximately equal chance of being sampled (i.e. of opting in to the final sample of users). The greater source of error, in this conceptualisation of the target population, then shifts to the sampling itself. Individuals within the population of interested, internet-connected individuals will have differential exposure to information about a VAA, potentially biasing the sample of users (Çarkoğlu et al. 2012). Finally, many individuals within the target population (however defined) who become aware of a VAA will choose not to use it, whether due to lack of interest, resource constraints, reluctance to share

personal information or lack of internet-related skills (among any number of other reasons). The final nonresponse error in a sample of VAA users is largely unobservable, as a result of lack of information on the possible reasons for an individual to choose not to opt in to VAA use (as opposed to choosing not to respond to a survey following an attempt at recruitment).

Despite these systematic errors within VAA-generated data, a growing number of political science studies employ VAA-generated data to test national population-based hypotheses. These studies often include caveats about external validity, or maintain that post hoc adjustment of the data allows population inference. To date, peer-reviewed studies have analysed VAA-generated data with regard to party–voter congruence (Talonon & Sulkava 2011, Lees-Marshment et al. 2015), voters in ideological space (Mendez & Wheatley 2014, Mendez 2017), voter behaviour in the ‘Brexit’ referendum (Antonucci et al. 2017), nationalist voting behaviour (Loewen et al. 2015), and ideology and intolerance (van Prooijen & Krouwel 2017). Few of these papers validate their findings with reference to probabilistically sampled data; among those that do, Johnston (2017) found similar distributions between VAA-generated data and Canadian Election Study data on sociodemographic measures, but stark differences in political attitudes and behaviours over the course of a campaign. Generally, the external validity of VAA-generated data as a measure of social and political attitudes remains largely unaddressed.

2 Data and methods

This study compares data generated from a probabilistically sampled survey of social and political attitudes with data from a sample of VAA users. It asks whether the VAA-generated sample can validly measure a range of attitudes, using the probabilistic data as a benchmark in the absence of known population parameters on attitudinal questions. To this end, the study focuses on forms of representation error in the VAA-generated data: coverage error, sampling error, nonresponse error and adjustment error. The analysis first compares the distributions of sociodemographic and attitudinal variables from the two samples, with the variables both unadjusted (i.e. unweighted) and adjusted (i.e. weighted) against population benchmarks. Second, the internal relationships between variables are compared, with regression models providing an additional measure of the external validity of the VAA-generated data.

The VAA-generated data were collected in December 2016, entirely online. In July 2016, the Australian media company Fairfax partnered with Dutch voting advice company Kieskompas and the University of Sydney to host a VAA (branded as YourVote) on the websites of two Fairfax-owned newspaper outlets: *The Age* and *The Sydney Morning Herald*. The VAA attracted approximately 240 000 respondents. Of these, 9895 individuals provided their email address, thereby consenting to be contacted by Kieskompas in the future. No data are available on the demographic profile of these 9895 respondents. In November 2016, Kieskompas emailed the 9895 VAA users a link to an online survey with questions on a range of social and political attitudes. A total of 3123 respondents completed the survey by the end of November 2016. The data are adjusted against population benchmarks with raking weights by respondents' state, sex and highest educational attainment.

The benchmark data were collected in December 2016 from a sample of respondents recruited

through dual-frame (70% mobile; 30% landline) random digit dialling. Respondents agreed to join the Social Research Centre's newly established Life in Australia research panel; the survey analysed here was the first questionnaire administered to these respondents. Respondents receive an incentive worth A\$10 for completing one questionnaire per month. The initial recruitment rate for the panel was 15.5%, and the completion rate for this survey was 78.8%, creating a cumulative response rate of 12.2%. Most (87%) respondents completed the survey on the internet; the remaining 13% (most of whom do not have internet access at home) completed it by telephone. The data are adjusted against population benchmarks, with raking weights derived from respondents' state, sex, telephone status, home internet use and highest educational attainment.

The probability-based and VAA sample questionnaires contained an identical module of issue statements, with which respondents are asked to strongly disagree, disagree, neither agree nor disagree, agree or strongly agree (that is, a standard 5-item Likert scale response frame), with responses scored from 1 (strongly disagree) to 5 (strongly agree). The statements consistent across both questionnaires are shown in Table 1. They include a range of attitudinal and personal value measures, selected as part of a project to generate a multidimensional typology of Australian adults.¹ The study is not concerned with any measurement error in either sample, and constraining the measures to identically worded statements in both datasets effectively holds measurement error constant. Some measurement error may be incurred in the different modes of administration within the probability-based sample (i.e. online and telephone); for that reason, online and telephone-administered responses are modelled separately as appropriate.

¹ The probability-sampled data were used as the basis of Fairfax Media's Political Personas Project, on which the author was an academic adviser.

Table 1 Social and political attitude measures

Measure	Statement wording
Career	I want to get to the very top in my career
Change	Everything is changing too often and too fast
Confidence	I have confidence in society
Differences	I like different people, cultures, ideas and lifestyles around me
Disillusion	I am disillusioned with politics in this country
Fitness	It is important to me to stay fit
Food	It is important to me that the food I eat has been sourced ethically
Future	I sometimes feel that the future holds nothing for me
Goals	I constantly set myself higher goals, which I strive to achieve
Housing	The price of housing is creating a class system in Australia
Incomes	I think that the difference between high and low incomes should be smaller
Job	It is important that my job provides a sense of personal fulfilment
Leadership	Australia needs a strong leader, who can quickly make decisions
Let down	I feel let down by society
Look	The way I look is very important to me
Luxury	I believe that buying luxury goods is wasting money
Manufacturing	We rely too heavily on foreign imports and should manufacture more in Australia
Money	I am feeling pretty good these days about how much money I can spend
Nanny state	Australia has become a nanny state
Organised	I like my life to be organised and predictable
Pride	I am very proud of the country I live in
Schools	Private schools offer a superior education to public schools
Security	Financial security is an important goal in my life
Spend	I prefer to spend my money now, rather than saving it for later
Values	I think that there is too little emphasis on traditional values in Australia
Welfare	All of us together have a responsibility to make sure that everyone has enough money to get by

3 Analysis

3.1 Demographic distributions

Studies in political science that rely on VAA-generated data commonly draw attention to similarities between the demographic compositions of VAA users and those of more established data sources such as national election studies (Carson et al. 2016, Johnston 2017). Beginning with age, the demographic composition of VAA respondents, online plus telephone probability respondents, and online-only probability respondents varies substantially (Figure 1). The smoothed distributions shown in Figure 1 reveal some overrepresentation of the very oldest respondents in the probability online-only sample. Otherwise, the full and subset probability samples closely resemble each other. The VAA sample is underrepresented at both ends of the age distribution. The proportion of

respondents 35 years and younger is remarkably lower in the VAA sample than in the probability samples. To the extent that young adults are underrepresented in survey samples generally, this particular VAA sample does not redress that bias. The mean age of respondents in the VAA sample is 52 years, with a standard deviation of 19 years. Among all probability sample respondents, the mean age is 52 years, with a standard deviation of 20 years; among online respondents in the probability sample, the mean age is 50 years, with a standard deviation of 19 years.

Adjusting the three samples using raking weights brings the age distribution of respondents in the VAA sample closer to that in the full probability sample (Figure 2). Younger respondents are still underrepresented in the adjusted VAA sample,

Figure 1 Unadjusted distribution of age among three samples (local regression smoothed)

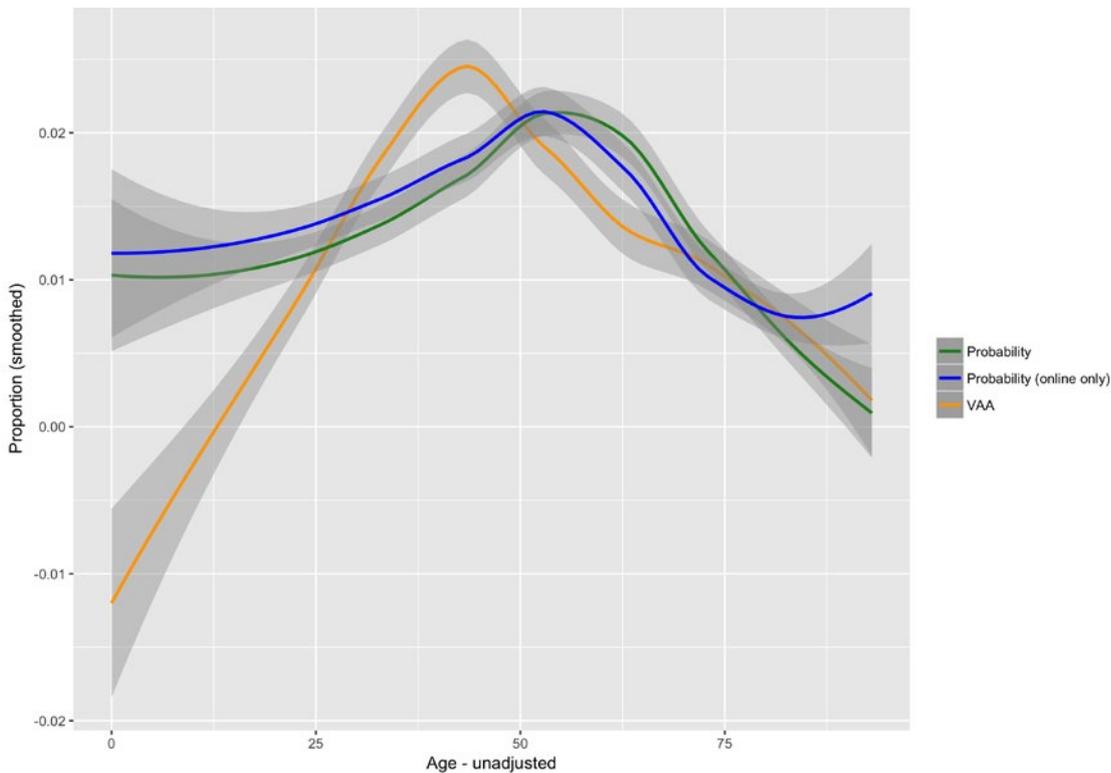
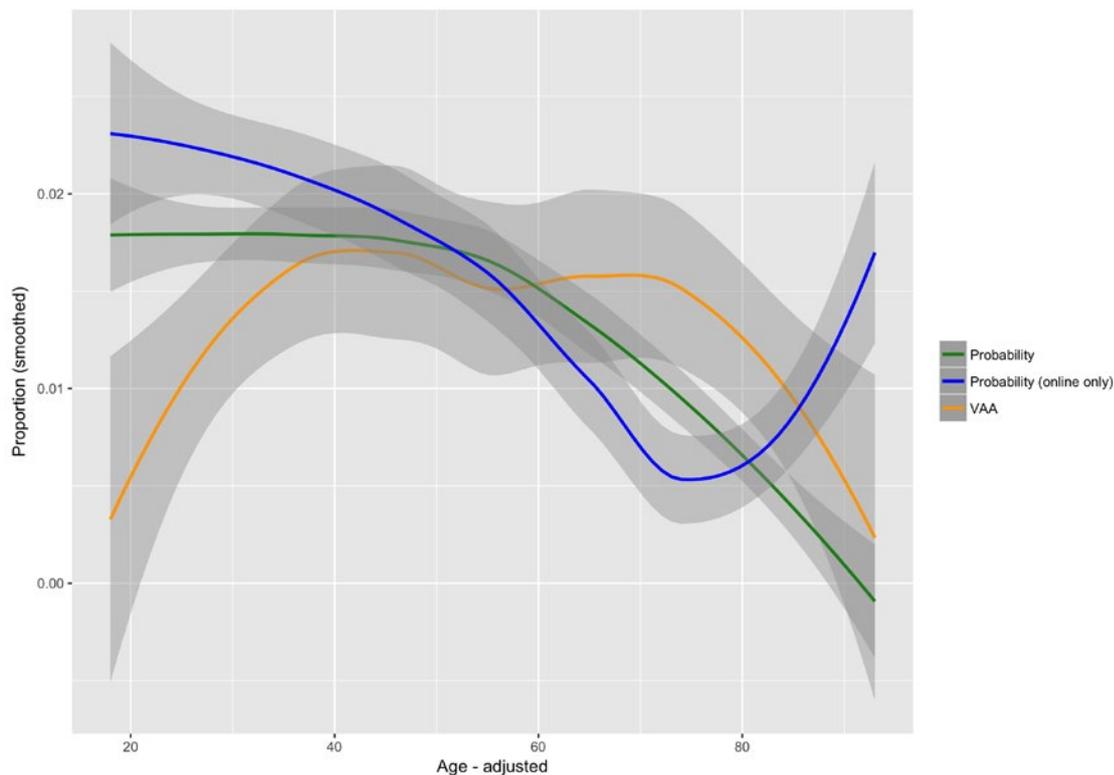


Figure 2 Adjusted distribution of age among three samples (local regression smoothed)



but the subsequent distribution is the closest to normal among the three adjusted samples. Indeed, adjustment through post-stratification appears to make the distribution of the probability-based samples less externally valid; the overrepresentation of older respondents in the online-only probability sample is skewed further after adjustment, while younger respondents are overweighted in the full probability sample. Post-stratification based on the multistage raking weight has the desired effect of increasing external validity of the VAA sample, but fails for the probabilistically sampled respondents. The mean age of the adjusted VAA sample is 56 years (standard deviation of 18 years) – that is, 4 years older than in the unadjusted sample. The mean age of the full probability sample falls by 4 years, to 47 years (standard deviation of 18 years). Among the online subset of the probability sample, the adjusted mean age falls by 8 years, to 42 years (standard deviation of 16 years). Overall, post-stratification of the VAA sample takes the age distribution of that group further away from the probability-based samples, and appears to exacerbate the nonresponse bias among younger Australians in the VAA sample.

The distribution of sex among VAA respondents presents an even tougher challenge to the post-stratification process: only 30% of the unadjusted VAA sample self-reports as being male, compared with 47% of the two probability-based samples (Figure 3). On this measure, the probability-based samples are vastly closer to the Australian population distribution (50% male) than the VAA sample. After post-stratification, each sample accurately reflects the population benchmark, within 1 percentage point (Figure 4). This is achieved in the VAA sample with post-stratification weights trimmed to a range of 0–5, but with high numbers of respondents weighted either out of the sample entirely (14% of the sample) or weighted at the maximum of 5 (14%). The distribution of assigned weights in the probability samples resembles a normal distribution much more closely.

Educational attainment presents another source of bias within the VAA sample (Figure 5). Before adjustment, 66% of VAA respondents report having at least an undergraduate university education (including 33% who report holding a postgraduate qualification). Fewer members of the probability samples (38% of the full sample,

Figure 3 Unadjusted distribution of female and male respondents, by sample

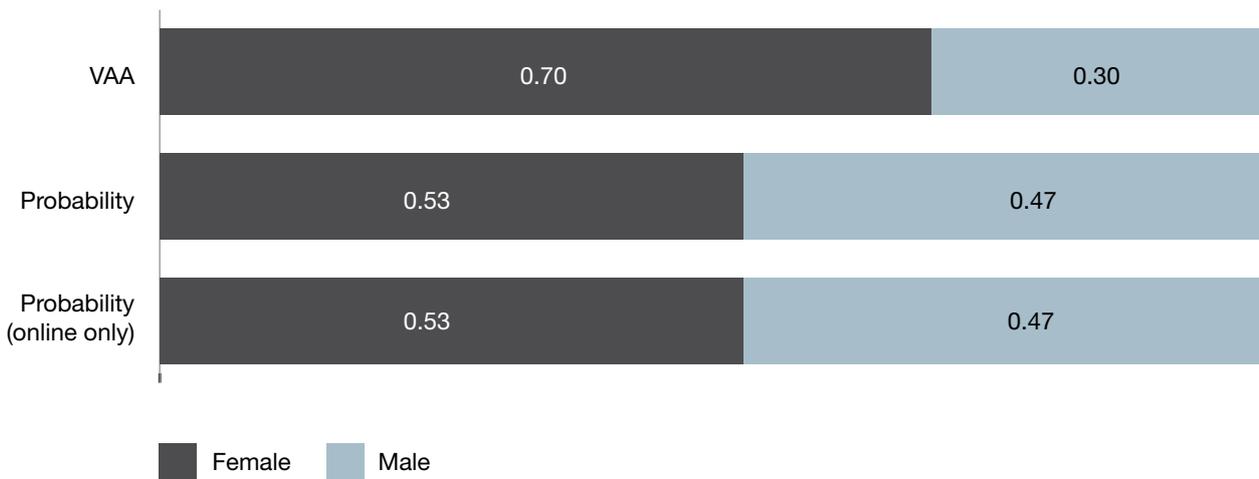


Figure 4 Adjusted distribution of female and male respondents, by sample

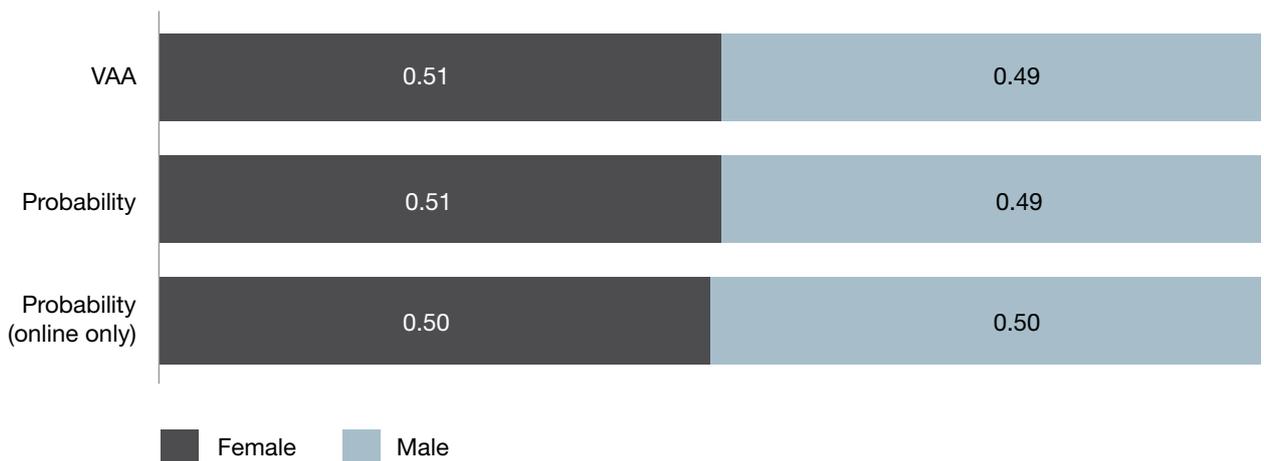
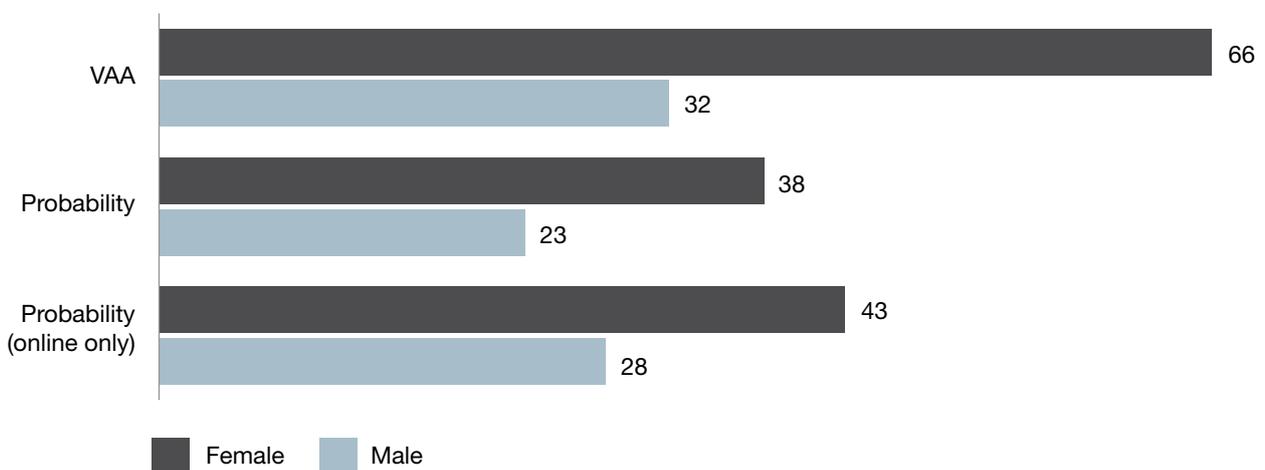


Figure 5 Unadjusted and adjusted frequencies (percentage) of respondents with a bachelor degree or higher, by sample



and 43% of the online-only sample) report having a bachelor degree or higher. With approximately 25% of the Australian population possessing a bachelor degree or higher, each sample reported here overrepresents degree holders, by between 15 percentage points (full probability sample) and 41 percentage points (VAA sample). After adjustment, the full probability sample comes within 2 percentage points of the population frequency, the online-only probability sample within 3 percentage points and the VAA sample within 8 percentage points. However, the adjustment to the VAA sample has moved the percentage of graduates by 34 percentage points, and excluded 2506 respondents from that sample (leaving an effective *n* of only 507). Consequently, adjusting the VAA sample to reflect educational attainment in the population has substantially reduced the quality of the subsequent sample on other measures.

3.2 Attitudinal distributions

Researchers using VAA-generated or similarly skewed data can reasonably argue that, after

adjustments to population benchmarks, the data more closely resemble certain population parameters (usually those specific parameters against which the data are adjusted). What happens, however, to the distribution of those measures for which we have no population parameters? Figure 6 shows the unadjusted distribution of responses to the statement ‘I think that the difference between high and low incomes should be smaller’. VAA respondents are much more likely to ‘strongly agree’ with this statement than respondents in the two probability-based samples, among which the modal response is ‘agree’. At the ‘disagree’ end of the Likert scale, all three samples are closely distributed. After post-stratification, the distribution of responses is practically unchanged (Figure 7); no response category frequency moves by more than 2 percentage points after adjustment. Adjustment on the basis of observable variables does not increase the external validity of the VAA sample, nor does it reduce the nonresponse error in the VAA sample.

Looking at a second social attitude – national pride – reveals a closer distribution between the unadjusted VAA and probability-based

Figure 6 Unadjusted distribution of attitudes to income equality, by sample

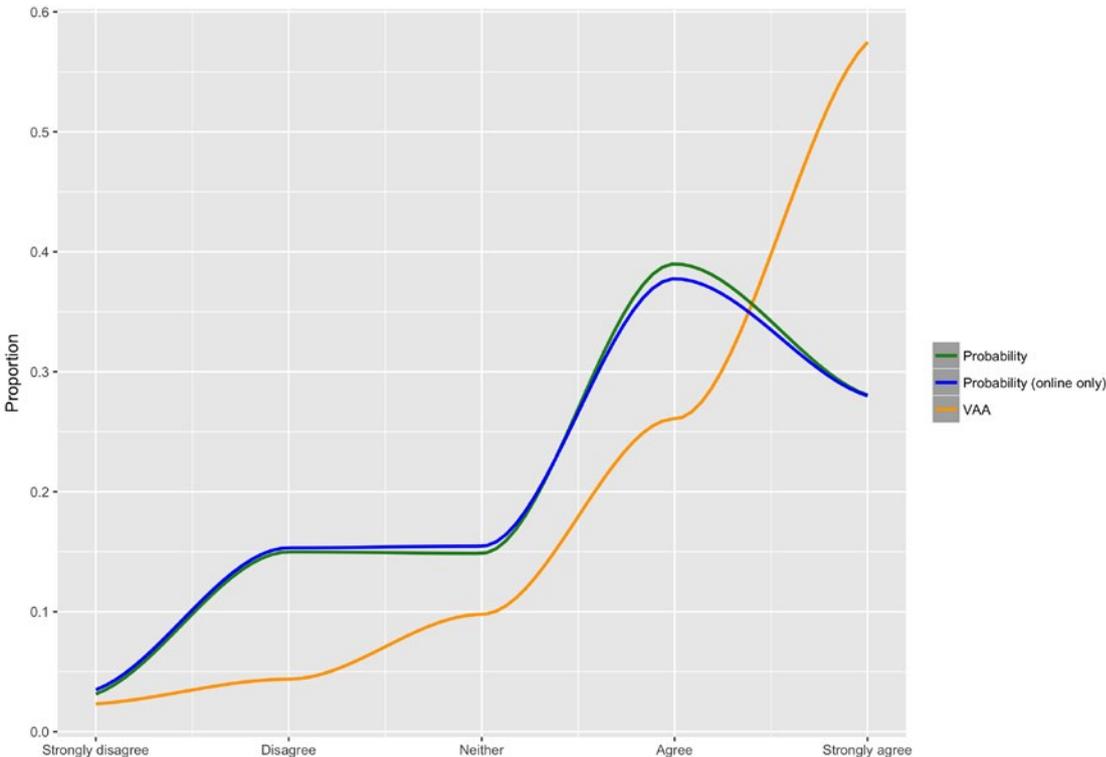


Figure 7 Adjusted distribution of attitudes to redistribution, by sample

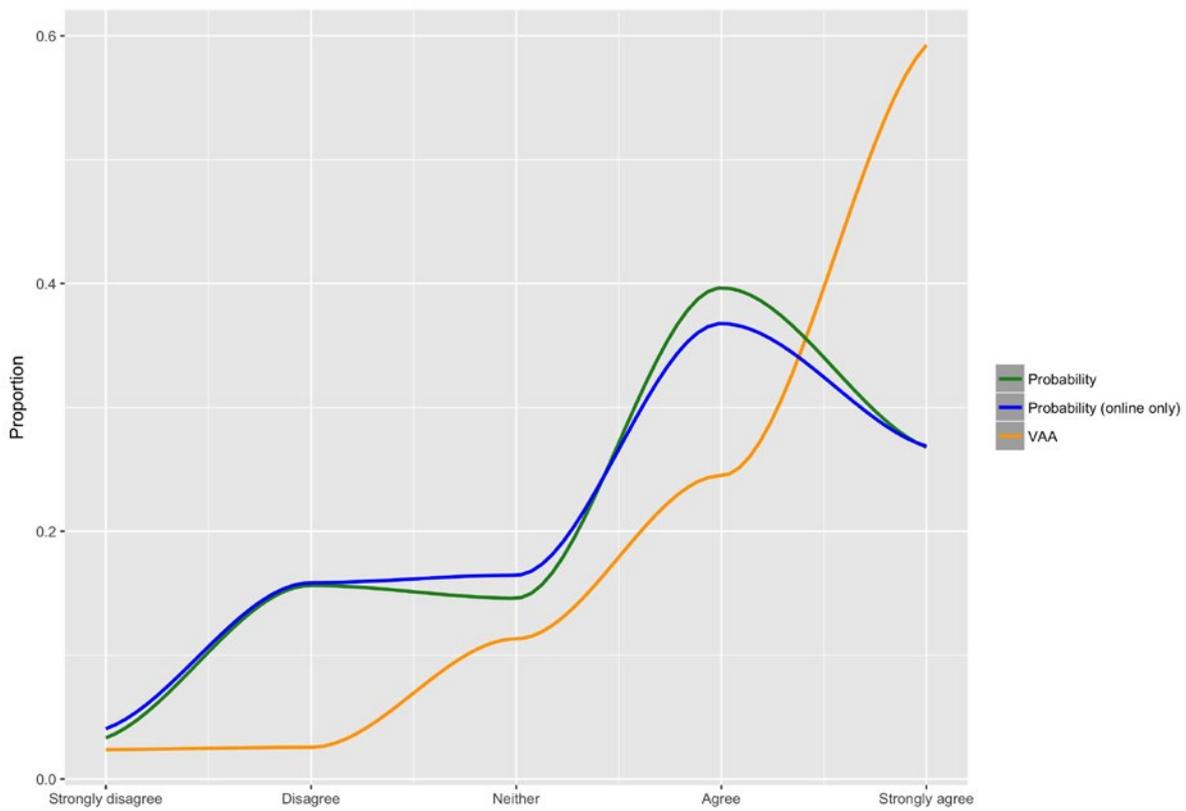


Figure 8 Unadjusted distribution of attitudes to national pride, by sample

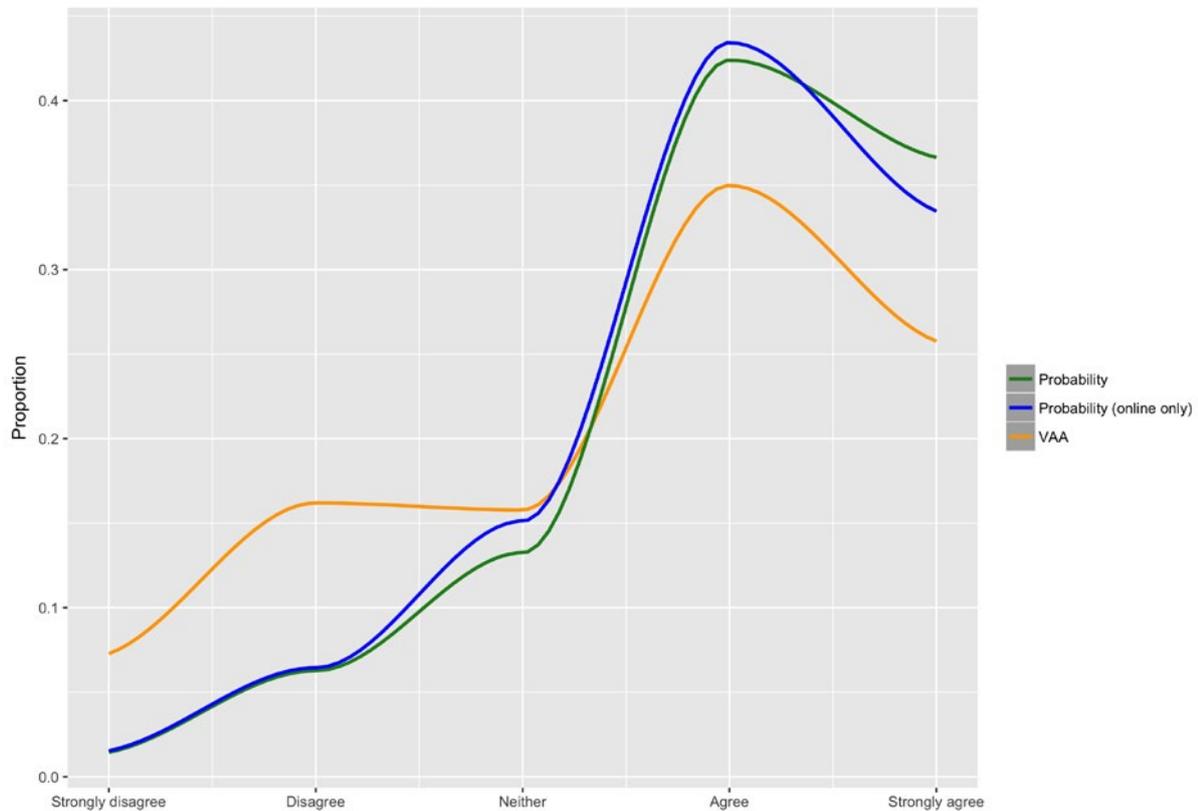
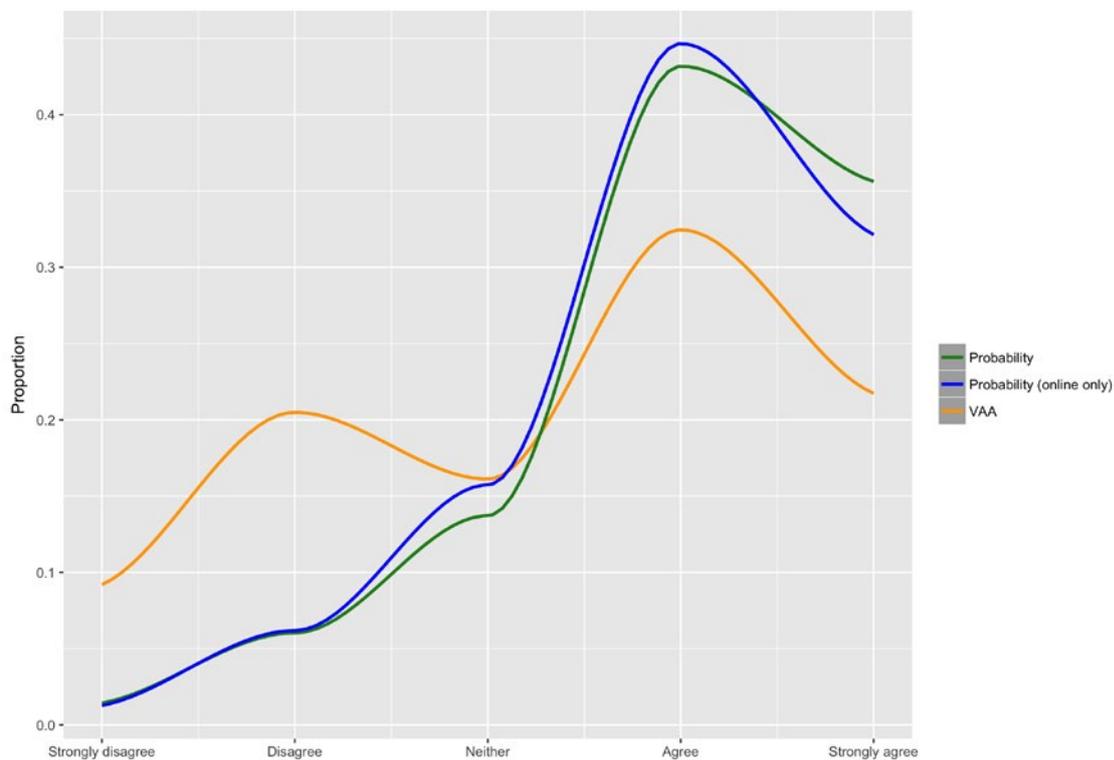


Figure 9 Adjusted distribution of attitudes to national pride, by sample



samples (Figure 8). Although respondents in the VAA sample are more likely to disagree with the statement ‘I am very proud of the country I live in’ than are probability-based sample respondents, the modal category (‘agree’) is consistent across all three samples. Broadly, the shape of the distribution is the same, and, after adjusting the VAA sample data by education and sex, we might expect that the VAA and probability samples would become even more similar. However, the post-stratified distribution in Figure 9 shows that the opposite occurs: adjustment based on observable variables and known population parameters makes the VAA and probability-based samples less similarly distributed. Little changes with regard to the two probability-based samples, whereas respondents answering ‘disagree’ become overrepresented in the VAA sample. As with attitudes towards income equality, post-stratifying the VAA sample data has the effect of both decreasing the external validity and magnifying the nonresponse error in the dataset, compared with the probability-based dataset. However, again, the changes in response category frequencies are small, with

no movement beyond 2 percentage points after adjustment

3.3 Relationships between variables

Researchers are not always concerned with the external validity of their findings; often, we are more interested in the relationships between variables within a dataset, or the internal validity of the data. In this study, we compare the findings from regression models predicting two further measures of social attitudes: the belief that society is changing too fast, and the belief that society has a collective responsibility to provide welfare. Both attitudes are measured as a function of respondents’ sex (a dummy variable in which ‘male’ equals 1) and age. Other sociodemographic variables (e.g. education, occupation, location) are excluded from the analyses because they are measured and scored slightly differently across datasets. Further, the attitudes are modelled using the survey-weighted normal regression function in the Zelig package for R, which assumes that the outcome variable

is normally distributed (Carnes 2017). Clearly, the distributions of responses to the attitudinal measures here are not uniformly normal. However, the internal comparisons are more salient than external reliability in this case, and so the most generally applicable model is used.

Modelling the effects of respondents' age and sex on their agreement with the statement that 'Everything is changing too often and too fast' reveals further substantial differences between the VAA and probability-based samples. Figure 10 shows regression coefficients from the model run on the unadjusted data. The coefficients from the VAA sample show inverse relationships between respondents' sex and age and their attitude towards change from the relationships evident in the two probability-based samples. Among VAA respondents, being male has a positive but insignificant effect on a person's agreement that everything is changing too often and too fast, while age has a negative and significant effect. The two probability-based samples report relationships that are more expected: age has small but positive and significant effects on agreement, and male

respondents are significant less likely to agree that things are changing too fast.

After post-stratifying the data, the coefficients from the VAA sample look closer to what might be expected for attitudes towards change, although age still has a negative and significant effect (Figure 11). The effect of sex has inverted, so that the male binary variable has a negative effect (although insignificant, with very large standard errors). The positive effects of age among members of the probability samples remain significant, but the sex coefficients are effectively nulled by the adjustment. Overall, adjusting the VAA-generated data in an attempt to enhance the external validity of the survey estimates has resulted in small increases in the internal validity of these modelled relationships, but has also increased standard errors and added uninformative noise to the data.

Similarly, results from the models predicting respondents' agreement with the statement that 'All of us together have a responsibility to make sure that everyone has enough money to get by' – a measure of support for redistributive welfare – show that the unadjusted VAA sample data are

Figure 10 Unadjusted regression coefficients predicting agreement that 'Everything is changing too often and too fast'

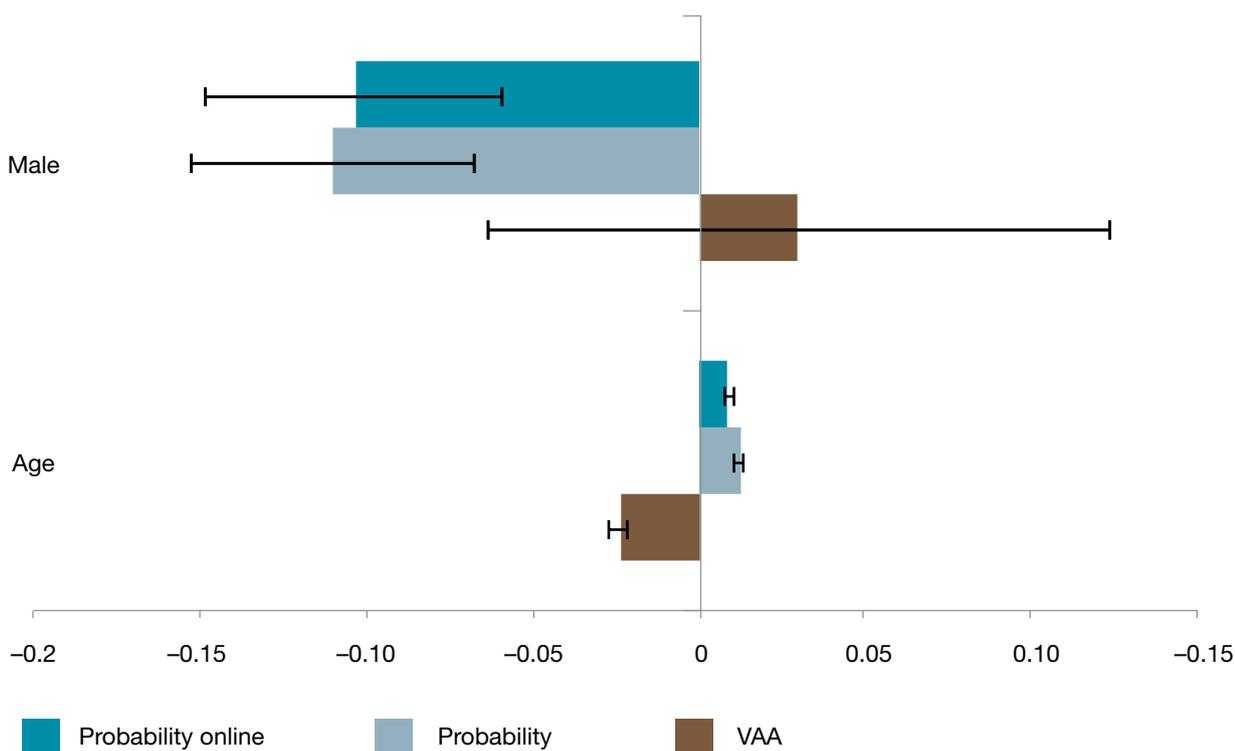
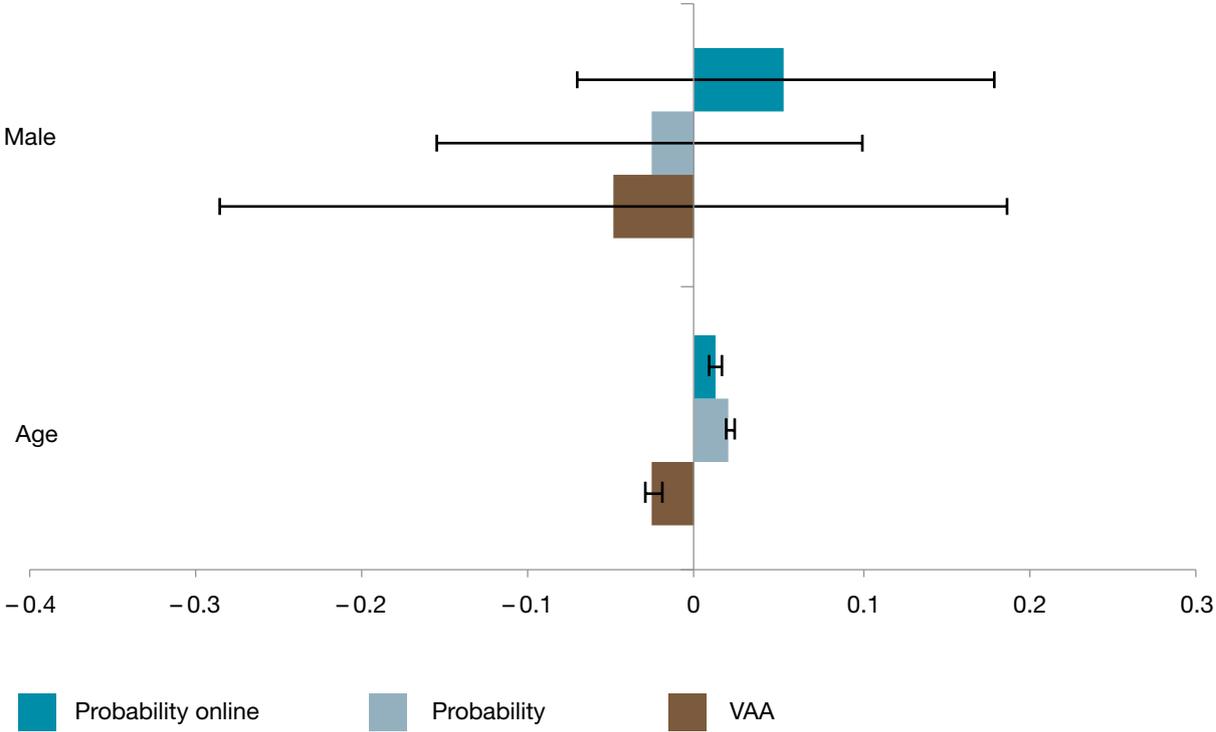


Figure 11 Adjusted regression coefficients predicting agreement that ‘Everything is changing too often and too fast’



starkly divergent from the two probability samples (Figure 12). Among VAA respondents, being male has a strong and positive effect on agreement with this statement, whereas male respondents from the probability samples are less likely than female respondents to agree. Age has very small but positive and significant effects in the probability samples, and a very small, negative effect in the VAA sample. Again, the coefficients have inverse directions between the probability and VAA samples overall.

After adjusting the data against observable population benchmarks, the relationships modelled within the VAA sample data do more closely resemble the more reliable probability sample results (Figure 13). Primarily, the strong positive effect of male sex on agreement that everyone has a responsibility to make sure that others have enough money to get by is decreased, but remains positive and significant. This is in contrast to the existing literature on attitudes towards welfare and redistributive policies, in which women are consistently shown to be more supportive than men (e.g. see Wolbrecht et al. 2008). Although adjustment does bring the VAA sample data

closer to that literature, the results do not cross into positive territory.

Overall, adjustments to the VAA sample data have very minimal effects on the external and internal validity of the relationships modelled here; they are not substantive or remarkable effects.

Figure 12 Unadjusted regression coefficients predicting agreement that ‘All of us together have a responsibility to make sure that everyone has enough money to get by’

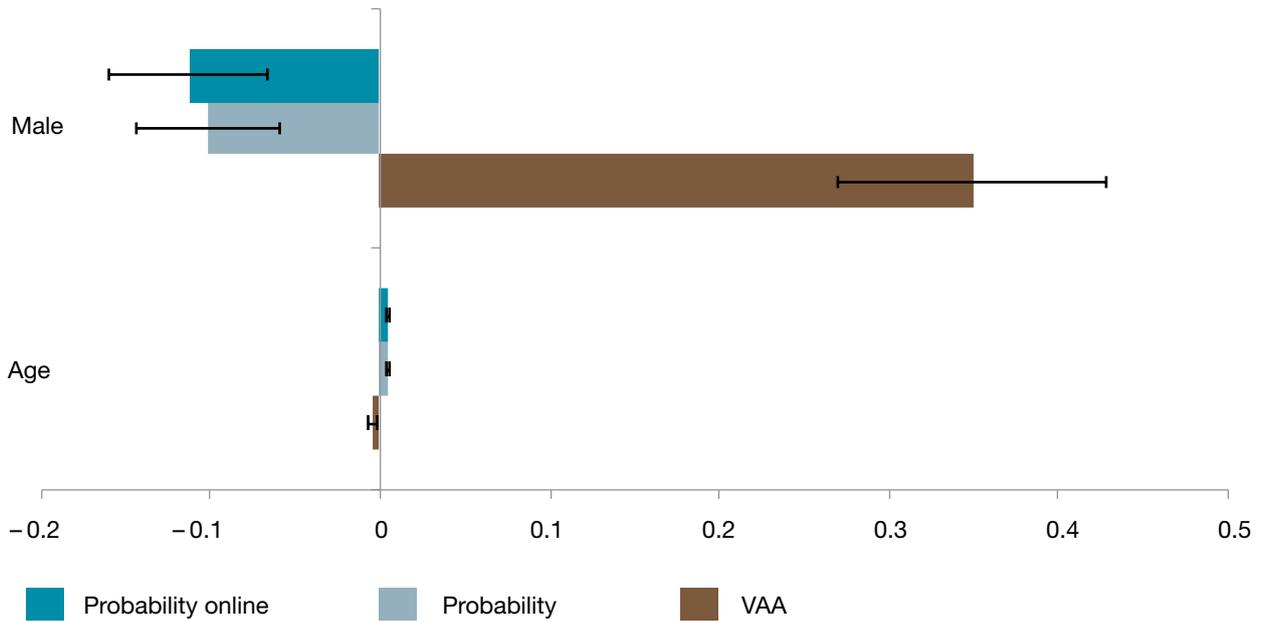
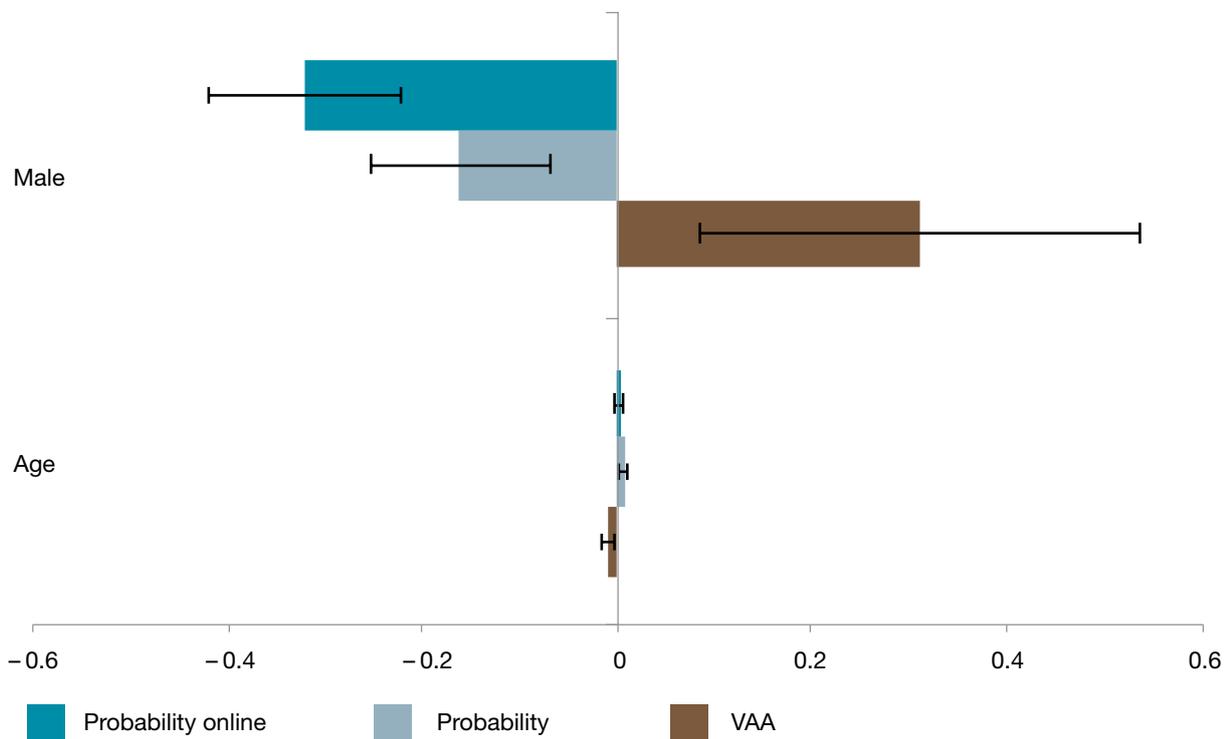
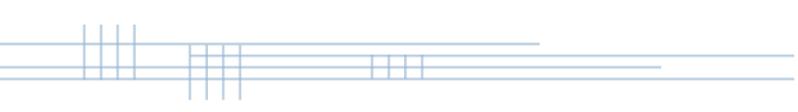


Figure 13 Adjusted regression coefficients predicting agreement that ‘All of us together have a responsibility to make sure that everyone has enough money to get by’





4 Conclusion

Survey researchers face many contemporary challenges: decreasing response rates; increasing costs; and imperatives to look to low-cost incidental, ‘big’, and other forms of noninferential data. Data generated from individuals’ responses to VAAs are slowly but steadily gaining popularity as a source of low-cost, very large n information. This paper argues that VAA-generated data should be subject to the same data quality frameworks as are conventional social surveys – namely the TSE framework (Biemer 2010). By identifying and examining the specific sources of error that can be incurred through the life cycle of any survey, researchers can compare and assess datasets that differ on any number of dimensions. This paper compares types of representation error in two datasets: one composed of a sample of Australian respondents recruited from a larger population of VAAs, and one of Australian adults recruited probabilistically (using random digit dialling). An identical questionnaire was administered to each sample in late 2016.

Focusing on the coverage, sampling, nonresponse and adjustment errors in the two datasets, we found that the VAA sample is substantially less externally valid than the probability-based sample (including both the full and the online-only subset of the sample). The unadjusted (i.e. unweighted) VAA sample data are unrepresentative with regard to sex, age and educational attainment. Adjusting the data against population benchmarks (by respondents’ state of residence, sex and educational attainment) moves the VAA sample closer to these observable population parameters, but excludes a large number of individuals from the analysis. With regard to measures of social attitudes, for which no population parameters are known, the distributions of responses between the VAA and probability-based samples show greater similarities. However, adjusting on the basis of observable population parameters does little to change the distribution of the VAA sample, and in some cases increases the

differences between the VAA and probability sample distributions. Finally, with regard to the relationships between variables, the unadjusted VAA sample reports relationships with inverse effects to the probability-based samples (and to existing literature). Adjustment makes very small improvements on some measures, but introduces additional adjustment error on others.

These findings suggest that researchers analysing data generated from VAA users’ responses to what are ostensibly voter education tools should be both wary and honest about the likelihood of substantial representation error in these data. Caveats that the data have been adjusted according to observable population benchmarks bely the possible introduction of adjustment error in such a process. Although new and more sophisticated methods of post-stratification (including calibration and model weighting) promise some improvements on traditional inverse probability adjustments such as raking weights (Chang & Kott 2008), the absence of knowable population parameters for social attitudes renders any attempt to adjust nonprobabilistic data a vast challenge. Even in the presence of parameters on any given attitude (say, support for same-sex marriage), issue attitudes are more dynamic and inherently less predictable than voting intention or party identification (Wang et al. 2014). On the other hand, the days of reliable, high-response, inexpensive probabilistic survey methods are seemingly behind us, and incidental data provide innumerable opportunities for creative researchers. The findings from this study, alongside others, suggest that we should take heed and be honest about the limitations as we begin to incorporate – or even embrace fully – new forms of incidental survey data.

References

- Andreadis I (2014). Data quality and data cleaning. In: Garzia D & Marschall S (eds), *Matching voters with parties and candidates: voting advice applications in comparative perspective*, ECPR Press, Colchester, 79–91.
- Antonucci L, Horvath L, Kutiyski Y & Krouwel A (2017). The malaise of the squeezed middle: challenging the narrative of the 'left behind' Brexiter. *Competition & Change* 21(3):211–229, <https://doi.org/10.1177/1024529417704135>.
- Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, Dillman D, et al. (2010). *Research synthesis: AAPOR report on online panels*. *Public Opinion Quarterly* 74(4):711–781, <https://doi.org/10.1093/poq/nfq048>.
- Berinsky AJ, Huber GA & Lenz GS (2012). *Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk*. *Political Analysis* 20(3):351–368.
- Biemer PP (2010). Total survey error: design, implementation, and evaluation. *Public Opinion Quarterly* 74(5):817–848, <https://doi.org/10.1093/poq/nfq058>.
- Çarkoğlu A, Vitiello T & Moral M (2012). Voter advice applications in practice: answers to some key questions from Turkey. *International Journal of Electronic Governance* 5(3–4):298–317, <https://doi.org/10.1504/IJEG.2012.051307>.
- Carnes N (2017). Zelig-normal-survey-class: normal regression for continuous dependent variables with survey weights. In: *Zelig: everyone's statistical software*, <https://www.rdocumentation.org/packages/Zelig/versions/5.1.6/topics/Zelig-normal-survey-class>.
- Carson A, Dufresne Y & Martin A (2016). Wedge politics: mapping voter attitudes to asylum seekers using large-scale data during the Australian 2013 federal election campaign. *Policy & Internet* 8(4):478–498, <https://doi.org/10.1002/poi3.128>.
- Chang T & Kott PS (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* 95(3):555–571, <https://doi.org/10.1093/biomet/asn022>.
- Dillman DA, Smyth JD & Christian LM (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*, John Wiley & Sons.
- Dinas E, Trechsel AH & Vassil K (2014). A look into the mirror: preferences, representation and electoral participation. *Electoral Studies* 36(December):290–297, <https://doi.org/10.1016/j.electstud.2014.04.011>.
- Dumont P & Kies R (2012). Smartvote.lu: usage and impact of the first VAA in Luxembourg. *International Journal of Electronic Governance* 5(3–4):388–410, <https://doi.org/10.1504/IJEG.2012.051302>.
- Dutwin D & Buskirk TD (2017). Apples to oranges or Gala versus Golden Delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples. *Public Opinion Quarterly* 81(S1):213–239, <https://doi.org/10.1093/poq/nfw061>.
- Fivaz J & Nadig G (2010). Impact of voting advice applications (VAAs) on voter turnout and their potential use for civic education. *Policy & Internet* 2(4):167–200, <https://doi.org/10.2202/1944-2866.1025>.

- Fossen T & Anderson J (2014). What's the point of voting advice applications? Competing perspectives on democracy and citizenship. *Electoral Studies* 36(December):244–251, <https://doi.org/10.1016/j.electstud.2014.04.001>.
- Gelman A (2007). Struggles with survey weighting and regression modeling. *Statistical Science* 22(2):153–164, <https://doi.org/10.1214/088342306000000691>.
- Gemenis K & Rosema M (2014). Voting advice applications and electoral turnout. *Electoral Studies* 36(December):281–289, <https://doi.org/10.1016/j.electstud.2014.06.010>.
- Groves RM (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 70(5):646–675, <https://doi.org/10.1093/poq/nfl033>.
- & Lyberg L (2010). Total survey error: past, present, and future. *Public Opinion Quarterly* 74(5):849–879, <https://doi.org/10.1093/poq/nfq065>.
- , Fowler FJ Jr, Couper MP, Lepkowski JM, Singer E & Tourangeau R (2009). *Survey methodology*, John Wiley & Sons.
- Johnston R (2017). Vote Compass in British Columbia: insights from and about published polls. *Journal of Elections, Public Opinion and Parties* 27(1):97–109, <https://doi.org/10.1080/17457289.2016.1269112>.
- Kamoen N, Holleman B, Krouwel A, Van de Pol J & de Vreese C (2015). The effect of voting advice applications on political knowledge and vote choice. *Irish Political Studies* 30(4):595–618, <https://doi.org/10.1080/07907184.2015.1099096>.
- Kleinnijenhuis J, Van de Pol J, van Hoof A & Krouwel A (2017). VAAs as sources of volatility and fragmentation: self-selection effects and genuine effects. *Journal of Elections, Public Opinion and Parties* 27(1):75–96, <https://doi.org/10.1080/17457289.2016.1268143>.
- Lees-Marshment J, Dufresne Y, Eady G, Osborne D, van der Linden C & Vowles J (2015). Vote Compass in the 2014 New Zealand election: hearing the voice of New Zealand voters. *Political Science* 67(2):94–124, <https://doi.org/10.1177/0032318715609076>.
- Loewen PJ, Heroux-Legault M & de Miguel C (2015). Nationalism and ethnic heterogeneity: the importance of local context for nationalist party vote choice. *Electoral Studies* 39(September):129–141, <https://doi.org/10.1016/j.electstud.2015.04.005>.
- Marschall S & Schultze M (2012). Voting advice applications and their effect on voter turnout: the case of the German Wahl-O-Mat. *International Journal of Electronic Governance* 5(3–4):349–366, <https://doi.org/10.1504/IJEG.2012.051314>.
- & Schultze M (2015). German e-campaigning and the emergence of a 'digital voter'? An analysis of the users of the Wahl-O-Mat. *German Politics* 24(4):525–541, <https://doi.org/10.1080/09644008.2014.949681>.
- Massey DS & Tourangeau R (eds) (2012). *The nonresponse challenge to surveys and statistics*, SAGE Publications, Los Angeles.
- Mendez F (2012). Matching voters with political parties and candidates: an empirical test of four algorithms. *International Journal of Electronic Governance* 5(3–4):264–278, <https://doi.org/10.1504/IJEG.2012.051316>.
- (2017). Modeling proximity and directional decisional logic: what can we learn from applying statistical learning techniques to VAA-generated data? *Journal of Elections, Public Opinion and Parties* 27(1):31–55, <https://doi.org/10.1080/17457289.2016.1269113>.

- & Wheatley J (2014). Using VAA-generated data for mapping partisan supporters in the ideological space. In: Garzia D & Marschall S (eds), *Matching voters with parties and candidates: voting advice applications in comparative perspective*, ECPR Press, Colchester, 161–196.
- Pennay D, Neiger D, Sebastian M, Honey N & Lavrakas PJ (2015). *Online Panels Benchmarking Study, 2015*, Social Research Centre, Melbourne, <https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.4225/87/FSOYQI>.
- Pianzola J (2014). Selection biases in voting advice application research. *Electoral Studies* 36(December):272–280, <https://doi.org/10.1016/j.electstud.2014.04.012>.
- Rosema M, Anderson J & Walgrave S (2014). The design, purpose, and effects of voting advice applications. *Electoral Studies* 36(December):240–243, <https://doi.org/10.1016/j.electstud.2014.04.003>.
- Schultze M (2014). Effects of voting advice applications (VAAs) on political knowledge about party positions. *Policy & Internet* 6(1):46–68, <https://doi.org/10.1002/1944-2866.POI352>.
- Stoker L & Bowers J (2002). Designing multi-level studies: sampling voters and electoral contexts. *Electoral Studies* 21(2):235–267, [https://doi.org/10.1016/S0261-3794\(01\)00021-X](https://doi.org/10.1016/S0261-3794(01)00021-X).
- Talonen J & Sulkava M (2011). Analyzing parliamentary elections based on voting advice application data. In: *Advances in intelligent data analysis X*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 340–351, https://doi.org/10.1007/978-3-642-24800-9_32.
- Van Camp K, Lefevere J & Walgrave S (2014). The content and formulation of statements in voting advice applications: a comparative analysis of 26 VAAs. In: Garzia D & Marschall S (eds), *Matching voters with parties and candidates: voting advice applications in comparative perspective*, ECPR Press, Colchester, 11–31, <http://dare.uva.nl/search?metis.record.id=416380>.
- Van de Pol J, Holleman B, Kamoen N, Krouwel A & de Vreese C (2014). Beyond young, highly educated males: a typology of VAA users. *Journal of Information Technology & Politics* 11(4):397–411, <https://doi.org/10.1080/19331681.2014.958794>.
- van Prooijen J-W & Krouwel APM (2017). Extreme political beliefs predict dogmatic intolerance. *Social Psychological and Personality Science* 8(3):292–300, <https://doi.org/10.1177/1948550616671403>.
- van Veen F, Göritz AS & Sattler S (2016). Response effects of prenotification, prepaid cash, prepaid vouchers, and postpaid vouchers. *Social Science Computer Review* 34(3):333–346, <https://doi.org/10.1177/0894439315585074>.
- Wagner M & Ruusuvirta O (2012). Matching voters to parties: voting advice applications and models of party choice. *Acta Politica* 47(4):400–422, <https://doi.org/10.1057/ap.2011.29>.
- Walgrave S, van Aelst P & Nuytemans M (2008). ‘Do the vote test’: the electoral effects of a popular vote advice application at the 2004 Belgian elections. *Acta Politica* 43(1):50–70, <https://doi.org/10.1057/palgrave.ap.5500209>.
- Wall M, Krouwel A & Vitiello T (2014). Do voters follow the recommendations of voter advice application websites? A study of the effects of Kieskompas.nl on its users’ vote choices in the 2010 Dutch legislative elections. *Party Politics* 20(3):416–428, <https://doi.org/10.1177/1354068811436054>.

- Wang W, Rothschild D, Goel S & Gelman A (2014). Forecasting elections with non-representative polls. *International Journal of Forecasting* 31(3):980–991, <https://doi.org/10.1016/j.ijforecast.2014.06.001>.
- Weisberg HF (2009). *The total survey error approach: a guide to the new science of survey research*, University of Chicago Press.
- Wolbrecht C, Beckwith K & Baldez L (eds) (2008). *Political women and American democracy*, Cambridge University Press.
- 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, <https://doi.org/10.1093/poq/nfr020>.

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

CRICOS PROVIDER NO. 00120C