Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election

October 2020

Corresponding author: darren.pennay@srcentre.com.au
Media enquiries: rochelle@thirdavenue.com.au
Industry enquiries: sarah@amsro.com.au
Report prepared by:

Polling Inquiry Committee
Darren Pennay, The Social Research Centre, A subsidiary of the Australian National University
Emeritus Professor Murray Goot, Macquarie University
Dr Phil Hughes, Engine Australia
Dr Dina Neiger, The Social Research Centre, A subsidiary of the Australian National University
Dr Jill Sheppard, Australian National University
John Stirton, Independent Consultant

Advisory Group
Dennis Trewin, AO, Independent Consultant and Former Australian Statistician and Australian Electoral Commissioner
Dr John Henstridge, Data Analysis Australia / Statistical Society of Australia
Professor Ian McAllister, Australian National University
Kerry O’Brien, Journalist
Travyn Rhall, Independent Consultant

International Advisers
Dr Paul J. Lavrakas, Independent Consultant, Past AAPOR President
Dr Patrick Moynihan, Pew Research Center
Professor Kristen Olson, University of Nebraska-Lincoln
Professor Patrick Sturgis, London School of Economics


Contents

Contents ........................................................................................................................................ iii
List of figures ................................................................................................................................. v
List of tables ................................................................................................................................. vi
Foreword ......................................................................................................................................... vii
Acknowledgements ......................................................................................................................... ix
Acronyms ........................................................................................................................................ x
Executive Summary ......................................................................................................................... xi

1. Introduction ............................................................................................................................... 1
   1.1. Background .......................................................................................................................... 1
   1.2. How this Inquiry was conducted ....................................................................................... 2
   1.3. About this report .................................................................................................................. 3
   1.4. Definition of election and other political polling .............................................................. 4
   1.5. Caveats and limitations ...................................................................................................... 5

2. The importance of election polling ......................................................................................... 6

3. The election polling environment in Australia ...................................................................... 7
   3.1. Context ............................................................................................................................. 7
   3.2. Recent changes in survey research .................................................................................... 7

4. How election polls are conducted ......................................................................................... 10
   4.1. A variety of methods ......................................................................................................... 10
   4.2. A hierarchy of survey methods ......................................................................................... 13
   4.3. Polling practices at the 2019 election ................................................................................. 14

5. The performance of the polls ................................................................................................. 18
   5.1. How often do the election polls ‘get it right’? ................................................................. 18
   5.2. 2019 poll results ............................................................................................................... 20
   5.3. Absolute error .................................................................................................................. 22
   5.4. Margin of sampling error ................................................................................................. 24
   5.5. Performance of the polls over a longer time frame .......................................................... 27
   5.6. International Comparisons ............................................................................................... 28
   5.7. The performance of the polls by methodology ................................................................. 30
   5.8. Is there partisan bias in the polls? .................................................................................... 31

6. Factors that may have contributed to inaccuracies in the 2019 election polls .................... 33
   6.1. Commentary from the pollsters ......................................................................................... 33
   6.2. Analytical framework: Total Survey Error ...................................................................... 34
   6.3. Possible measurement errors ........................................................................................... 35
      6.3.1. How voting intentions are measured ....................................................................... 35
      6.3.2. Ballot order effects .................................................................................................... 41
      6.3.3. Device effects ............................................................................................................ 42
      6.3.4. Were pollsters deliberately misled by respondents? .............................................. 42
      6.3.5. Shy conservatives .................................................................................................... 42
      6.3.6. Early voting .............................................................................................................. 43
      6.3.7. Late deciders ............................................................................................................ 45
   6.4. Possible errors of representation ...................................................................................... 48
6.4.1. Coverage ........................................................................................................................................48
6.4.2. Sample design and sample size ........................................................................................................50
6.4.3. Non-response ....................................................................................................................................52
6.4.4. Weighting ..........................................................................................................................................57
6.5. The convergence of the polls .................................................................................................................61
   6.5.1. Herding ...........................................................................................................................................61
   6.5.2. Herding by suppression .................................................................................................................62
   6.5.3. Was herding a factor in the performance of the polls in 2019? ................................................63
7. Setting standards for election polling .........................................................................................................64
8. Reporting the polls ....................................................................................................................................67
   8.1. The polls are big news .......................................................................................................................67
   8.2. Disclosure standards and the reporting of the polls ........................................................................67
9. Our expectations of election polling .........................................................................................................70
   9.1. The perils of prediction .....................................................................................................................70
   9.2. Developing more reasonable expectations of the polls .....................................................................71
10. Findings and recommendations .............................................................................................................72
    10.1. Polling ‘miss’ or polling ‘failure’? ....................................................................................................72
    10.2. Factors that contributed to the failure of the published national election polls in 2019 73
       10.2.1. Contextual factors ....................................................................................................................73
       10.2.2. Statistical and methodological factors ......................................................................................73
    10.3. Reporting the polls ........................................................................................................................75
    10.4. Recommendations ........................................................................................................................75
Appendix 1: Terms of reference ......................................................................................................................77
Appendix 3: Technical notes on the calculation of error metrics ..................................................................79
Appendix 4: Error estimates in the final polls 1993-2019 ...........................................................................83
Appendix 5: Total Survey Error framework ................................................................................................84
Appendix 6: Weighting case studies ............................................................................................................86
Appendix 7: The convergence of the polls – what are the odds? .................................................................94
Appendix 8: International and Australian standards ..................................................................................95
Appendix 9: Source materials – primary reporting of the polls ................................................................110
Appendix 10: Recommended disclosure standards for Australian election polls ..................................112
References .....................................................................................................................................................114
List of figures

Figure 1: ‘Typical’ response rates for telephone surveys in Australia, 2009 – 2019. ................................. 9
Figure 2: MoSE 2PP LNP Vote ...................................................................................................................... 26
Figure 3: MoSE LNP Primary Vote ................................................................................................................. 26
Figure 4: MoSE LABOR Primary Vote ........................................................................................................... 26
Figure 5: MoSE Green Primary Vote ............................................................................................................... 26
Figure 6: Australian polls, Mean Absolute Error, Primary Vote – All parties with ≥3% of the vote, 1980 – 2019*. ............................................. 27
Figure 7: Australian polls, Mean Absolute Error on the Margin, 1980 – 2019** ................................................. 28
Figure 8: International polls, Mean Absolute Error, Primary Vote – All parties with ≥3% of the vote, 1980 – 2019*. ........................................... 29
Figure 9: International polls, Mean Absolute Error on the Margin, 1980 – 2019*. ............................................. 30
Figure 10: Pre-poll voting over time, 1993 to 2019.......................................................................................... 44
Figure 11: Cumulative early voting patterns by day, 2013, 2016 and 2019....................................................... 45
Figure 12: Level of interest in politics by party voted for, unweighted data (%)................................................. 55
Figure 13: Education and vote choice at the 2019 election, unweighted data (%).......................................... 56
Figure 14: The evolution of polling methods in Australia, 1993 – 2019 ......................................................... 78
Figure 15: Total Survey Error ........................................................................................................................ 85
Figure 16: Comparison of weighting solutions impact on measures of absolute error ................................. 91
List of tables

Table 1: Polling methods used for public release national election polls in 2019 ................................................. 15
Table 2: Number of final polls ‘getting it right’, 1993-2019(a) .......................................................................................... 20
Table 3: Final national polls for the House of Representatives, 2019 ................................................................. 21
Table 4: Absolute error (percentage points) ...................................................................................................................... 22
Table 5: Summary of the polls performance by error metric, 2019 .................................................................................. 24
Table 6: Margins of Sampling Error for LNP, ALP and Greens ...................................................................................... 25
Table 7: Summary of poll performance by methodology, 1993 to 2019 ......................................................................... 31
Table 8: How each pollster measures voting intentions ............................................................................................... 35
Table 9: How each pollster derived their two-party-preferred estimate ........................................................................... 38
Table 10: Preference flows at the 2019 and 2016 elections ......................................................................................... 39
Table 11: A comparison of primary vote and two-party-preferred errors, in percentage points ................................. 40
Table 12: Number and proportion of seats in which Labor or the Coalition drew first position ..................................... 41
Table 13: Number and proportion of seats in which Labor or the Coalition drew a higher ballot position ......................... 41
Table 14: Estimated Margin of Error given a 2PP gap of 3 percentage points .............................................................. 51
Table 15: Estimated margins of error for a single poll and two independent polls ........................................................... 51
Table 16: Primary reports that met basic disclosure requirements .................................................................................. 68
Table 17: The number of ‘primary reports’ that met basic disclosure requirements by pollster and media/reporting outlet ......................................................................................................................... 69
Table 18: Two-party-preferred vote by seats won and election winner, 1993 – 2019 ....................................................... 70
Table 19: MAE calculations example, Australia 2019 .............................................................................................. 82
Table 20: Error in estimates of the Liberal-National Party share of the two-party-preferred vote, final polls, 1993-2019 (percentage points) ...................................................................................... 83
Table 21: The impact of weighting by educational attainment on the vote choice estimates produced from the CSES Australia survey, 2019 ................................................................................. 86
Table 22: The impact of weighting by educational attainment on the vote choice estimates produced from the 2019 AES ........................................................................................................... 88
Table 23: Summary of the impact of various weighting solutions used for the CSES Australia survey on selected error metrics ................................................................................................................. 90
Table 24: Population benchmarks used for weighting the CSES Australia and AES case studies ........................................ 92
Table 25: Comparison of selected international, national and Australian disclosure standards for election polling .............................................................................................................................. 107
Table 26: Source materials used to analysis the disclosure standards of the primary reporting of the polls ......................... 110
Foreword

The result of the Australian Federal Election on 18 May 2019 caught most observers by surprise – the key reason being the mismatch between the election result and the estimates of the election polls conducted by Australia’s leading polling organisations over the preceding three years.

The Association of Market and Social Research Organisations (AMSRO), as the main industry body representing Australian research, data and insights organisations (and in the absence of any other representative organisation of polling companies in Australia), felt it was important to understand the nature of, and reasons for, the ‘polling miss’. As a result, on the Monday after the election, AMSRO announced the establishment of this Inquiry. The Statistical Society of Australia (SSA) was also quick to announce their own investigation. The SSA process was quickly joined with that of AMSRO and this Inquiry is the result. We are grateful for the contribution of the SSA’s members.

Key to the credibility of this Inquiry was establishing a process that was independent, impartial, robust and comprehensive. The independence and impartiality of the Inquiry were seen as critical in encouraging the polling companies to participate, given that most of the polling companies were not AMSRO members at the time. The Inquiry Panel comprised technical specialists in polling, sampling, statistics, political science and related fields. An independent Advisory Board was also established to provide suggestions and advice to the Inquiry Panel. The Advisory Board comprised senior figures with a range of relevant backgrounds and included experts in official statistics, political science, commercial market research and the media. The Advisory Board was chaired by Dennis Trewin, AO, a former Australian Statistician (head of the Australian Bureau of Statistics) and former Australian Electoral Commissioner.

None of the members of the Inquiry Panel or Advisory Board worked for organisations currently involved in pre-election polling. This was to assure the polling companies that their competitors were not judging them and to encourage their participation in the Inquiry. The Inquiry was funded by, and periodically reported to, AMSRO but otherwise operated entirely independently. The Terms of Reference were determined by AMSRO and included examining the accuracy of the polls, determining the cause/s of any inaccuracies, and suggesting improvements to how polls are conducted and how their results are communicated to the public.

Darren Pennay, founder and immediate past CEO of the Social Research Centre (SRC), a subsidiary of the Australian National University (ANU), agreed to chair the Panel and led five other members with relevant experience, all of whom are highly regarded in their fields. We were also fortunate to access some of the best-qualified international experts in polling and survey statistics, including those who had worked on similar polling inquiries in the US and the UK.

An earlier Discussion Paper – Disclosure standards for election and political polling in Australia (May 2020) – preceded the release of this final report. The key recommendation from the Discussion Paper was that a new and comprehensive minimum set of disclosure standards needed to be developed.

AMSRO will closely consider the recommendations of this final Report, and is committed to working with all organisations that conduct or commission election polls, and other stakeholders, to improve standards and transparency in polling and the accuracy with which poll results are communicated to the public. In the meantime, we hope that readers of this report will find it sheds some light on the nature and causes of the ‘polling miss’ at the 2019 Federal Election, and that it ultimately helps polling companies in their collective efforts to improve the accuracy and transparency of election polling in Australia. This is an important endeavour because election polls play an important and prominent role in underpinning a modern, well-informed democracy.

AMSRO and the SSA are exceedingly grateful to all those who have volunteered their time and expertise to undertake this Inquiry – by serving on the Inquiry Panel or Advisory Board – and to the
polling companies that cooperated with the Inquiry. In particular, we are immensely thankful to Darren Pennay for his leadership of the Inquiry Panel and his commitment to seeing this process through and to the SRC for their support and administrative assistance throughout.

AMSRO is also indebted to Craig Young (former AMSRO President 2017-2019) for having the foresight to see how important it was to undertake an independent review of the Australian election polls in the aftermath of the 2019 Federal Election. He moved swiftly to establish the Inquiry and was instrumental in assembling the Inquiry Panel, its Advisory Board and its terms of reference, and has been a tenacious advocate and supporter of the Inquiry’s work and its independence.

George Zdanowicz
AMSRO President
October 2020
Acknowledgements

The work on this Inquiry was conducted on a voluntary basis by the members of the Inquiry Panel and Advisory Group and would not have been possible without the generous support of the following organisations and individuals:

- The pollsters who contributed to this Inquiry.
- The employers of the Inquiry Panel, Advisory Board and International Advisers.
- The Social Research Centre, in particular, for providing the research environment, in-kind administrative and technical support and time of three Inquiry members (Darren Pennay (chair), Dina Neiger, and Paul Lavrakas); as well as via:
  - Providing meeting space to the Inquiry.
  - Bianca Smith assisting with the formatting and desktop publishing of the Inquiry reports.
  - Jack Barton for running the various weighting scenarios to illustrate alternative weighting schemes. (See Appendix 6 – Weighting Case Studies.)
  - Wendy Guo for setting up the coding tool to enable the coding of the 16 original source documents and primary media primary reports used in section 8 and Stephen Cuttriss for conducting qualitative data coding.
  - Access to Life in Australia™ data and Charles Dove for assisting with access and use of the data.
- Sarah Campbell, Executive Director, AMSRO.
- The Australian Election Study founder, Professor Ian McAllister, and the Australian Election Study Team for making the data from the 2019 Australian Election Study and the 2019 Comparative Study of Electoral Systems Survey (Australia) available through the Australian Data Archives.
- The Australian National University for funding Grace Sixsmith to undertake associated Research Assistant tasks.
Acronyms

AAPOR – American Association for Public Opinion Research
ABC – American Broadcasting Company
ABS – Australian Bureau of Statistics
ACMA – Australian Communication and Media Authority
AEC – Australian Electoral Commission
AES – Australian Election Study
ALP – Australian Labor Party
AMSRO – Association of Market and Social Research Organisations
APC – Australian Press Council
AStat – Accredited Statistician as recognised by Statistical Society of Australia
BPC – British Polling Council
CATI – Computer Assisted Telephone Interviewing
CNN – Cable News Network
CRIC – Canadian Research Insights Council
CSES – Comparative Study of Electoral Systems
CStat – Chartered Statistician
DFRDD – Dual-frame Random Digit Dialling
GRN – The Greens
IPND – Integrated Public Number Database
IVR – Interactive Voice Recognition
ISO – International Standards Organisation
LNP – Liberal National Party coalition (‘the Coalition’)
MoE/MoSE – Margin of Error/Margin of Sampling Error
NCPP – National Council for Public Polls
PHON – Pauline Hanson’s One Nation (“One Nation”)
QPR – Qualified Professional Researcher
RANZ – Research Association of New Zealand
RS – The Research Society
SRC – Social Research Centre
SSA – Statistical Society of Australia
TSE – Total Survey Error
UAP – United Australia Party
WAPOR – World Association for Public Opinion Research
2PP – Two-party-preferred
Executive Summary

All of the national election polls published during the 2019 election campaign purported to show that Labor had the support of the majority of Australian voters in terms of the two-party-preferred vote. The Coalition went on to win the election with 51.5% of the vote compared to Labor with 48.5%, almost the mirror opposite of what the final polls found; all missing the result in the same direction and by a similar margin. After the election, the polls attracted widespread criticism. There was also considerable hand-wringing within the polling community.

This Inquiry was established to undertake an independent, objective and non-partisan review of what, if anything, had gone wrong with the national polls and what improvements could be made. The Inquiry sought the cooperation of the four pollsters responsible for these polls: Essential Research, Ipsos, Roy Morgan Research and YouGov (responsible for both YouGov Galaxy and Newspoll). Lonergan Research was also approached to provide some insights into polling via the use of Interactive Voice Recognition (IVR) technology (i.e. robopolls) and because it went public with a claim of being involved with the suppression of a poll that was out of alignment with other polls at the time, a form of ‘herding’. Informal discussions were held with some pollsters from outside this group and all pollsters and other interested parties were able to make submissions to the Inquiry via the public submissions process established by AMSRO. The governance structure for the Inquiry comprised the Inquiry Panel itself, an Advisory Board and International Advisers.

While the Inquiry Panel is grateful for the cooperation it received, the lack of access to data sets and detailed descriptions of the survey methods and statistical techniques used by the pollsters materially affected our ability to identify the specific factors that contributed to the inaccuracy of the polls. We have had to supplement the information provided by the pollsters with incomplete publicly available methodological descriptions, compare against best practice and draw inferences based on our analyses of overseas inquiries and other data. The information provided by the pollsters to this Inquiry was much less than was provided to similar inquiries conducted into the performance of recent election polls in the US and the UK. This came as somewhat of a shock to our International Advisers.

The performance of the national polls in 2019 met the independent criteria of a ‘polling failure’ not just a ‘polling miss’. The polls: (1) significantly—statistically—erred in their estimate of the vote; (2) erred in the same direction and at a similar level; and (3) the source of error was in the polls themselves rather than a result of a last-minute shift among voters.

We rule out as contributing factors to the poor performance of the polls not only a late swing after the final polls were conducted (except possibly to a very minor extent); but also the impact of ‘shy conservatives’, measurement error arising from the voting intentions questions, respondents deliberately misleading pollsters, early voting, and ballot order effects. The allocation of preferences led to a slight increase in overall poll error in some estimates of the two-party-preferred vote, but was not a major contributing factor to the failure overall.

The Inquiry Panel could not rule out the possibility that the uncommon convergence of the polls in 2019 was due to herding. This could not be ruled out largely because, despite our requests, the pollsters provided no raw data to enable us to attempt to replicate their results.

Our conclusion is that the most likely reason why the polls underestimated the first preference vote for the LNP and overestimated it for Labor was because the samples were unrepresentative and inadequately adjusted.

- The polls were likely to have been skewed towards the more politically engaged and better educated voters with this bias not corrected.
- As a result, the polls over-represented Labor voters.
Such a skew has been evident in recent election cycles, with 17 of the 25 final poll results since 2010 (68%) overestimating 2PP support for Labor.

This finding stands independent of methodology because even though the methods used by the pollsters differ they share a common difficulty in struggling to establish contact with and gain the cooperation of a representative sample of voters. This conclusion is broadly similar to that reached by the reviews into the performance of the 2015 UK polls and the 2016 US polls.

The Inquiry Panel also found some evidence that the reporting of the polls failed to consistently meet the basic disclosure guidelines for editors and journalists set out by the Australian Press Council. The view of the Inquiry Panel is that the various reporting and disclosure standards as they apply to the publicly released election polls in Australia are in need of consolidation and clarification. This would best be achieved by putting in place a new code and new governance structure. We concur with the conclusion reached by Mansillo and Jackman (2020, p. 145) in their review of the 2019 polls, that ‘the quality of political discussion, the salience and content of the nation’s policy agenda—indeed, the health of Australian democracy—would gain from a commitment to … transparency around the polling of Australian public opinion.’

In relation to industry governance, the conduct of polling, and standards of disclosure, the Inquiry Panel makes a number of recommendations:

**Governance**

**Recommendation 1 – Establish a Code of Conduct for Election Polling:** AMSRO, as the initiators of this Inquiry, help facilitate the establishment of a Code of Conduct for Election Polling in order to provide an oversight, regulatory and disclosure regime for election polling in Australia. This could be achieved by working in tandem with the recently announced Australian Polling Council (YouGov, 2019). This Code should be in place before the next federal election polling cycle.

**Recommendation 2 – Consult the experts:** The development of this Code could be led by pollsters but should be informed by the views of experts from AMSRO, the Statistical Society of Australia, The Research Society, political scientists and the Australian Press Council and/or interested media outlets.

**Recommendation 3 – Ensure compliance:** Fundamental to the integrity and reputation of any such Code are the disclosure requirements it establishes, how these are monitored, and how compliance is ensured. We recommend a similar approach to that of the British Polling Council (BPC) in ensuring that a broad constituency sign up to the Code and that an appropriate arbitration and sanctions process is in place. BPC membership comprises almost every market research organisation in the UK that publishes political polls. Its management committee and officers are drawn from the member organisations, but its Sub-Committee on Disclosure (which provides the technical advice) comprises representatives from research organisations, academia and the media as well as member organisations. The rules of the BPC include a complaints-handling mechanism and enable sanctions to be determined including a period of probation, suspension or expulsion from the BPC (Appendix 8).

**Methodology**

**Recommendation 4 – Develop more effective sample balancing and/or weighting strategies:** Pollsters to identify and better understand the biases in their samples and to develop more effective sample balancing and/or weighting mitigation strategies to improve representativeness. Weighting or balancing by education seems promising and this report suggests several other variables that may warrant further consideration (section 6.4.4 and Appendix 6 – Case Studies).

**Recommendation 5 – Trial new calculation method for the two-party-preferred vote:** Pollsters should not rely solely on the preference flow from the previous election to arrive at a two-party-preferred estimate, not least because some parties may be new, and for very small parties and independents preference flows from the previous election may be difficult to trace. Models that involve
stated preferences with imputation of missing data or the use of stated preferences in combination with preferences from previous election could be experimentally trialled by pollsters.

**Recommendation 6 – Clarify measures of uncertainty:** Pollsters could use more robust methods of estimating the variability associated with their results beyond the currently inadequately calculated and inadequately reported ‘Margin of Error’ heuristic (see section 5.4). In addition, pollsters should routinely report the proportion of respondents who are ‘undecided’ about their vote choice and identify those who are only ‘leaning’ towards a particular party.

**Disclosure**

**Recommendation 7 – Establish disclosure standards:** The 24 disclosure standards set out in Appendix 10 act as a starting point for the development of a coordinated set of publicly available disclosure standards for election polls in Australia.

**Recommendation 8 – Extend beyond election polling:** AMSRO and The Research Society endorse the disclosure standards so that they apply to all Members and all publicly released research, not just election polling.

**Recommendation 9 – Get the media outlets onside:** The support of major news outlets and others who commission or publish polls should be sought to help ensure compliance with any new disclosure standards. Australian media organisations should play an active role in supporting these standards.

**Resources**

**Recommendation 10 – Provide educational resources:** Educational resources about polling methods and standards should be made available to journalists and other interested parties. This could be along similar lines to election polling resources provided by AAPOR in the US, and by the British Polling Council in Britain (AAPOR, 2020; BPC, 2019a). The Research Society, in its submission to this Inquiry, stated that it was willing to make such resources available. We recommend that AMSRO work with The Research Society and the Australian Press Council to make educational resources and training available.
1. Introduction

1.1. Background
The most recent federal election in Australia was held on 18 May 2019. At that time the main opposition party, the Australian Labor Party (ALP), had ‘won’ 55 Newspolls in a row (Edelman.com, 2019).

All the published national election polls conducted throughout the campaign, including the last polls before polling day, estimated, without exception, that Labor had the support of the majority of Australian voters in terms of the two-party-preferred vote. Most commentators declared there would be a change of government.

Below are some of the views expressed based on the final poll results.

<table>
<thead>
<tr>
<th>What the pollsters said …</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Essential</strong>: ‘The final survey of 1,201 voters has Labor in front of the Coalition 51.5% to 48.5% on the two-party-preferred measure, which is the same as last week. The Coalition’s primary vote is 38.5% (up from 38% a week ago) and Labor’s is 36.2% (up from 34%). While Labor has been in front of the Coalition for three years, and in front in every poll during the campaign, strategists believe there will be different results in different parts of the country on Saturday night rather than a uniform national swing either to Labor or against the Coalition – a dynamic that makes the final result difficult to predict.’ (Murphy, 2019a)</td>
</tr>
<tr>
<td><strong>Ipsos</strong>: ‘The national poll of 1,842 respondents (who are all enrolled and certain to vote) shows the Labor Party on 51% (down one point since the mid-campaign poll on 1-4 May 2019), with the Coalition on 49% (up one point since 1-4 May), based on 2016 election preferences. The poll findings suggest that this is a close election with Labor narrowly ahead. While the results show a 1.4% swing against the Coalition Government since the July 2016 election, this assumes a universal national swing, which is unlikely to happen. There are many parliamentary seats where the race between the candidates is very close and the outcome in each of these electorates will determine who gets to form the next government. However, despite the close two-party-preferred figures, 55% think Labor will win, Ipsos Director, Jessica Elgood, said.’ (Ipsos, 2019a)</td>
</tr>
<tr>
<td><strong>Newspoll</strong>: ‘The exclusive Newspoll, conducted for The Weekend Australian, shows Labor heading to the polls with a two-party-preferred vote lead over the Coalition of 51.5% to 48.5%. Bill Shorten is on track to return Labor to power six years after the collapse of the Rudd-Gillard government, with an election-eve Newspoll showing a 1.9% swing against the Coalition, despite voters declaring Scott Morrison to be their preferred prime minister. Despite the half-point break towards Labor in the final week of the campaign, party strategists on both sides believe the swing is patchy and “hand-to-hand” battles in about 20 seats will decide the election.’ (Benson, 2019)</td>
</tr>
<tr>
<td><strong>Roy Morgan Research</strong>: ‘The last face-to-face Roy Morgan Poll before the election shows the ALP regaining the initiative and pulling away from the L-NP with an election winning two-party-preferred lead: ALP 52% cf. L-NP 48%.’ (Roy Morgan Research, 2019a)</td>
</tr>
<tr>
<td><strong>YouGov Galaxy</strong>: ‘Australians may think Scott Morrison has better intentions than his political rival and is a man of action but Bill Shorten remains in the box seat to be this country’s next prime minister, according to a YouGov Galaxy poll conducted for News Corp this week. The poll shows Labor ahead 51-49% on a two-party-preferred basis. “The poll confirms things have tightened since the start of the campaign, but it really is too little, too late,” Mr Briggs (Managing Director of YouGov Galaxy) said.’ (Blickers, 2019)</td>
</tr>
</tbody>
</table>
The commentary accompanying the final Ipsos, Newspoll, Roy Morgan and YouGov Galaxy polls all declared Labor in an election-winning position, albeit in a close contest, whereas the Essential Report had Labor in front but declared the result hard to predict. Overall, the commentary was nuanced and far from categorical.

History tells us the Coalition government comprising the Liberal and National parties (LNP) was returned to office, attracting 51.5% of the two-party-preferred vote compared to 48.5% for Labor. On average, in two-party-preferred terms, the national polls had underestimated support for the Coalition by 2.9 percentage points with a corresponding overestimate of support for Labor.

The 2019 polls reversed the very strong track record of the pollsters in recent elections in ‘calling the right result’ and, as is demonstrated later in this report (Section 5.5), the 2019 polls stand out as being the least accurate since 1998.

The immediate post-election view reflected in the headlines and by most of the commentariat was that the polls had failed spectacularly, adding to the list of polling failures in recent years, prominent examples of which include ‘the 2015 Israeli parliamentary election …, the 2016 Brexit referendum …, and the 2016 US presidential election where the majority of polls predicted that Hillary Clinton would defeat Donald Trump’ (Cornesse, et al., 2020).

Pollsters were left to ponder what went wrong and some announced their own reviews: Essential (Lewis, 2019); Ipsos, (2019b); Newspoll (White, 2019); and YouGov (2019). In the immediate aftermath this Inquiry was also announced by the Association of Market and Social Research Organisations (AMSRO), the membership-based body representing companies in the Australian market and social research industry (AMSRO, 2019a) and a similar call was issued by the Statistical Society of Australia (SSA).

The upshot was one joint inquiry rather than two. This inquiry is undertaken under the auspices of AMSRO with the support of the SSA.

### 1.2. How this Inquiry was conducted

AMSRO announced the Terms of Reference for this Inquiry and the members of the Inquiry Panel on 7 June 2019 (AMSRO, 2019b) (see Appendix 1).

The governance structure comprised the Inquiry Panel itself, an Advisory Board and International Advisers. A discovery process commenced almost immediately with the aim of gathering the views of the main polling organisations. The Advisory Board first met on 23 September 2019 to review initial progress and the Inquiry Panel met shortly thereafter to begin ‘filling in the gaps’ and to commence drafting this report. Subsequent meetings were joint meetings between the Inquiry Panel and the Advisory Board, and members of the Advisory Board contributed not just in an oversight role but also to substantive sections of this report. Given that the Inquiry Panel comprised already busy people contributing their time on a voluntary basis we came to an early realisation that this report would take some time to complete. However, the early work of the Inquiry identified that the disclosure regime required of Australian pollsters was inadequate by international standards and that if this issue were to be addressed in time for the next federal election, one that might be held as early as the second half of 2021, then a Discussion Paper specific to this issue ought to be released. The result was the Discussion Paper: Disclosure standards for election and political polling in Australia released by AMSRO in May 2020 (AMSRO, 2020a). The Discussion Paper invited submissions from interested parties. Ten submissions were received and reviewed. Those submissions, along with previous and subsequent contributions from pollsters, and the Panel’s own research, analysis and expertise, inform this report.
Our findings – regarding the factors most likely to have contributed to the relatively poor performance of the published national election polls in 2019, the adequacy of the reporting on these polls and our subsequent recommendations – are based on the collective decisions of the Inquiry Panel.

1.3. About this report

The task before this Inquiry was to undertake an independent, objective and non-partisan review of the performance of the polls at the 2019 federal election. We have approached this task in the following manner. In section 2 of this report we look at the role of election polls in our society to encourage readers to reflect upon whether election polling actually matters. The implication here is that if election and political polling matters then the accuracy of such polling also matters. From this starting point, it is important in terms of understanding the findings and recommendations of this report that the polling industry, its interrelationship with the media, and the context within which this interrelationship plays out is also understood. This will help the readers of this report and the sponsors and consumers of election polling to form a view as to what are reasonable expectations of the polling companies producing these polls and of the media in reporting these polls. These issues, along with the recent trends in survey research generally and election polling specifically, are explored in section 3. In section 4 we describe how election polling was undertaken in Australia in 2019 and examine these practices against what could be considered a quality hierarchy of polling practices.

The performance of the election polls is explored in section 5. We use several metrics, put the performance of the 2019 election polls into an historical context, look at the performance of the 2019 polls from an international perspective, and present research that examines whether or not there is a partisan bias in the Australian election polls. The main reference period we use for this analysis is 1993 to 2019. The 1993 election was the first election for which all the polls produced an estimate of the two-party-preferred vote and this is a period when polling methodology developed from face-to-face and telephone interviewing via a landline to the range of methods, predominantly online, employed today.

The main diagnostic section of the report is section 6. Here we use the Total Survey Error framework to guide our analysis. In this section we explore, as best we can, those factors that may have contributed to the relatively poor performance of the polls in 2019, including the possibility of ‘herding’. The Inquiry’s Terms of Reference (Appendix 1: Terms of Reference #3) require ‘the causes of any inaccuracies [in the polls] to be investigated’. Potential causes to be considered include: (1) the possible impact of late changes in vote preferences; (2) the extent to which sample frames (sample sources) provided adequate population coverage; (3) sampling methods; (4) interview methods; (5) data weighting or other data adjustments undertaken; (6) differential availability and willingness of respondents to participate; (7) question order and question wording; and (8) data item refusal (aka ‘missing data’ or ‘item-nonresponse’). This section also attempts to address most of the other possible explanations advanced including the ‘shy conservative’ effect, ballot order effects, the treatment of undecideds and don’t knows, the possibility of a late swing and the method used to calculate the two-party-preferred vote.

The focus of section 7 is the transparency and disclosure standards as they apply to publicly released election polls in Australia. This section draws heavily on the Discussion Paper: Disclosure standards for election and political polling in Australia released by the Inquiry Panel in May 2020 (see https://www.amsro.com.au/amsro-polling-inquiry/). While many of the findings and recommendations contained in the Discussion Paper are unchanged others have been added, deleted or modified to take account of reactions to the Discussion Paper and further considerations by the Inquiry Panel. This section, in conjunction with Appendix 8, also examines some of the models that have been used overseas to monitor election and other polling standards, and we reflect on what these tell us about the adequacy of the applicable standards in Australia at the time of the 2019 election.
It is generally recognised that the publication of (pre-)election polls help set expectations among the
general public and commentariat regarding the outcome of an election and ‘the large and extensive
influence of polls on mass and elite politics is well documented in the political science literature’
(Mansillo and Jackman (2020, p. 125). In light of this, we have reviewed a selection of the original
source material and primary media reports that accompanied the release of each of the election polls
to assess whether or not the relevant Australian Press Council (APC) disclosure standards were met
(section 8). In section 9 we tackle the issue of making predictions based on the polls and consider:
What are reasonable expectations to have regarding the precision of the polls?

The findings of this Inquiry and the resultant recommendations are set out in section 10.

We hope that the pollsters that undertake these polls and the media that report these polls appreciate
this review for what it is – a well-intentioned attempt to improve election polling practices in Australia
and to support the accurate reporting of polling results.

1.4. Definition of election and other political polling

An important first step in establishing the remit of this Inquiry was to settle on a working definition of
the types of polling activity covered by the term ‘published opinion polls’ as set out in the Inquiry's
Terms of Reference. We found the typology used by the House of Lords Select Committee into
Political Polling and the Digital Media useful (House of Lords, 2018, p. 13):

- Voting intention poll: This refers to pre-election polls or surveys that aim to gauge how
people intend to vote at any one time or in a particular election.

- Policy issues poll: This refers to polling or surveying undertaken to assess people’s views on
issues that might relate to social policy or politics, such as views on same-sex marriage or
the government’s response to the bushfires crisis or the Coronavirus pandemic, but which do
not involve estimating voting intention.

- Private poll: This refers to the polling or surveying undertaken by political parties, individuals,
or private and public companies, where the results are only selectively released to the public.

- Exit poll: This is a poll conducted of voters as they leave the polling booth.

- Informal poll (sometimes called a ‘snap poll’ or ‘straw poll’): This refers to a poll that has
been conducted without using robust sampling techniques and where the representativeness
of the sample is questionable. An example of this would be a television station or newspaper
running a limited poll of their own readers on an issue. There is nothing inherently wrong with
this approach unless the poll findings are presented as being representative of the wider
population.

- Social survey: This term refers to more comprehensive, longer-running exercises conducted
by governments, independent research agencies, academics and think tanks to measure
social and policy issues.

The focus of this Inquiry is on national voting intention polls when the results of these polls are
released to the public. In this report, such polls are referred to as ‘election polls’ or ‘pre-election polls’.
A strong argument also exists that when privately commissioned polls are made public by the
organisations or researchers involved, they should be required to meet the same disclosure standards as set out in this report.¹

1.5. Caveats and limitations

This Inquiry did not canvass the views of all pollsters nor all polling commentators or experts. Given the focus on publicly released national election polls, the Inquiry Panel sought the cooperation of the four pollsters responsible for these polls; Essential Research, Ipsos, Roy Morgan Research and YouGov (responsible for the YouGov Galaxy and Newspoll polls). Lonergan Research was also approached in the hope that it would provide some insights into polling via the use of Interactive Voice Recognition (IVR) technology (i.e. robopolls) and because it went public with a claim of being involved with the suppression of a poll that was out of alignment with the prevailing view at the time, a form of ‘herding’, (see section 6.5). Beyond this, informal discussions were held with some pollsters from outside this group and all pollsters and other interested parties were able to make submissions to the Inquiry via the public submissions process (AMSRO, 2019c). For the most part, pollsters were reluctant to fully cooperate with the Inquiry; some refused. This reluctance seemed mainly to stem from a concern that they would be asked to disclose what they regard as proprietary or commercial-in-confidence information or because they were embarking upon their own reviews.

Essential, Ipsos, Lonergan Research and YouGov participated in interviews with the Chair and one other member of the Inquiry Panel (some of these being off-the-record) and Essential and YouGov made written submissions. Only Ipsos and Lonergan provided copies of their questionnaires and only Ipsos provided cross-tabulations of their polling results. None were prepared to provide data sets to the Inquiry and Roy Morgan Research chose not to participate at all. In general, the level of information provided by pollsters to this Inquiry was much less than was provided to similar inquiries conducted into the performance of recent election polls in the US (AAPOR, 2017) and Britain (Sturgis et al., 2016). While we are grateful for the cooperation received, the lack of access to data sets and to detailed descriptions of the survey methods and statistical techniques that were used materially limits the ability of this review to definitively identify the specific factors that contributed to the inaccuracy of the published national election polls in 2019.

¹ The view of the British Polling Council is that their disclosure standards apply to the privately commissioned polls of member organisations ‘in the event that the results of (such polls) are made public by the organisation that commissioned the survey (including its employees or agents)’ (British Polling Council, n.d.).
2. **The importance of election polling**

Implicit in the commissioning of this Inquiry is the notion that having accurate election polls matters. This certainly seems to be the case, based on the widespread criticism the polls received following the election.

The Inquiry Panel believes that well-funded, well-designed, well-executed and carefully interpreted election and political polls are capable of providing a *guide* as to:

- Which of the political parties is ahead and which is behind, whether the gap between the government and opposition is widening or narrowing, which of the two is likely to win the majority of seats or to form government, support for leaders (and alternative leaders) and the public’s position on issues.
- The strength of minor parties – those that are not the parties of government – and of independents.
- The likely flow of preferences from minor parties and independents to the major parties.

The polls are often influential in many ways:

- Among politicians, across the media, and among voters, political polls can create expectations, especially about which party or parties are likely to win, and expectations can drive a range of actions – changes in party leadership, resignations in advance of an election, closer policy scrutiny of a party by the media, and activism among the voting public.
- Businesses, the public service, and non-government organisations may also be affected by the polls, their actions or contingency plans being adjusted or maintained depending on what the polls appear to be telling them – more so when all the polls, or the most reputable of them, seem to be pointing in the same direction, and remain unchanged over long periods.
- Politicians and political candidates pay close attention to the publicly released polls, especially from the most prestigious pollsters. Typically politicians outside the leadership group of their party only have access to the public polls. A party’s faith in the findings of their private research may be strengthened if the public polls seem to support their findings, or raise doubts if there is a divergence in findings.
- For political journalists, broadcasters, bloggers and others operating in the public sphere, political polls have become central to how they discuss electoral politics. The polls help the media organise their coverage, prioritise stories, and prepare for political outcomes of various kinds. Polls help attract audiences; hence, the media’s willingness to commission or simply publicise polls. During election campaigns, poll results are front-page stories (see section 8.1). They may also be used to help drive online subscriptions. The premier election polls also help to drive stories in other media.

Voters may use election polls – national polls but also polls in particular seats – to get a sense of: whether they are out of step with the wider public, raising the possibility of a poll-induced ‘bandwagon’ effect; whether the outcome seems a forgone conclusion, raising the possibility of a poll-induced ‘underdog’ effect, with voters not wanting the party that is ahead to win by too big a margin or not wanting it to win in both Houses of Parliament; and, in those electorates for which individual seat polls are published, whether they should vote tactically or sincerely.

For these reasons, there is an obligation upon pollsters and those reporting the findings of polls ‘to disclose sufficient information about how their research was conducted to allow for independent review and verification of research claims’ (AAPOR, 2015). This is an obligation that Australian pollsters and those reporting the polls do not meet routinely.
3. The election polling environment in Australia

3.1. Context

During the 2019 election campaign, 16 national polls were published, each with its own estimate of how Australians would vote. Five were commissioned by The Australian, the national daily newspaper owned by News Corp Australia; two by The Australian Financial Review, The Sydney Morning Herald and The Age – formerly owned by Fairfax, now owned by Nine Entertainment; and two by News Corp metropolitan mastheads – The Daily Telegraph (NSW), Herald-Sun (Vic), The Courier-Mail (Qld), The Advertiser (SA), and The Mercury (Tas). The other seven were conducted by market research firms with no media contracts: four by Roy Morgan Research and three by Essential Research, which released its results to The Guardian.

Election polling accounts for only a small part of Australia’s approximately $1 billion market and social research industry (ESOMAR, 2017) yet the interviews with the pollsters who cooperated with this Inquiry revealed that they felt that election pollsters had been made ‘scapegoats for the industry’. There were many factors, they insisted, that had made polling for the 2019 election both more difficult and more expensive than in previous years; in particular, issues associated with declining contact rates, cooperation rates, and response rates. The general view expressed by the pollsters was that the level of funding for election polls was barely adequate, and certainly not commensurate with the levels required if pollsters were to adopt more rigorous methods. In terms of its financial returns, election polling represents, at best, a marginal proposition. Under these circumstances, pollsters argued that if the impact of the recommendations from this Inquiry were to drive up the costs of polling, it was unclear how these additional costs would be met. One possible consequence could be that even fewer organisations would be prepared to undertake publicly released election polls.

While the pollsters’ primary motivation may not be a direct financial return, it is important to point out that polling is one means by which market research firms advertise their existence and seek to promote their skills. Even in the absence of media sponsorship, opinion polls would still be conducted ahead of an election. The fact that some pollsters conducted election polls of their own volition speaks to this point.

The Inquiry Panel is of the view that the polling companies, having made the decision to publicly release the results of their polls, have a commensurate responsibility to conduct their polls in ways that are consistent with defensible methodological standards and to do so with a level of transparency that allows them to withstand proper scrutiny. The level of disclosure to this Inquiry suggests there is still some way to go in this regard.

3.2. Recent changes in survey research

Survey research methods, including those used for election and political polling, have undergone considerable change in recent years, arising largely from changes in the communications and technology landscape. The 2018-19 Communication Report produced by the Australian Communications and Media Authority (ACMA, 2020) reveals the following. As of June 2019:

- 91% of Australian adults had a home internet connection during the six months prior to May 2019 (ACMA, 2020, p. 1).
- Only 49% of Australian adults had a fixed in-home telephone line, down from 70% four years earlier (ACMA, 2020, p.4), and 51% were mobile-only for voice (ACMA, 2020, p. 71), up from 7% in 2008 (Social Research Centre, 2019).
- 96% of Australian adults had a mobile phone and 83% had a smartphone (ACMA, 2020, p. 5).
Survey research methods have evolved in response to these changes in the technology and communications landscape and in response to changes in the dynamics of conducting surveys of the general community.

Until recently, and for many years, Computer Assisted Telephone Interviewing (CATI) had been the go-to method for election and political polling in Australia. After a long period of relative stability for telephone surveying in Australia, a notable methodological shift started in 2010 – somewhat earlier in USA, the UK and Europe – with the introduction of dual-frame random digit dialling (DFRDD) (Pennay, 2010, Lavrakas et al., 2008, 2017). DFRRDD, which added mobile phone numbers to landlines, was first used for national election polling in Australia by Ipsos in 2016.

Prior to the introduction of DFRRDD most telephone surveys relied on sampling households only via landline telephone numbers that were either randomly generated (landline RDD) or listed in directories. The development of DFRRDD approaches meant that the mobile-phone-only population (persons who have a mobile phone and live in a household without a landline telephone connection) could be included in telephone survey estimates for the first time. When DFRRDD surveys first began in Australia in 2010, 13% of the population had access only to a mobile phone (Pennay, 2010).2

In addition to the decline in the proportion of households with a landline telephone connection (accelerated in recent years by the rollout of the National Broadband Network) and the huge increases in mobile phone and smartphone penetration, other changes that added to the cost and degree of difficulty of conducting telephone surveys include the increasing use of answering machines (on landlines) and Voicemail (on mobile phones) to screen incoming calls, the introduction of the Do Not Call register (in 2003)3, and the more recent introduction of call-blocking functionality and SPAM filters on mobile phones. These changes, and others such as the use robopolling (see section 4.1), the increased volume of telemarketing calls and the use of predictive dialling technology (which means the telephone answerer is often greeted by a silent ‘delay’ when answering the phone), have led to a situation where contact rates, cooperation rates and response rates for telephone surveys have declined dramatically.4 As a result, the cost of undertaking high-quality telephone surveys has increased exponentially (Social Research Centre, 2019).

Data compiled by the Social Research Centre (SRC) and presented at the Telephone Surveying in the Postmodern Era workshop (with 2019 data added for this report) plots the decline in contact rates, cooperation rates and response rates over the last decade for a major Australian national telephone survey.5 The cumulative effect on the overarching measure of survey participation, the survey

---

2 Another 2% of the adult population are thought not to be contactable by phone at all (Phillips et al., 2019)
3 By asking for their number to be included on The Do Not Call Register, members of the community are opting out from receiving unsolicited telemarketing calls and faxes. People sometimes think that being on the Do Not Call Register exempts them from receiving legitimate market and social research calls, a belief that, even though this is not the case, can be a factor in driving up refusal rates to calls made for legitimate research purposes.
4 Crudely defined, the ‘contact rate’ can be thought of as the percentage of working telephone numbers that are answered by a person. The ‘cooperation rate’ can be thought of as the percentage of in-scope persons who answer the incoming call to their phone and then go on to complete an interview. The response rate is akin to an amalgam of the contact rate x the cooperation rate and gives a sense of the proportion of the in-scope sample actually interviewed. The rates reported here are AAPOR CON2, AAPOR COOP2 and AAPOR RR3 (AAPOR, 2016).
5 Contact rates from landline telephone numbers fell from 85% in 2009 to 47% in 2019, while contact rates from mobile phone numbers fell from 50% in 2013 to 34% in 2019. Cooperation rates were more stable but dipped sharply in the last couple of years. Cooperation rates from calls to landline numbers sat at between 40% and 50% over the period 2009 to 2015 but had dropped to 24% by 2019. Cooperation rates for calls to mobile phones were slightly higher at 30% in 2019 (Social Research Centre, 2019).
response rate, is shown in Figure 1. This shows a decline in survey response rates from 35% in 2009 to 11% in 2019; almost half the rate observed at the time of the previous election in 2016 (20%).

Figure 1: ‘Typical’ response rates for telephone surveys in Australia, 2009 – 2019.

The rapid changes in technology and communications, as well as the erosion in conditions for telephone surveying, have contributed to the uptake of lower-cost (and quicker) survey modes, particularly the use of non-probability online panels and telephone surveys where the in-person telephone interviewer has been replaced by a scripted voice recording or a computer-generated voice (i.e. robopolls). Data compiled from the ESOMAR Global Market Research reports show that since 2010 opt-in sampling and online research has been the dominant mode of data collection in the Australian market and social research industry, supplanting CATI. In Australia in 2016 (the most recent data available to the Inquiry), online research accounted for 44% of the revenue generated by the industry, up from 29% in 2011. Over the same period the percentage of research industry revenue attributable to telephone surveys declined from 22% in 2011 to 6% in 2016 (Social Research Centre, 2019).
4. How election polls are conducted

4.1. A variety of methods

The evolution of polling methodology in recent years reflects the trends outlined above. Regardless of the method used, the objective is the same: ‘to efficiently [and accurately] estimate the characteristics of a large population based on measurements of a small subset of the population’ (Cornesse et al., 2020, p. 7).

When researchers are considering which survey methods to use it is not just the mode of data collection (e.g. interviewer-administered or self-administered) that is important but also the representativeness of the unweighted sample, based on the extent to which the sample source being used covers the population of interest and whether there is a non-response bias. Generally, the decision as to how to best proceed should strive to balance these two factors against other considerations such as timelines, budget and the degree of confidence needed in the resultant estimates.

For example, the reason for choosing a telephone-based mode of data collection would probably be driven mainly by the desire to use the high-coverage telephone sampling frames available in Australia (when using dual-frame RDD designs) rather than by a particular desire to use telephone as the mode of data collection. Having decided to use a telephone sampling frame, the subsequent decision to use a live interviewer in preference to an automated robopoll would likely be driven mainly by the ability for a live interviewer to generate a higher and more representative response rate, thereby reducing the likelihood of non-response bias. On the other hand, the reason for choosing an online self-completion mode is likely to be driven by the relatively low cost of using such samples rather than by the extent of their coverage of the population of interest.

Computer Assisted Telephone Interviewing (CATI)

CATI involves an interviewer calling either a landline or mobile phone telephone number, gaining cooperation from an eligible respondent, asking the scripted questions displayed on a computer screen, and entering the respondents’ answers directly into the computerised data collection program. Nowadays, many survey research firms engaged in CATI use automated systems for dialling the telephone numbers and this often results in a silent ‘delay’ when the call is answered. Over the main period of focus for this Inquiry, 1993-2019, CATI was the dominant mode of data collection for election polling – though this was not the case in 2016 or 2019.

Face-to-face interviewing

The face-to-face (or personal visit) mode of data collection is where interviewers doorknock a sample of addresses (usually based on a geographically clustered stratified sample design) to gain cooperation from an eligible respondent, and then administer a questionnaire. Over the period 1993-2019, only one pollster, Roy Morgan, has used this method.

Online surveying via non-probability panels

The main method used by Australian pollsters for publicly released national election polls in the two most recent elections, 2016 and 2019, was online surveying via non-probability panels. This is now the main method worldwide (Callegaro, Villar, Yeager, and Krosnick, 2014a; Callegaro, Baker, Bethlehem, Göritz, and Krosnick, 2014b). While attractive in terms of cost and a relatively quick turnaround, non-probability online panels face several challenges, including the non-coverage of the population...
population without internet access, and selection bias due to the reliance on convenience samples of volunteers (Bethlehem, 2017), often not representative of the general population.

Non-probability online panels were first used for national election polling in Australia in 2001 (ANU/YouGov) and have been a regular feature since 2010 (see Appendix 2). In 2019, three of the five publicly released national election polls used non-probability panels alone or in combination with another sampling frame. In their review of trends in polling accuracy and their exploration of the causes of inaccuracy, Prosser and Mellon (2018, p. 760) noted that a key problem observed by researchers using non-probability panels was that respondents on such panels, ‘are generally more politically engaged than the general population’ and generally more politically engaged than those reached through probability samples (Fahimi, 2015, p. 5).

When considering non-probability sampling methods used for Australian elections, the most informative comparison is between non-probability methods and probability-based methods. Probability sampling draws its strength from a set of established mathematical principles that makes it possible to compute the accuracy of estimates (e.g. in the form of confidence intervals or margins of error). For non-probability sample surveys, there is no general statistical underpinning. Instead, the justification for non-probability sampling relies on ‘four basic types of claims: (i) that any sample examining a particular question will yield the same inferences; (ii) that the specific design of the sample, as related to the questions at hand, will produce conclusions that mirror the population of interest; (iii) that a series of analytical steps will account for any differences between the sample and the population; and (iv) that the particular combination of sample and/or analytic approaches will produce accurate population estimates’ (Cornesse, et al., 2020, p. 8). Ideally, those using non-probability samples should disclose their assumptions and their methods in order to demonstrate why others should have confidence in the accuracy and reliability of their estimates (as should those using probability sampling methods who have a stronger theoretical claim to accuracy).

Interactive Voice Recognition (Robopolls)

The technology behind robopolling – known as Interactive Voice Recognition (IVR) – has been used in survey research since the early 1980s (Tourangeau et al., 2002) to recruit and/or gather data from people sampled via telephone numbers. The IVR technology can be used with either probability or non-probability samples. Appendix 2 shows that robopolls were first used for national election polling in Australia in 2013. The pioneer of this method for polling in Australia was JWS Research, who used IVR to survey voters via their landline for single-seat polls in 2010. In 2013 ReachTEL and Lonergan Research both used IVR to landlines. Also, in 2013, four poll organisations – Galaxy, JWS, ReachTEL and Lonergan – used IVR to poll in marginal seats (Goot, 2014, p. 26; 2015, pp. 133-36). In 2016, ReachTEL conducted IVR to landlines and Lonergan conducted IVR to both landlines and mobile phones, and Newspoll and YouGov Galaxy used IVR to landlines in conjunction with a non-probability online panel. Newspoll persisted with this hybrid IVR/non-probability panel method in 2019.7. While some IVR polling has been used in national election polls in Australia since 2013, its main use has been for polling in marginal seats (Goot, 2014, Goot 2018, Goot, 2020a, Goot, 2020b).

In traditional CATI polling, interviewers speak with respondents over the telephone, generally asking to interview either a randomly selected adult from within the household (based on next or last birthday) or with the youngest male (or female) adult in the household as young people are generally underrepresented in telephone surveys.8 With robopolls, what respondents hear when they answer the

---

7 Following the 2019 election, YouGov announced that they are abandoning this method for the Newspoll in favour of a ‘100 per cent online approach’ (White, 2019).

8 A respondent selection routine is usually used when calling landline telephone numbers but the individual who answers the phone, provided they are in scope, is generally the person selected for interview when calling mobile phone numbers.
phone is a recorded voice (or nowadays, a computer-generated voice) inviting the person who has picked up the phone to participate in an interview provided they are aged 18 years or over. The respondent answers the recorded questions by pressing numbers on their telephone keypad.

Despite its widespread use for election polling in Australia and the United States there has been a conspicuous lack of attention paid to IVR in the methodological research. This relative dearth of attention may be related to a belief among survey researchers that high-quality surveys cannot be conducted via IVR (Traugott and Lavrakas, 2016, p. 73).

Other issues commonly associated with IVR include very low response rates (when compared with interviewer-administered modes of survey administration), a high-level of item non-response (i.e. missing data) and substantial ‘primacy’ effects; that is, a heightened tendency for respondents to select response options from earlier in the response list (Lavrakas and Richards, 2020) – issues that most who use IVR do not try to control/reduce. Whether this was true of the (limited) IVR polling conducted for the 2019 election, the dearth of evidence provided to the Inquiry means that we cannot say. There may also be issues relating to data quality arising from mis-keying an answer and respondents not being able to go back to correct the error. And there is the issue of whether the questionnaire is programmed in such a way as to enable ‘barging in’, that is, whether a respondent is allowed to key in a response to a question before all options have been read out. The argument in favour of allowing barging in is that it enables respondents to answer a question as soon as they have heard the option they want to choose, thereby preventing the interview from becoming too tiresome (which can result in an increase in mid-survey terminations/breakoffs). It may also be the case that if respondents become ‘annoyed’ with the robopoll ‘environment,’ incorrect data are more likely to be entered, and satisficing in the form of non-differentiated answers and ‘speeding’ is more likely to occur. Of course, enabling a respondent to answer a question before all response options have been read out also risks a less accurate answer.

On the other hand, to the extent that participating in a robopoll affords as much privacy as participating in an online survey, robopoll polls should not differ from online polls in terms of social desirability or acquiescence bias.

Campbell White, Head of Public Affairs and Polling (Asia-Pacific) at YouGov, has argued that ‘response rates for robocalls in Australia have been falling and many people find them annoying and invasive. They tend to be answered largely by older people or those who are very interested in politics. Busy people who are less interested in politics either don’t answer or hang up. We believe this was a significant contributor to the inaccuracies seen at the (2019) election’ (White, 2019). Chris Lonergan described IVR to the Inquiry ‘as a business and reputational risk’, noting that ‘commercially there is not much upside. If industry banned robopolling, the industry will [sic] come up with something better’. Ipsos and Essential told the Inquiry that robopolling was damaging the industry.

**Mixed modes of Sampling, Recruitment and/or Data Collection**

The use of different survey modes applies not just to the mode in which data is collected from respondents but also to the sampling and recruitment of respondents. To achieve the optimal design for a given survey/poll, researchers and pollsters are increasingly turning to mixed-mode designs for sampling, recruitment and/or data collection.

---

9 The most common strategy used to control for, or at least minimise, the sometimes-strong primacy effects associated with robopolling is to reverse the order of the response alternatives for two split-half/randomised subsets of respondents.

10 Social desirability bias is the tendency for some respondents to report an answer in a way they deem to be more socially acceptable than their ‘true’ answer. Acquiescence bias is the tendency for respondents to agree with statements — (Lavrakas, 2008, pp. 3, 825-26).
For their 2016 polls, both Newspoll and Galaxy adopted hybrid methods, combining different modes of sampling (telephone for the landline sample and online for the non-probability panel) with different modes of data collection (IVR for the telephone sample and web for the online panel). YouGov continued with this method for Newspoll in 2019. In their written submission to the Inquiry, YouGov explained that ‘Deficiencies in early online panels meant that when we were awarded the contract to conduct Newspoll on behalf of *The Australian*, we decided that Galaxy should ‘include automated or robopolls as part of the mix’.

However, when combining modes in the hope of reducing one source of error (in this case, the coverage and self-selection biases associated with the use of a non-probability online panel), other factors need to be considered. At issue here is whether survey questions read by respondents and answered online via desktop/laptop/tablet or mobile phone are understood and answered in the same way as questions administered via robopolls (i.e. asked over the telephone by a recorded voice or computer-generated voice and answered by keying in a response via a telephone keypad).

Another factor that warrants careful consideration when mixing modes is the proportion of completed questionnaires obtained via each mode of sampling, recruitment and data collection. In determining the share of sample across modes, account needs to be taken (to the extent possible) of the sampling and non-sampling errors from the modes and the associated weighting schemes. If this is not carried out, the objective of the mixed-mode design (to lower total error) may well fail. Of the five national polls conducted by Newspoll over the course of the 2019 election campaign, the proportion of completed questionnaires obtained from robopolling ranged from 43% to 55%. In email correspondence to the Inquiry, YouGov was not able to shed any light on this historical design issue but did note that this anomaly, along with others, was one of the reasons for abandoning this approach.

Mixed-mode surveys that do not adopt a fixed mode share risk introducing a design-based variation to their survey estimates.

Section 5.7 sets out the available Australian and international evidence regarding the relative accuracy of the various methods used for election polling.

### 4.2. A hierarchy of survey methods

Survey methodologists generally accept that when accuracy is the only criterion, there is a survey quality hierarchy. At the lowest-quality end of the hierarchy are surveys based on non-probability samples that are not weighted or only have post-stratification weights applied to align the sample profile with standard demographic characteristics such as age, gender and location (the same variables for which quotas are usually set). This approach will usually suffer from non-ignorable sample bias, even after weighting. It is also difficult to calculate meaningful confidence estimates or margins of sampling error from such surveys. Estimates from non-probability samples generated after applying more sophisticated sample balancing and/or weighting techniques based on characteristics associated with the outcome variable/s of interest are generally considered more likely to yield accurate findings.

On the next rung of the survey quality hierarchy are survey estimates from non-probability samples when combined and/or calibrated with a good quality probability sample and/or high-quality benchmarks.

On the top rung are probability samples that use advanced weighting methods and are drawn from sample frames with ignorable coverage error.

This hierarchy should not be interpreted as implying that all surveys based on non-probability samples are inherently bad and all surveys based on probability samples are inherently good. While the literature over the last 15 years or so shows that surveys using probabilistic sampling techniques
generally produce less biased estimates than surveys using non-probability samples (Cornesse et al., 2020, pp. 15-17), this is not always the case.

4.3. Polling practices at the 2019 election

Table 1 uses available information to summarise the methods used to undertake the publicly released national election polls in 2019. Essential and YouGov Galaxy used non-probability online panels. Newspoll combined estimates from a non-probability online panel with robocalls to landline telephone numbers. Ipsos and Roy Morgan used probability-based methods; DFRDD for Ipsos and face-to-face interviewing for Roy Morgan.
Table 1: Polling methods used for public release national election polls in 2019

<table>
<thead>
<tr>
<th>Description of method</th>
<th>Essential</th>
<th>Ipsos</th>
<th>Newspoll</th>
<th>Roy Morgan</th>
<th>YouGov Galaxy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population/Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample frame</td>
<td>Non-probability online panel</td>
<td>Probability-based DFRDD design. Landline and mobile phone telephone numbers</td>
<td>Non-probability online panel and landline IVR</td>
<td>Geographically stratified sample of residential addresses</td>
<td>Non-probability online panel</td>
</tr>
<tr>
<td>Non-coverage</td>
<td>Excludes the off-line population</td>
<td>Excludes the &lt;2% of the population not contactable either by a landline or mobile phone</td>
<td>Excludes the off-line population</td>
<td>Excludes people living in non-residential dwellings and people of no fixed address</td>
<td>Excludes the off-line population</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td>English</td>
<td>English</td>
<td>English</td>
<td>English</td>
</tr>
<tr>
<td>Achieved sample size (for final poll)</td>
<td>1,201</td>
<td>1,842</td>
<td>3,038</td>
<td>1,265</td>
<td>1,004</td>
</tr>
<tr>
<td>Sample design: including quotas and stratification scheme and other elements of sample design as applicable/appropriate (e.g. cluster or multi-stage design)</td>
<td>Quotas used</td>
<td>ABS-based quotas on capital cities, and rest of State, for each State/Territory, and also age and gender</td>
<td>Quotas used</td>
<td>Unknown</td>
<td>Quotas used</td>
</tr>
<tr>
<td><strong>Fieldwork</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019 fieldwork dates (for final poll)</td>
<td>10-14 May</td>
<td>12-15 May</td>
<td>14-17 May</td>
<td>?? -12 May</td>
<td>12-15 May</td>
</tr>
<tr>
<td>Use of respondent incentives or other means for gaining cooperation</td>
<td>Yes</td>
<td>No</td>
<td>Yes, for online sample</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Description of method</td>
<td>Essential</td>
<td>Ipsos</td>
<td>Newspoll</td>
<td>Roy Morgan</td>
<td>YouGov Galaxy</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------</td>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode of data collection (e.g. online/CATI/IVR, Mixed mode, other). See item 20 for expanded requirements</td>
<td>Self-administered Online</td>
<td>Interviewer-administered CATI</td>
<td>Self-administered Online +IVR</td>
<td>Interviewer-administered Face-to-face</td>
<td>Self-administered Online</td>
</tr>
<tr>
<td><strong>Survey questions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The exact wording of relevant questions and response options and the order in which they were asked</td>
<td>Some questions available from public reports</td>
<td>Full questionnaire provided to Inquiry</td>
<td>Some questions available from public reports</td>
<td>Some questions available from public reports</td>
<td>Some questions available from public reports</td>
</tr>
<tr>
<td><strong>Response rates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provision of response rates or similar statistics including description of how these metrics have been calculated</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td><strong>Weighting</strong>&lt;sup&gt;(a)&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation methods to be described, including whether the survey has been weighted and if the survey has been weighted, what weighting techniques were used (e.g. post-stratification, raking, etc.); the weighting variables used and benchmarks used for the weighting parameters</td>
<td>The data are weighted against ABS parameters for age, gender and location</td>
<td>Weighting by age, sex, location and income Results filtered on enrolled to vote and certain to vote</td>
<td>Weighting not described</td>
<td>Weighting not described</td>
<td>(1) age interlocked with gender and region (2) education (3) previous election voting intention &lt;sup&gt;(b)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Description of method</td>
<td>Essential</td>
<td>Ipsos</td>
<td>Newspoll</td>
<td>Roy Morgan</td>
<td>YouGov Galaxy</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Analysis and reporting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measures of uncertainty such as margin of sampling error (MoSE) and how estimated</td>
<td>MoSE</td>
<td>MoSE</td>
<td>MoSE</td>
<td>MoSE</td>
<td>MoSE</td>
</tr>
</tbody>
</table>
| How the two-party-preferred vote was calculated (e.g. by allocating preferences as per previous election(s), or asking intending minor party voters their major party preference or other means) | The two-party-preferred estimate was calculated by distributing the votes of the other parties according to their preferences at the 2016 election | Two methods were used and reported: 1) using 2016 preferences flows; and 2) using stated respondent preferences | Recent federal and state elections | Stated preferences | Preference flows are calculated using a formula derived by YouGov from recent elections a)
| Quantifying item-level non-response, including the definition and treatment of unknowns, unstated and undecided | Don't knows/Undecided were not reported | Don't knows/Undecided were reported | Don't knows/Undecided were reported | Don't knows/Undecided were reported | Don't knows/Undecided were reported |
| The proportion of telephone interviews undertaken on a mobile phone and/or the proportion of online surveys completed on a mobile or small screen browser | Not provided | 50% landline / 50% mobile | Spilt varied from 43-55% IVR. Proportion of online panel questionnaires completed on mobile/small screen devices is unknown | Not relevant | Not provided |
| Membership and accreditation of key professionals with the relevant professional bodies (e.g. QPR or AStat) | Individual pollsters are members of The Research Society (RS) | Polling team comprises TRS (RS) members, QPR, member of the Statistical Society of Australia | Not provided | Not provided | Not provided |

Source: Compiled from publicly available information and information provided by the pollsters. Essential and Ipsos confirmed/contributed to the descriptions used in this table. The other pollsters were given the same opportunity but did not confirm the contents of this table.

a) The specific age group and geographic dimension used for weighting are not available to the Inquiry.
5. The performance of the polls

In this section we assess the performance of the polls, first against the most commonly used measure of whether they picked the winner. Where polls have picked the winner, criticism of their performance has been quite limited – even if they have substantially underestimated or overestimated the vote shares, especially of the LNP Coalition (‘the Coalition’) and Labor. We also use a range of statistical metrics to assess the performance of the polls relative to the historical performance of election polls in Australia and internationally. A summary of poll performance by methodology is also undertaken as well as an analysis of any evidence of a partisan skew in the polls.

5.1. How often do the election polls ‘get it right’?

How often have the pollsters ‘got it right’ in national elections since they all began reporting a two-party-preferred vote in 1993? The breakout box details the final poll results against the election outcome, and Appendix 4 summarises the percentage of the final polls over this period that ‘called’ the election result correctly. Calling the right result involves, most typically: (a) having the winning party or parties ahead on the two-party-preferred vote when the winning party or parties did indeed secure the majority of the two-party-preferred vote; or (b) much less often, calling it for the party or parties that won the election, even when they had the party or parties that won level-pegging or even behind on the two-party-preferred vote and the party or parties did win with 50% or less of the two-party-preferred vote.

How often the pollsters ‘got it right’, 1993-2019

The 1993 election was won by Labor with a two-party-preferred vote of 51.4% compared to 48.6% for the Coalition. The Morgan Poll (50.5%) and Saulwick (51%) had Labor ahead; AGB McNair and Newspoll had the Coalition ahead, 50.5%. Conclusion: Two of four called the result.

The 1996 election was won by the Coalition with 53.6% of the two-party-preferred vote. ACNielsen, Newspoll and Quadrant had the Coalition winning between 51% (Nielsen and Quadrant) and 53.5% (Newspoll). The Morgan Poll had the Coalition on just 45%. Conclusion: Three of four called the result.

The 1998 election was won by the Coalition with 49% of the two-party-preferred vote. Even though Labor won the popular vote with a two-party-preferred vote of 51.0%, it won only 67 out of the 148 seats in the House of Representatives compared with 80 seats for the Coalition, with one independent. Neither the Morgan Poll, which underestimated the Coalition by one percentage point, nor Newspoll, which underestimated the Coalition vote by two percentage points, tipped the Coalition to win. Conclusion: Four of six called the result.

The Coalition retained government in 2001 with a two-party-preferred vote of 50.9% compared to 49.1% for Labor. Three of the four pollsters had the Coalition ahead, the Morgan Poll underestimating the Coalition vote by 5.4 percentage points. Conclusion: Three of four called the result.

The 2004 election saw the Coalition returned with 52.7% of the two-party-preferred vote compared with 47.3% for Labor. Five pollsters (ACNielsen, the Australian National University, Galaxy, Newspoll and Roy Morgan) conducted six final election polls – Roy Morgan producing results from separate surveys, one using face-to-face interviewing and the other by telephone. Of the six polls, three had the Coalition with a popular vote of 50% or more (ACNielsen, Galaxy and Newspoll) while both ANU and Roy Morgan polls had the Coalition short of a majority. Conclusion: Three of six called the result.

Polls that estimated the two-party-preferred vote at 50:50, or numbers close to that, which the pollster said made it too close to call, are not counted as wrong calls where the actual result was from 49-51 to the LNP and 51-49 to Labor.
The 2007 election was won by Labor with 52.7% of the two-party-preferred vote compared with 47.3% for the Coalition. The final polls of all four pollsters (Galaxy, Newspoll, ACNielsen and Roy Morgan) called the correct result. Roy Morgan conducted telephone and face-to-face polls, ACNielsen conducted separate telephone and online polls. Conclusion: Six of six called the result.

The 2010 election saw Labor secure a two-party-preferred vote of 50.1% compared with 49.9% for the Coalition. Six pollsters (Essential, Galaxy, Newspoll, ACNielsen, Roy Morgan and JWS) conducted seven polls with Roy Morgan again conducting a face-to-face poll and a telephone poll. Each of the polls estimated Labor to have more than 50% of the vote. The election actually resulted in a hung parliament with 72 seats each to Labor and Coalition. Labor was able to form a minority government with the support of four of the six Independents. Conclusion: Seven out of seven pollsters came down on the right side of the result.

The 2013 election saw a Coalition victory with 53.5% of the two-party-preferred vote compared with 46.5% for Labor. In the final polls (AMR, the Essential Report, Galaxy, Lonergan, Newspoll, ACNielsen, ReachTEL and the Morgan Poll) estimates of the Coalition vote ranged from 50.8% to 54%. Conclusion: Eight of eight called the result.

The Coalition retained government at the 2016 election with 50.4% of the two-party-preferred vote compared with 49.6% for Labor. In a very tight election, four out of five pollsters (Essential, Galaxy Newspoll and ReachTEL) called it correctly with Ipsos estimating that the Coalition would get 49% of the two-party vote. Conclusion: Four of five called the result.

The 2019 election resulted in the return of the Coalition government with 51.5% of the two-party-preferred vote compared to 48.5% for Labor. The final polls of all five pollsters (Essential, Ipsos, Newspoll, Roy Morgan and YouGov Galaxy) produced the mirror opposite of this result. Conclusion: Five out of five called it wrongly.

Source: The election results were sourced from https://www.aec.gov.au/Elections/Federal_Elections/tpp-results.htm viewed on 24 May 2020. The results for the 55 polls undertaken over this period are from Goot, in press).

To summarise (Table 2), over the period 1993 to 2019, there have been 40 ‘right calls’ from the 55 final election polls being considered – a success rate of 73%. As a point of comparison, the record of US pollsters in getting it right over a similar period, 1998 to 2018, was 79% (Silver, 2018a). Over the period 1993 to 2019, 2019 is the only election where all the pollsters involved in national election polling called the wrong result.

In the four elections preceding the 2019 election – 2007, 2010, 2013, and 2016 – there was a 96% success rate (25 correct calls out of 26 polls). This period of sustained success may have led to pollsters being less alert to the rapidly changing conditions for polling and survey research outlined in section 3.2.
Table 2: Number of final polls ‘getting it right’, 1993-2019

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of polls</th>
<th>No. Called Correctly</th>
<th>% Called Correctly</th>
<th>Winner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>4</td>
<td>2</td>
<td>50</td>
<td>ALP</td>
</tr>
<tr>
<td>1996</td>
<td>4</td>
<td>3</td>
<td>75</td>
<td>LNP</td>
</tr>
<tr>
<td>1998</td>
<td>6</td>
<td>4</td>
<td>67</td>
<td>LNP</td>
</tr>
<tr>
<td>2001</td>
<td>4</td>
<td>3</td>
<td>75</td>
<td>LNP</td>
</tr>
<tr>
<td>2004</td>
<td>6</td>
<td>3</td>
<td>50</td>
<td>LNP</td>
</tr>
<tr>
<td>2007</td>
<td>6</td>
<td>6</td>
<td>100</td>
<td>ALP</td>
</tr>
<tr>
<td>2010</td>
<td>7</td>
<td>7</td>
<td>100</td>
<td>ALP</td>
</tr>
<tr>
<td>2013</td>
<td>8</td>
<td>8</td>
<td>100</td>
<td>LNP</td>
</tr>
<tr>
<td>2016</td>
<td>5</td>
<td>4</td>
<td>80</td>
<td>LNP</td>
</tr>
<tr>
<td>2019</td>
<td>5</td>
<td>0</td>
<td>0%</td>
<td>LNP</td>
</tr>
<tr>
<td>Total polls</td>
<td>55</td>
<td>40</td>
<td>73%</td>
<td></td>
</tr>
</tbody>
</table>

Source: Compiled from Goot (in press, Table 2)

5.2. 2019 poll results

Since the early 1990s, when all the Australian election pollsters began presenting their results in terms of a two-party-preferred vote, this has become the principal way of evaluating their performance. Prior to this, polls were assessed by measuring their performance against the parties’ first preferences – or primary vote. While the two-party-preferred is a measure unique to Australia, the first preference vote has its analogues in every democracy and is the only measure that is used internationally. The final national poll results for 2019 are shown in Table 3.
### Table 3: Final national polls for the House of Representatives, 2019.

<table>
<thead>
<tr>
<th>Date</th>
<th>Poll</th>
<th>Mode</th>
<th>Sample size</th>
<th>Primary Vote Result(^{12})</th>
<th>2PP result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LNP %</td>
<td>ALP %</td>
</tr>
<tr>
<td>18 May 2019</td>
<td>Election Outcome</td>
<td>Online</td>
<td>1,201</td>
<td>41.4</td>
<td>33.3</td>
</tr>
<tr>
<td>10–14 May</td>
<td>Essential Online</td>
<td>1,201</td>
<td>38.5</td>
<td>36.2</td>
<td>2.3</td>
</tr>
<tr>
<td>12–15 May</td>
<td>Ipsos CATI + DFRDD</td>
<td>1,842</td>
<td>39</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>15–16 May</td>
<td>Newspoll IVR +</td>
<td>3,038</td>
<td>38</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>?? – 12 May</td>
<td>Roy Morgan F2F</td>
<td>1,265</td>
<td>38.5</td>
<td>35.5</td>
<td>3</td>
</tr>
<tr>
<td>13–15 May</td>
<td>YouGov Galaxy Online</td>
<td>1,004</td>
<td>39</td>
<td>37</td>
<td>2</td>
</tr>
</tbody>
</table>


---

\(^{12}\) LNP: Liberal Party of Australia and National Party combined (including Liberal National Party (QLD) and Country Liberals (NT); ALP: Australian Labor Party; PHON: Pauline Hanson's One Nation; UAP: United Australia Party.

\(^{13}\) NA: either not asked or not reported
5.3. Absolute error

The simplest way of evaluating the performance of a poll is by comparing the poll estimate to the election result in absolute terms; that is, the difference between each poll’s estimate of a party’s share of the vote and the share of the vote the party received – regardless of whether the difference represents an overestimate or an underestimate. The absolute errors for the final polls of the campaign for each pollster are shown in Table 4. Absolute error is measured for each poll in the following ways: (1) on the primary vote for each party on which all the polls reported; (2) on the primary vote for each party averaged across all the parties reported by all the polls – this is the average absolute error on the unweighted primary vote; (3) on the primary vote reported for each party averaged across all parties reported by all the polls, weighted by share of primary vote – this is the average absolute error weighted by share of primary vote; (4) on the difference, in primary votes, between the Coalition (LNP) and Labor (ALP) – this is the absolute error on the primary vote margin; and (5) on the difference between the Coalition (LNP) and Labor (ALP) poll estimates and the two-party-preferred outcome – this is the absolute error on the two-party-preferred vote.

<table>
<thead>
<tr>
<th>Poll</th>
<th>Primary vote (Absolute error) (1)</th>
<th>Primary vote (average absolute error) (2) and (3)</th>
<th>Absolute error on the primary vote margin (4)</th>
<th>Absolute error on the 2PP vote (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LNP</td>
<td>ALP</td>
<td>GRN</td>
<td>PHON</td>
</tr>
<tr>
<td>Essential</td>
<td>2.9</td>
<td>2.9</td>
<td>1.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Ipsos</td>
<td>2.4</td>
<td>0.3</td>
<td>2.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Newspoll</td>
<td>3.4</td>
<td>3.7</td>
<td>2.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>2.9</td>
<td>2.2</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>2.4</td>
<td>3.7</td>
<td>1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Average poll error</td>
<td>2.8</td>
<td>2.6</td>
<td>1.4</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

### Absolute error on the primary vote (1)

At the election, the Coalition achieved a first preference vote of 41.4%. Ipsos and YouGov Galaxy underestimated the Coalition’s primary vote by 2.4 percentage points. The Morgan Poll, which rounded its estimate to the nearest half of a percentage point, and the Essential Report, which produced estimates to one decimal point, underestimated the Coalition’s first preference vote by 2.2 to 2.9 percentage points. Newspoll underestimated the Coalition’s first preference vote by 3.4 percentage points.

If every poll underestimated the Coalition’s vote, every poll – save one – overestimated Labor’s share of the first preference votes. The Morgan Poll overestimated Labor’s share by 2.2 percentage points; the Essential Report by 2.9 points and YouGov Galaxy and Newspoll by 3.7 points. The exception was

---

14 The absolute error on the primary vote margin is computed as the absolute value of the margin (%ALP-%LNP) in the poll minus the same margin (%ALP-%LNP) at the election. This statistic is always positive, providing a sense of how much polls differed from the final vote margin but not indicating whether they missed more toward one party or another.
Ipsos, which underestimated Labor’s primary vote share by 0.3 percentage points – effectively a zero error.

**Average absolute error on the primary (unweighted and weighted by poll share) (2 and 3)**

There are five parties for which every poll produced an estimate: the LNP, Labor, the Greens, Pauline Hanson’s One Nation, including unpublished figures provided by Ipsos, and a category of Others.\(^{15}\)

The unweighted primary vote average absolute error of the polls across these parties ranged from 1.3 percentage points recorded by the Morgan Poll to 2.4 points recorded by the Essential Report. In between came Ipsos, with an average error of 1.4 points; YouGov Galaxy, 1.7 points; and Newspoll, 2.1 points.

A number of polls measured the minor parties’ vote more accurately than the major parties’ vote. The weighted primary vote poll average, where the absolute error on the primary vote for each party is weighted by the share of the vote for that party, therefore, may be a better overall measure of how well each poll performed relative to the election outcome.\(^{16}\) When weighted by the share of the vote, the average poll error increases for all pollsters and the ranking of the pollsters changes, with Ipsos now being well ahead of the other pollsters with just half of the Newspoll weighted average error – Newspoll being the worst-performing poll on this measure.

**Absolute error on the primary vote margin between the LNP and Labor (4)**

The absolute error on the primary vote margin between the LNP and Labor, ranged from 2.1 percentage points for Ipsos to 7.1 percentage points for Newspoll. This is the measure most commonly used to evaluate the performance of the polls in the US and the UK.

**Absolute error on the two-party-preferred vote (5)**

The right-hand column of Table 4 (Column 5) shows the absolute error for each party based on the two-party-preferred vote. The two pollsters with the smallest absolute error on the two-party-preferred vote for the Coalition were Ipsos and YouGov Galaxy. Both of these polls showed a two-party-preferred vote of 49% for the Coalition against an eventual outcome of 51.5%, underestimating the Coalition’s lead over Labor by 2.5 percentage points.

The final Essential Report and Newspoll had slightly larger absolute errors in terms of the two-party-preferred vote. The Essential Report underestimated the Coalition’s position relative to Labor by 3 percentage points. Newspoll, using a sample more than twice as large, also underestimated the Coalition’s position relative to Labor by 3 percentage points.

The poll that registered the largest absolute error on the two-party-preferred vote was the Morgan Poll. The final Morgan Poll estimate of the Coalition’s two-party vote fell short by 3.5 percentage points – one percentage point more than the estimates provided by Ipsos and YouGov Galaxy.

---

\(^{15}\) Ipsos unpublished data from ‘Fairfax Ipsos Poll M19Y0032_Tables_15_02_2019_Weighted_Client Income_Weights.pdf.’ Others’ includes the United Australia Party (UAP). The UAP is excluded from separate consideration because even though it attracted over 3% of the vote, Essential Research, unlike the other polls, did not provide a stand-alone estimate for the UAP.

\(^{16}\) E.g. The weighted primary vote error for Essential is calculated by multiplying the primary vote share in Table 3 by the primary vote error in Table 4. Thus LNP (.385 x 2.90) + ALP (.362 x 2.9) + GRN (.091 x 1.3) = PHON (.0606 x 3.5) + OTH (.08 x 1.2) = 2.6
**Overall assessment**

Assessing the relative performance of the polls across the four overarching error metrics, Ipsos was the first ranked or equally first ranked on all measures; and YouGov Galaxy and Roy Morgan on one out of four. Essential and Newspoll were not top ranked for any of the metrics used.

**Table 5: Summary of the polls performance by error metric, 2019.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Error range (%)</th>
<th>Polls with the lowest error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute error on the primary vote</td>
<td>1.3 – 2.0</td>
<td>Ipsos, Roy Morgan</td>
</tr>
<tr>
<td>Weighted absolute error on primary vote</td>
<td>1.5 – 3.1</td>
<td>Ipsos</td>
</tr>
<tr>
<td>Absolute error on the primary vote – margin</td>
<td>2.1 – 7.1</td>
<td>Ipsos</td>
</tr>
<tr>
<td>Absolute error 2PP</td>
<td>2.5 – 3.5</td>
<td>Ipsos, YouGov Galaxy</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

**5.4. Margin of sampling error**

To acknowledge the uncertainty associated with any sample survey, and to contextualise the difference between poll results and the final election results, pollsters and media often refer to a ‘margin of error’ (MoE) or ‘margin of sampling error’ (MoSE). In calculating their margin of sampling error the pollsters assume that their polls are based on a simple random probability sample of the voting population. The calculations used to determine the margin of error for each pollster under this assumption are shown in the Technical Notes provided at Appendix 3.

While all of the polls reported a margin of sampling error, none of the polls used a simple random sample and none factored into their reported margin of error calculations the complex nature of their sample design, the impact that weighting had on sample variance or the underlying nature of their sample. In most situations, weighting will increase sampling variability; however, this is not always the case. Sturgis et al. (2018, pp 763-765) use the 2015 UK election polls to show that when weighting variables that are highly correlated with voting intention are used, such as party affiliation or past vote choice, a design effect of less than 1 can be achieved, ‘indicating that the sampling variability is smaller than would be expected under simple random sampling’. To the best of our knowledge, none of the weighting schemes used for the 2019 Australian election polls would have produced this effect. Instead, their margins of error were calculated under the assumption of a simple random sample that only reflected the error associated with the sample size for each poll and not any differences in accuracy from the different sampling and weighting methods.
Non-probability samples do not allow the use of the traditional means of calculating sampling error that is afforded when one has a probability sample. This means that the margin of error calculations reported by Essential, Newspoll and YouGov Galaxy have no valid statistical foundation. On the other hand, the margins of error reported by Ipsos and Roy Morgan for their probability-based samples do have a valid underlying statistical foundation. However, both of these pollsters mis-report their true margins of error by not factoring in the increase in variance associated with their sample designs and weighting solutions. These are important limitations to note because the ‘margins of error’ reported by the pollsters are relied upon by journalists and readers to help them interpret the poll results and, as currently reported, these margins of error convey a false impression of the underlying statistical accuracy of these polls.

Rather than relying on ‘hypothetical’ sampling errors that are clearly not supported by polls’ sampling and weighting methods, more robust methods of estimating the variability associated with the polls should be encouraged. These days, an emerging consensus is that model-based approaches combined with repeated/replicate sampling methods provide options for estimating variances and confidence intervals for the non-probability samples; e.g., Sturgis et al. (2018), Elliott and Valliant (2017); and Baker et al. (2013). These methods, while more complex and more time-consuming to implement, will ‘better reflect the nature of the sampling and estimation procedures in the election polls’ and therefore provide more realistic expectations for the precision of poll estimates (Sturgis et al., 2018, p. 763).

Notwithstanding the above, a comparison of the polls in terms of their respective margins of sampling error, as they report them, is shown in Table 6; Figures 2 to 5 show the margins of error for selected poll estimates as shaded confidence intervals. YouGov Galaxy is the only pollster to have produced an estimate of the two-party-preferred vote for the Coalition (LNP) within their reported margin of sampling error. YouGov Galaxy is also inside of their stated margin of sampling error for their estimate of the Coalition’s primary vote. Ipsos was the only poll that estimated the Labor (ALP) primary vote within their stated margin of sampling error and YouGov Galaxy, Essential and Roy Morgan all estimated the Green (GRN) vote within their stated margin of sampling error.

Table 6: Margins of Sampling Error for LNP, ALP and Greens

<table>
<thead>
<tr>
<th>Poll</th>
<th>Reported Sample Size</th>
<th>LNP %</th>
<th>MoE %</th>
<th>ALP %</th>
<th>MoE %</th>
<th>GRN %</th>
<th>MoE %</th>
<th>2PP LNP %</th>
<th>MoE %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential</td>
<td>1,201</td>
<td>38.5</td>
<td>2.8</td>
<td>36.2</td>
<td>2.7</td>
<td>9.1</td>
<td>1.6</td>
<td>48.5</td>
<td>2.83</td>
</tr>
<tr>
<td>Ipsos</td>
<td>1,846</td>
<td>39</td>
<td>2.2</td>
<td>33</td>
<td>2.1</td>
<td>13</td>
<td>1.5</td>
<td>49</td>
<td>2.28</td>
</tr>
<tr>
<td>Newspoll</td>
<td>3,038</td>
<td>38</td>
<td>1.7</td>
<td>37</td>
<td>1.7</td>
<td>9</td>
<td>1.0</td>
<td>48.5</td>
<td>1.78</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>1,265</td>
<td>38.5</td>
<td>2.7</td>
<td>35.5</td>
<td>2.6</td>
<td>10</td>
<td>1.7</td>
<td>48</td>
<td>2.75</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>1,004</td>
<td>39</td>
<td>3.0</td>
<td>37</td>
<td>3.0</td>
<td>9</td>
<td>1.8</td>
<td>49</td>
<td>3.09</td>
</tr>
<tr>
<td>Election Result</td>
<td></td>
<td>41.4</td>
<td></td>
<td>33.3</td>
<td></td>
<td>10.4</td>
<td></td>
<td>51.5</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ calculations
Figure 2: MoSE 2PP LNP Vote
2PP LNP Election Result (51.5%)

- Newspoll (48.5%)
- YouGov/Galaxy (49%)
- Ipsos (49%)
- Essential (48.5%)
- Roy Morgan (48%)

Figure 3: MoSE LNP Primary Vote
LNP Election Result (41.4%)

- Newspoll (38%)
- YouGov/Galaxy (39%)
- Ipsos (39%)
- Essential (38.5%)
- Roy Morgan (38.5%)

Figure 4: MoSE LABOR Primary Vote
ALP Election Result (33.3%)

- Newspoll (37%)
- YouGov/Galaxy (37%)
- Ipsos (33%)
- Essential (36.2%)
- Roy Morgan (36.5%)

Figure 5: MoSE Green Primary Vote
GRN Election Result (10.4%)

- Newspoll (9%)
- YouGov/Galaxy (9%)
- Ipsos (13%)
- Essential (9.1%)
- Roy Morgan (10%)
5.5. Performance of the polls over a longer time frame

Controversy about the performance of election polls is not unique to Australia, with the 2015 UK general election and the 2016 US presidential election just some of the high-profile examples that have resulted in reviews of polling methodology and comprehensive reports into what went wrong (Durand and Blais, 2020). Jennings and Wlezien (2018) undertook an analysis of election polls around the world to investigate whether the recent performance of polls has been ‘outside the ordinary’. For their main analysis, Jennings and Wlezien compiled data on polls conducted during the last week of election campaigns in 220 national elections in 32 countries between 1942 and 2017. For our analysis we start with the Australian polling data compiled by Jennings and Wlezien from 1980 to 2016 and supplement it with polling data compiled by the Inquiry Panel for 1983, 1984, 2013 and 2019 so as to have a complete Australian time series from 1980 onwards.

In the charts that follow, it is not individual poll results that are shown but the ‘mean error’ of the poll results across all in-scope polls for those parties with a primary vote share of at least 3% – this being the same cut-off used by Jennings and Wlezien (2018). The results are plotted with a smoothing (LOWESS) curve with a bandwidth of 0.3. Details of the calculations of these measures are shown in Technical Notes in Appendix 3.

The average absolute poll error of 1.7 percentage points in 2019 (Figure 6), whilst not out of the ordinary, is the largest since 1998. The mean absolute error on the margin (i.e. the average poll error on the primary vote gap between the Coalition and Labor) was 5.2 percentage points in 2019, the least accurate result since 1987 (7.4 percentage points). So, whichever of these metrics is used the 2019 polls stand out as being the least accurate for some considerable time.

Figure 6: Australian polls, Mean Absolute Error, Primary Vote – All parties with ≥3% of the vote, 1980 – 2019**

Source: Supplemented Jennings and Wlezien

5.6. **International Comparisons**

Figure 8 and Figure 9 show that Australian polls from around the mid-1990s onwards have been, on average, more accurate (i.e. had less error) than the international trend.

---

17 The Canadian election polls for 2004 have been excluded on the advice of the paper’s authors due to an error that was detected when we were analysing this data.
Figure 8: International polls, Mean Absolute Error, Primary Vote – All parties with ≥3% of the vote, 1980 – 2019*

Source: Supplemented Jennings and Wlezien


All countries smoothed is based on the year average across all countries
Inquiry into the performance of the opinion polls at the 2019 Australian federal election

AMSRO Inquiry Panel

5.7. The performance of the polls by methodology

To what extent have different sampling and data collection methods affected the accuracy of the polls? Here we rely on data presented in Appendix 4 sourced from Goot (in press, Table 2). These data show the error in terms of the two-party-preferred vote for the Coalition for each pollster’s final poll for each election from 1993 to 2019, and the method used for each poll. Of the 55 polls, 34 used telephone sampling frames and Computer Assisted Telephone Interviewing for recruitment and data collection, eight used an online mode of data collection, six – all conducted by Roy Morgan – used area-based face-to-face interviewing, three used telephone sampling frames with an IVR mode of data collection, three used IVR in combination with an online panel, and one poll undertaken by Roy Morgan in 2013 used an unspecified mixed-mode design.

The average absolute error of each method is based on the two-party-preferred votes; CATI (1.4 percentage points, based on 34 observations), non-probability online panels (2.4 percentage points, 8 observations), face-to-face (3.2 percentage points, 6 observations), IVR (1.3 percentage points, 3 observations), and IVR in combination with online (1.2 percentage points, 3 observations). This data set is too limited to draw definitive conclusions about the relative accuracy of polling methods in...
Australia, apart from noting the good track record of CATI over a long period and the fact that if we look at the period since 2010, during which both CATI and online polling have operated, the gap between the methods has closed (average error for CATI =1.1 percentage points; online = 1.6 percentage points). More generally, what appear to be mode effects may be effects of other kinds. As Goot observes, ‘so dominant has been the use of one mode – CATI via landlines – that comparisons across modes need to be treated with caution, the more so when most modes are closely associated with a particular pollster and no allowance is made for house effects’ (Goot, in press).

American and British research may shed some light, although in both Britain and the US voting is not compulsory. US poll aggregator and poll analytics website, FiveThirtyEight, suggests that ‘the clearest trends are that telephone polls — including both live caller and IVR polls — have outperformed online polls in recent elections and that polls using mixed or hybrid methods haven’t performed that well’ (Silver, 2018b). In relation to the 2015 British general election, the British review found ‘no difference in accuracy between online and telephone polls’ (Sturgis et al., 2017, p. 59) — though no British pollsters use IVR. Comparing mode differences over two elections, 2010 and 2015, the British review concludes that ‘it is not unreasonable to assume that the generally higher Conservative share in the phone polls indicates that they were somewhat more accurate during the campaign period but this is, to be clear, only an assumption’ (Sturgis et al., 2017, p. 60).

Table 7: Summary of poll performance by methodology, 1993 to 2019.

<table>
<thead>
<tr>
<th>Methodology for Recruitment and Data Collection</th>
<th>First used</th>
<th>No. of final polls</th>
<th>Avg. absolute error LNP 2PP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993-2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATI</td>
<td>1993</td>
<td>34</td>
<td>1.4</td>
</tr>
<tr>
<td>Online</td>
<td>2001*</td>
<td>8</td>
<td>2.4</td>
</tr>
<tr>
<td>Face-to-face</td>
<td>1998</td>
<td>6</td>
<td>3.2</td>
</tr>
<tr>
<td>IVR</td>
<td>2010</td>
<td>3</td>
<td>1.3</td>
</tr>
<tr>
<td>IVR/Online</td>
<td>2016</td>
<td>3</td>
<td>1.2</td>
</tr>
<tr>
<td>Other mixed mode</td>
<td>2013</td>
<td>1</td>
<td>_</td>
</tr>
<tr>
<td>2010-2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATI</td>
<td>11</td>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>Online</td>
<td>5</td>
<td></td>
<td>1.6</td>
</tr>
</tbody>
</table>

Source: Derived from Goot (in press, Table 2)

# Mainstream use from 2010.

5.8. Is there partisan bias in the polls?

The fact the polls in 2019 underestimated the Coalition’s vote share raises the possibility that there is an anti-Coalition, or pro-Labor, bias built into the polls. This Inquiry found no conclusive evidence to support this claim. We do note, however, that Goot (in press, Table 2) shows that in 33 of the 55 final polls conducted before the last 10 elections, the Coalition’s two-party-preferred vote was underestimated; in only 21 was it overestimated, with one poll getting it exactly right (see Appendix 4). Breaking this down further, 16 of the 30 final polls from 1993 to 2007 (53%) underestimated the LNP vote compared to 17 out of 25 (68%) from 2010 onwards.

Goot notes that ‘some pollsters are more likely than others to have underestimated the Coalition’s vote – Roy Morgan in particular; since 1993, it has underestimated the Coalition vote on 10 out of 13 occasions. YouGov Galaxy (four out of six), Essential (three out of four), and Ipsos (two out of two) are
also more likely than not to have underestimated the Coalition’s vote. Newspoll has been as likely to underestimate as to overestimate the Coalition’s vote share (Goot, in press).
6. Factors that may have contributed to inaccuracies in the 2019 election polls

6.1. Commentary from the pollsters

As has been the case with other ‘polling misses’ around the world over the last five or so years, there was no shortage of opinions about why the Australian polls in 2019 were less accurate than usual. Media commentators and some academics, as well as the pollsters themselves, put forward a range of reasons, with some of the pollsters going one step further and announcing their own reviews so that some of their speculation could be tested.

The Ipsos Director of Polling, Jessica Elgood, suggested several explanations to Nick Bonyhady of *The Sydney Morning Herald* (Bonyhady, 2019):

- A last-minute swing to the Coalition after the final poll.
- Incorrect weighting of the raw data.
- An oversampling of politically engaged people.
- More voters opting for minor parties and independents than was reflected in the polls.

A late swing being missed by the polls was an explanation also put forward by Peter Lewis, Director of Essential Research, in *The Guardian* (Lewis, 2019). In addition, he pointed to declining voter engagement, shifting demographics and technological change. Lewis also pointed to the changed conditions for polling as a possible explanatory factor, it having become harder to get representative samples of the electorate due to the reduction in households with landlines, increasing refusal rates and declining response rates – though none of this could have explained the miss by Essential since its samples were based on a non-probability online panel.

David Briggs, Managing Director of YouGov at the time, seemed to believe that the problem was less one of the polling methodologies than one of interpretation, with those presenting and interpreting the polls failing to give sufficient weight to the seat-based polls showing a much closer result and a possible pathway to a Coalition victory (Briggs, 2019). That the single-seat polls were a better guide to the outcome is contestable as ‘not for the first time in an Australian election, the single-seat polling proved neither particularly accurate in estimating vote shares (first preferences or the two-party-preferred) nor especially good at forecasting which party in the various seats would win’ (Goot, 2020a, p. 160).

In his submission to the Inquiry, Campbell White, Head of Public Affairs and Polling Asia Pacific at YouGov, noted the challenges facing pollsters caused by the call screening, low response rates, differential non-response and low budgets, among other things.  

The findings from reviews into the performance of the polls in Britain and the US may also provide some pointers.

---

18 White points out that the recent access granted to the Integrated Public Number Database (IPND) that, among other things, contains a list of all residential landline and mobile phone telephone numbers in Australia and is available for ‘Federal, State and Local Government electoral matters’ (ACMA, 2019), is unlikely to improve the situation because of the public’s reluctance to participate in surveys. Still, it is worth noting that the Liberal Party’s pollsters Crosby|Textor (known since 2018 as the C|T Group), and the Tasmanian-based polling firm Enterprise Marketing & Research Services Pty Ltd, (owned by the C|T Group), were both granted access to the unlisted numbers section of the IPND; in effect, the mobile phone database referred to by White in May 2019 (ACMA, 2019).
The main findings from the British review were:

- The 2015 polls were among the most inaccurate since election polling began in Britain in 1945.
- Previous polls had been nearly as inaccurate but did not attract as much attention as they called the right result.
- The primary cause of the miss was unrepresentative samples. ‘The methods the pollsters used to collect samples of voters systematically over-represented Labour supporters and under-represented Conservative supporters. The statistical adjustment procedures applied to the raw data did not mitigate this basic problem to any notable degree. The other putative causes can have made, at most, only a small contribution to the total error’ (Sturgis et al., 2017, p. 4).

The main findings from the US review were:

- The national polls were among the most accurate polls since 1936 in estimating the popular vote, a 3-percentage point lead for Clinton against a final margin of 2 percentage points.
- The State polls showed a competitive, uncertain race but underestimated support for Trump in the Upper Midwest because of:
  - Late deciders – about 13% of voters in Wisconsin, Florida and Pennsylvania decided on their Presidential vote in the last week and broke decisively for Trump.
  - The failure of many polls to adjust for the over-representation of college graduates in their samples (AAPOR, 2017, pp. 2-4).

6.2. Analytical framework: Total Survey Error

The Inquiry Panel relied on the Total Survey Error (TSE) framework (Groves and Lyberg, 2010; Groves, 2004) to structure our analysis of the methodological and statistical factors that may have contributed to errors in the published polls. The TSE framework is the most universally accepted approach for evaluating the accuracy (reliability and validity) of a given survey or poll.

Whether a poll is likely to be sufficiently accurate for its purpose depends on many methodological and statistical aspects of how the poll is designed, conducted and analysed. In evaluating all of these attributes, the overriding consideration is the potential for total survey error (i.e. bias and variance) across a survey's two main dimensions – the measurement dimension and the representativeness dimension.

- Measurement: This dimension covers all those factors related to the collection of data from a respondent, including: the survey questions; the overall design of the questionnaire; the mode of administration; interviewer and respondent error; and errors in data processing (e.g. mis-coding).
- Representation: This dimension covers all of the factors associated with obtaining a representative sample from the population of interest, including: the coverage of the population by the sampling frame; the sample design; non-response from the sample as initially drawn; and statistical adjustments, such as weighting.

An additional source of error is inferential error. This error results from the erroneous interpretation and reporting of survey findings; for example, attributing more precision to polling or survey results than is supported by the design of the survey, or by inferring causality when neither the survey design nor the analytical method supports such an inference. National election polls face a particular difficulty
in this regard, as a critical element in interpreting poll findings is to infer from a national poll the winner of an election decided by the voting outcome across 151 electorates (see section 9.1).

A more detailed explanation of Total Survey Error is provided in Appendix 5.

In 2019, the national polls varied considerably across the two dimensions. The next two sections explore the possible causes of measurement error (section 6.3) and of inaccuracies in the representativeness of the samples (section 6.4).

6.3. Possible measurement errors

6.3.1. How voting intentions are measured

Measuring the first preference vote for the House of Representatives

The first task of the pollsters is to ask the voting intention question of the correct population, this being eligible and enrolled voters. Eligibility is defined as being an Australian citizen or eligible British subject aged 18 years or over who has lived at their current address for one month or more (AEC, 2020).

The questionnaire provided to this Inquiry by Ipsos and other materials we have gathered indicate that at least some pollsters attempted to screen out ineligible and non-enrolled voters. Ipsos screened out those under 18, and those not enrolled to vote. Essential enabled non-voters to identify themselves via the option ‘do not intend to vote’ (Essential Research, 2019, p. 10), this question potentially capturing both ineligible voters and eligible non-voters. The approach taken by Newspoll, Roy Morgan and YouGov Galaxy is not known.

Having identified the in-scope population (refer to section 6.4.1 for a more detailed discussion), the voting intention questions asked are summarised below.

Table 8: How each pollster measures voting intentions

<table>
<thead>
<tr>
<th>Pollster</th>
<th>Voting intention questions</th>
</tr>
</thead>
</table>
| **Essential** | **Primary question:** If a Federal Election was held today to which party will you probably give your first preference vote?
| | **Leaning question** (If ‘don’t know’ to the above): ‘Which party are you currently leaning to?’
| | **Response list:** Liberal, National, Labor, Greens, Pauline Hanson’s One Nation, Other/Independent and a ‘don’t know’ option.
| | **Rotation of response list:** Response options were rotated. |
| **Ipsos** | **Primary question:** A Federal Election for the House of Representatives will be held on Saturday 18th May 2019. Which party will receive your first preference vote?
| | **Leaning question** (If ‘don’t know’ to the above): ‘Which party do you have a leaning towards at present?’
| | **Response list:** Labor Party, Liberal Party, National Party, Greens, Centre Alliance, One Nation, United Australia Party (previously known as the ‘Palmer United Party’), Some other party or candidate, Don’t know / refused, Not enrolled to vote.
<p>| | <strong>Rotation of response list:</strong> Response options 1 and 2 (Labor and Liberal) were rotated. |</p>
<table>
<thead>
<tr>
<th>Pollster</th>
<th>Voting intention questions</th>
</tr>
</thead>
</table>
| **Newspoll**     | **Primary question**: ‘If a Federal Election for the House of Representatives was held today, which of the following would you vote for?’  
|                  | **Leaning question** (if ‘uncommitted’ to the above): ‘To which one of these do you have a leaning?’  
|                  | **Response list**: Liberal, National, Labor, Greens, One Nation, United Australia Party, Other/Independent and a ‘don’t know’ option.  
|                  | **Rotation of response list**: Not known                                                                                                                                                                                  |
| **Roy Morgan**   | **Primary question**: ‘At the Federal election for the House of Representatives on May 18 – which party will receive your first preference?’  
|                  | **Leaning question**: Not known                                                                                                                                                                                            |
|                  | **Response list**: Not known                                                                                                                                                                                                |
|                  | **Rotation of response list**: Not Known                                                                                                                                                                                  |
| **YouGov Galaxy**| **Primary question**: If a federal election was held today, which one of the following would you vote for?  
|                  | **Leaning question**: (If ‘uncommitted’ to the above): To which one of these do you have a leaning?  
|                  | **Response list**: Labor, LNP, Greens, Pauline Hanson's One Nation, Katter's Australian Party, Others  
|                  | **Rotation of response list**: Response 1-5 rotated.                                                                                                                                                                        |

Source: Compiled from publicly available information and information provided by the pollsters. All participating pollsters were asked to comment on the accuracy of the descriptions contained in this table but only Ipsos did so. The YouGov Galaxy description was current as of August 2019 and may not have been the approach used in their federal election polls.

**Question wording – Voting intentions**

The voting intention questions used by the pollsters vary. Essential appears to be the only pollster that asked respondents to which party they would probably give their first preference vote. Roy Morgan says it asked ‘which party will receive your first preference?’ Essential, Newspoll and (maybe) YouGov Galaxy asked about voting intentions if the federal election was held today whereas Ipsos and Roy Morgan asked about voting intentions ‘on election day’.19

From the available information, it looks like Essential, Ipsos, Newspoll and YouGov Galaxy asked a follow-up question/probe of those respondents who offered a ‘don’t know’ response to the initial vote choice question. This is known as the ‘leaning’ question and these pollsters used very similar wording to try and get undecided respondents to declare who they would probably give their first preference vote to.

The voting intentions questions thereby elicit two levels of response. A more definitive response is provided to the first-level question ‘which party will receive your first preference vote’, and, for those unable to answer this question, a less definitive response to the ensuing question, ‘to which party are you leaning?’ These two levels of response are conflated when voting intention figures are reported thus masking a degree of uncertainty inherent in the responses. Of the 1,842 respondents to the final Ipsos poll, 12% were classified as ‘don’t know’ in response to the original voting intention question; after the ‘leaner’, only 7% were reported to be ‘don’t knows’.

19 Emphases added.
There is another issue to be considered in relation to the certainty with which voting intentions can be measured. It is not only those who choose a ‘don’t know’ option or only declare a voting intention when prompted by the leaning question who may be uncertain about their vote choice. A proportion of those who state a voting intention when first asked may also be uncertain/not totally committed to this choice.

The issue of how ‘undecided’ voters are best dealt with is further discussed later in this section and how the pollsters identify early voters in their samples and accommodate them in their questionnaires is examined in section 6.3.6.

Response lists

Response alternatives

In asking for whom respondents intended to vote, the pollsters provided respondents with a list of options. These lists named some parties but not others (the number of parties contesting seats in the House was very large), but all allowed respondents to nominate parties or others (independents, most notably) not on the list. Essential (and possibly YouGov Galaxy) was the only pollster not to include the United Australia Party (UAP), and the only pollster not to produce a separate estimate of support for the UAP. Ipsos was the only pollster to include the Centre Alliance in their list but it did not estimate a vote share for the Centre Alliance. The inclusion or exclusion of parties from these response lists is an important consideration; in survey research, it is generally agreed that having a particular option included in a response set will increase the proportion of respondents selecting that option. This is because including an item in the response list flags the option as a possibility – one that might otherwise be overlooked. Relying on the use of the ‘other specify’ option requires more cognitive effort on the part of respondents. Other things being equal, the inclusion or exclusion of minor parties on the vote choice response lists can influence the estimates generated for these parties.

The ordering of response options

The order in which items appear in a response list can trigger ‘response order’ effects. Such effects occur when the ordering of response options influences the option selected. Pollsters need to be careful not to convey some sort of implicit hierarchy or preference to respondents by the way in which the vote choice options are ordered. To mitigate response order effects pollsters, at a minimum, should rotate the positions of the Labor Party and the Liberal and National parties (always listed together and in that order) on the response list.

Response tailoring

Another issue is whether or not the lists are displayed or read out to respondents correctly and only reflect the options available in the state in which they live. We know that Ipsos tailored their questionnaires so that instead of displaying the generic ‘Liberal Party’ in each state and territory, the ‘Liberal-National Party’ was the option presented to their Queensland respondents and the ‘Country-Liberal Party’ to those in the Northern Territory. All the Ipsos respondents could have had read out to them by interviewers response options for the Centre Alliance, which ran candidates in only three South Australian seats, and One Nation, which ran candidates in only 59 seats, as Ipsos did not tailor the display of these options (Muller, 2020, p. 7).

Context/Conditioning effects

Essential and Roy Morgan included their election questions in omnibus surveys that covered a wider range of topics, most of them unrelated to the election. This means there is a possibility of context or
conditioning effects whereby preceding questions influence subsequent questions – here, the voting intention questions.

**Deriving the two-party-preferred vote**

There is also some variation in how the pollsters derived the two-party-preferred vote. Essential and YouGov Galaxy derived a two-party-preferred estimate by allocating minor party preferences in accordance with the 2016 election count; Newspoll based its distribution ‘on recent federal and state elections’; Ipsos used the 2016 vote count and respondents’ stated preferences; and Roy Morgan used respondents’ stated preferences (see Table 9).

**Table 9: How each pollster derived their two-party-preferred estimate**

<table>
<thead>
<tr>
<th>Pollster</th>
<th>Method used for deriving the two-party-preferred vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential</td>
<td>The 2PP estimate was calculated by distributing the votes of the other parties according to the distribution of preferences at the 2016 election.</td>
</tr>
<tr>
<td>Ipsos</td>
<td>A combination of 2016 preferences and stated preferences. The ‘stated preference’ question asked of those respondents who declared a first preference choice for other than the Labor or Liberal/National parties was: ‘At the Federal Election you will be required to vote for all candidates in your electorate in order of preference. Given this, will you give a higher preference to the Labor Party, the Liberal Party’ [Options rotated]. A ‘don’t know’ option was also offered. Where respondents did not provide a preference, the actual flow of preferences from the previous election was used.</td>
</tr>
<tr>
<td>Newspoll</td>
<td>Recent federal and state elections</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>The final 2PP vote estimates were based on respondents’ stated preferences but the time series data on the Roy Morgan website provides 2PP estimates using both the stated preference model and the past election model that draws on the allocation of preferences from the two previous federal elections (see <a href="http://www.roymorgan.com/morganpoll/federal-voting/2pp-voting-intention-recent-2016-2020">http://www.roymorgan.com/morganpoll/federal-voting/2pp-voting-intention-recent-2016-2020</a>, viewed on 30 September 2020).</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>The 2PP estimate was calculated by distributing the votes of the other parties according to the preference flows at the 2016 election.</td>
</tr>
</tbody>
</table>

Source: Compiled from publicly available information and information provided by the pollsters. All participating pollsters were asked to comment on the accuracy of the descriptions contained in this table but only Essential and Ipsos did so. The YouGov Galaxy description was current as of August 2019 and may not have been the approach used in their federal election polls.

In short, three models appear to have been used to derive the two-party-preferred vote: 1) previous elections; 2) stated preferences; or 3) a combination of these approaches. Essential and YouGov Galaxy utilised the ‘previous election’ model; so, too, did Newspoll, though how they combined recent federal and state elections is not reported. Roy Morgan used the ‘stated preference model’, but we do not know how preferences were allocated when respondents did not know/did not report a preference between Labor and the Liberal-National parties. Ipsos combined the two approaches using stated preferences when a preference was provided by the respondent and previous elections when not.

An added complication with the stated preference model is that respondents being asked to state their preferences may often be responding without knowledge of the preference policy (the How to Vote card) of their party of first choice that, once understood, could change their intention. Based on data from the 2019 the Australian Election Study, 29% of voters followed the House of Representative’s How to Vote card of their party of first choice (Cameron and McAllister, 2019b, p. 19). To what extent they would have made the same choice, independent of the party’s advice, is of course unknown.
The flow of preferences

Is the method used for deriving the minor party preference flow consequential for arriving at an accurate two-party-preferred estimate and is it consequential in explaining the inaccuracy of the 2019 polls? To address this question Table 10 shows the preference flows from the Greens, One Nation, UAP, and others, to the LNP and ALP at the 2019 and 2016 elections.

In the case of the Greens, the minor party with the greatest vote share, Essential and YouGov Galaxy allocated 82% of the preferences to Labor and 18% to the LNP, based on the actual distribution in 2016. This is effectively identical to the actual preference flow reported by the AEC in 2019. This means that, in the case of the Greens, the preference flow at the previous election (2016) was a very good proxy for the actual preference flow at the next election (2019).

In the case of One Nation, Essential allocated its preferences on an equal share basis in accordance with the flow of One Nation’s preferences at the 2016 election (50.5% Liberal, 49.5% Labor). Had it been able to correctly estimate the One Nation preference flow (65% Liberal, 35% Labor), Essential would have increased its estimate of the Coalition’s two-party-preferred vote from 48.5% to 49.5% with a corresponding decrease for Labor, from 51.5% down to 50.5%. This is a substantial change due solely to the preference allocation model used for one party.

Writing after the election, Peter Lewis, Director of Essential Research, conceded that ‘there was clearly a need to review [Essential’s] methodology, particularly the way we weighted our sample and distributed preferences’ (Lewis, 2020). In relation to preference allocations, he announced that ‘we will now be asking participants who vote for a minor party to indicate a preferred major party. Only when they do not provide a preference will we allocate based on previous flows’ (Lewis, 2020).

Table 10: Preference flows at the 2019 and 2016 elections

<table>
<thead>
<tr>
<th>Election result: Two-party-preferred preference flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
</tr>
<tr>
<td>Party</td>
</tr>
<tr>
<td>Greens</td>
</tr>
<tr>
<td>One Nation</td>
</tr>
<tr>
<td>UAP</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>2016</td>
</tr>
<tr>
<td>Party</td>
</tr>
<tr>
<td>Greens</td>
</tr>
<tr>
<td>One Nation</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>


Another issue for pollsters who use the previous election for deriving their two-party-preferred estimate is how to handle a new party; notably, in 2019, the United Australia Party. Almost half a million (488,817) UAP preferences were distributed and these favoured the Coalition over Labor by almost two to one (65:35). Apart from Ipsos, how the pollsters handled the UAP is unclear. Ipsos allocated preferences for all the minor parties, including UAP, the same way. What that way was is not known.

It is not just the preference allocation model but also the accuracy of the estimate of the minor parties’ vote that is important in ensuring the accuracy of pollsters’ two-party-preferred estimates. Table 11 shows, for each pollster: (a) the percentage point error in their primary vote estimate for each party;
(b) the percentage point error in their two-party-preferred estimates; and (c) the percentage change in the accuracy of their two-party-preferred estimate relative to their primary vote estimate.

Four of the five polls were slightly less accurate in their two-party-preferred estimates than in their primary vote estimates for the Coalition. For Essential, Ipsos and YouGov Galaxy, the increase in their error was very marginal (0.1 percentage points); for Roy Morgan the increase was 0.6 points. Only Newspoll was more accurate, the error on its two-party estimate for the Coalition (3.0 points) being 0.4 points smaller than the error on its primary vote estimate (3.4 points).

For Labor, three of the five polls were slightly less accurate in their two-party-preferred estimate than their primary vote estimate; the average error for the Labor vote increased from 2.6 percentage points on the primary vote to 2.9 percentage points on the two-party-preferred vote. For Essential, again, the difference (0.1 percentage points) was very marginal; their two-party-preferred estimate was better than would otherwise have been the case due to their overestimate of the One Nation vote; sometimes errors work to cancel each other out. For Roy Morgan (1.3 points) and especially for Ipsos (2.8 points), the inaccuracies were substantially greater; Ipsos suffered greatly from their overestimate of the Greens vote. By contrast, the accuracy of the Newspoll improved by 0.7 points, and the accuracy of YouGov Galaxy by 1.2 points.

Data reported on the Roy Morgan Research website following the election shows that using actual preference flows produced an estimated two-party-preferred vote distribution of 48.5% (LNP), 51.5% (ALP). This is closer to the election outcome than the two-party-preferred vote originally reported by Roy Morgan using the stated preference model, LNP – 48%, ALP – 52% (Roy Morgan, 2020).

Table 11: A comparison of primary vote and two-party-preferred errors, in percentage points

<table>
<thead>
<tr>
<th>Poll</th>
<th>Error in LNP estimate</th>
<th>Error in ALP estimate</th>
<th>Error on minor party estimates (+/-)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary vote (a)</td>
<td>2PP (b)</td>
<td>% change (+/-) (c)</td>
</tr>
<tr>
<td>Essential</td>
<td>-2.9</td>
<td>-3.0</td>
<td>+3</td>
</tr>
<tr>
<td>Ipsos</td>
<td>-2.4</td>
<td>-2.5</td>
<td>+4</td>
</tr>
<tr>
<td>Newspoll</td>
<td>-3.4</td>
<td>-3.0</td>
<td>-12</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>-2.9</td>
<td>-3.5</td>
<td>+21</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>-2.4</td>
<td>-2.5</td>
<td>+4</td>
</tr>
<tr>
<td>Absolute poll average</td>
<td>2.8</td>
<td>2.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

The treatment of don’t knows and undecideds

The accurate measurement, and subsequent treatment, of respondents who don’t know or refuse to state their voting intention is consequential not only in terms of obtaining an accurate measure of voting intentions but also because ‘reporting’ the proportion of ‘undecided’ voters helps to convey an element of uncertainty about the voting intention estimates that may, in turn, encourage greater caution in the interpretation of polling results. The proportion of ‘undecided’ respondents is routinely disclosed by the pollsters; only Essential failed to publish them before the election. The proportion of voters identifying as ‘undecided’ in each of the final public domain polls was: Newspoll – 4%; Roy Morgan – 5%; Ipsos and YouGov Galaxy – 7%; and Essential – 8% (Goot, in press). How pollsters deal with the undecided – traditionally, by ignoring them – could be another source of error.
This is another area where Peter Lewis has foreshadowed a change at Essential Research, moving from a position where ‘respondents who select “don’t know” are not included in the results’ (Essential Research, 2019, p. 4) to one where these respondents are left in the count. If only 93% nominated the Liberal, National or Labor Party as their first or two-party preference, then the reported two-party-preferred estimate might read ‘47-46%, with seven undecided’ (Lewis, 2020).

The measurement and treatment of undecided respondents in the election polls also relates to our discussion of the ‘shy conservative’ effect (section 6.3.5) and ‘late deciders’ (section 6.3.7).

### 6.3.2. Ballot order effects

Political scientists have long been interested in ballot order effects because some advantage could accrue to the candidate listed first on the ballot paper, particularly in the context of Australia’s compulsory voting system. King and Leigh (2009, p. 71) found that being placed first on the ballot increases a candidate’s vote share by about 1 percentage point. The ballot order for the House of Representatives is determined via a randomised allocation of a ballot position to all nominated candidates in each seat. If one of the major parties finishes up with more first positions on the ballot paper than the other, this could affect the election outcome but also the accuracy of the election polls.

In 2019, we find that ballot position advantaged neither the Coalition nor Labor party. Table 12 shows that Labor drew first position on the ballot 22 times compared to the LNP's 21.

#### Table 12: Number and proportion of seats in which Labor or the Coalition drew first position

<table>
<thead>
<tr>
<th>First position on ballot paper</th>
<th>Number of seats</th>
<th>% of all seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>22</td>
<td>14.6</td>
</tr>
<tr>
<td>Coalition</td>
<td>21</td>
<td>13.9</td>
</tr>
<tr>
<td>Total seats where ALP or LNP drew first place</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td><strong>Total seats</strong></td>
<td><strong>151</strong></td>
<td><strong>28.5</strong></td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

If securing first place is an advantage, then being higher on the ballot paper than the other major party may also carry an advantage. Jackman (2005b) calculated that in 2004 every place up the list for a Coalition candidate was worth 0.13 percentage points, net of other factors. Table 13 shows that in 2019 the Coalition may have benefited very slightly. The advantage to the Coalition is partly explained by there being 11 ‘three-cornered contests’ (contests involving Liberal, and National as well as Labor candidates plus, of course, several other candidates). In three-cornered contests, you would expect one of the LNP candidates to be higher on the ballot than the Labor candidate two out of three times, and this is almost exactly what happened. In non-three-cornered contests, the Coalition had a higher ballot position 72 times (51%) to Labor’s 68 times (49%). Any increase in vote attributable to this would have been negligible (effectively zero).

#### Table 13: Number and proportion of seats in which Labor or the Coalition drew a higher ballot position

<table>
<thead>
<tr>
<th>Position on ballot paper</th>
<th>Number of seats</th>
<th>% of seats</th>
<th>Non-3CC seats</th>
<th>% of seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP higher than all LNP candidates</td>
<td>72</td>
<td>47.7</td>
<td>68</td>
<td>48.6</td>
</tr>
<tr>
<td>At least one LNP candidate higher than ALP candidate</td>
<td>79</td>
<td>52.3</td>
<td>72</td>
<td>51.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>151</strong></td>
<td><strong>100.0</strong></td>
<td><strong>140</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

Source: Authors’ calculations
6.3.3. Device effects

Three of the five pollsters collected their data online, either entirely (Essential and YouGov Galaxy) or in part (Newspoll). Several overseas studies show that a large and growing number of respondents in online panels use smartphones to complete web surveys (Antoun et al., 2017, p. 96). Information shared with this Inquiry by the Social Research Centre reveals that for the June 2020 wave of Life in Australia™, 6% of the panel responded via telephone offline, 46% via a desktop or laptop computer, 42% via a mobile phone and 7% via a tablet. The pollsters using online methods have provided no information to this Inquiry nor is any information available in the public domain to explain what, if anything, is done to mitigate the measurement issues associated with device effects.

There is an extensive literature detailing the methodological challenges associated with gathering survey data in the same mode but where two very different interfaces/devices – desktop/laptop compared with mobile/tablet – are used (see Couper, Antoun and Mavletova 2017, pp. 133-150). Some of the effects associated with taking online surveys via small screen devices include longer questionnaire completion times, but quicker response times, higher break off rates, increased primacy effects (analogous to a ballot order effect), and increased use of responses that appear on the screen without the need for vertical or horizontal scrolling (Callegaro, Manfreda and Vehovar 2015). ‘Mobile optimisation’ is the method used to try and mitigate some of these potential effects. The Inquiry Panel has no way of knowing the extent to which such factors may have contributed to measurement error in those polls that used online panels.

6.3.4. Were pollsters deliberately misled by respondents?

Following the election, there was conjecture that some respondents may have deliberately misreported their voting intentions. From a Total Survey Error perspective, this would be an instance of respondent-related measurement error. If this misreporting was disproportionately attributable to supporters of a particular party, then the error would bias the data. The possibility of this form of measurement error is canvassed by two statisticians, Adrian Barnett and Scott Sisson, in The Conversation (Barnett and Sisson, 2019). ‘A number of Australian voters have admitted to giving fake answers to telephone polls’, they claim, ‘because they were annoyed by the call’. As they say: ‘Even a small percentage of annoyed people could ruin an otherwise well-designed opinion poll.’ But they fail to provide any evidence of ‘fake answers’ let alone deception on a scale, and with a sufficient skew, to support the view that it skewed overall results.

The review that followed the 2015 British polling miss concluded that ‘while it is very difficult to definitively rule out the possibility that respondents lied to pollsters before and after the election, circumstantial evidence from contemporaneous surveys suggests that it is unlikely to have been a contributory factor in the polling miss (Sturgis et al., 2016, p. 42).

Had deception been a factor in the Australian case, the proportion of respondents engaged in such misreporting would need to have been considerable, and evident across all polling methodologies. There would also need to be a case for supposing that this phenomenon was more prevalent in the 2019 election polls than in earlier elections, and that such misreporting was skewed in favour of Labor and against the Coalition. While it is true that the polls have been somewhat more likely to underestimate the Coalition’s share of the two-party-preferred vote rather than Labor’s in recent elections (Goot, Table 2, in press), there is no valid evidence to support the hypothesis that the cause of this has anything to do with respondents setting out to deceive.

6.3.5. Shy conservatives

Another form of respondent-related measurement error that has received attention is what has been called, in Britain, the ‘Shy Tory’ effect (Worcester et al., 2015, pp. 30-32) or ‘shame factor’ (Crewe
1992, p. 489); in the US, the ‘Shy Trump’ effect (AAPOR, 2017, p. 8); and most recently, in Australia, the ‘shy conservative’ effect. Respondents are said to be ‘shy’ or ‘ashamed’ because they assume that interviewers will not hold the same preference, and will look down on them for holding it.

In the US, the AAPOR Taskforce on the 2016 election polls concluded that '[a] number of Trump voters who participated in pre-election polls did not reveal themselves as Trump voters until callback studies conducted after the election (and they outnumbered late-revealing Clinton voters)', but noted that these ‘could be attributable to either late deciding or misreporting (Shy Trump) in the pre-election poll.’ However, '[a] number of other tests for the Shy Trump hypothesis yielded no evidence to support it, including differences between polls with live interviewers and those with no live interviewers’ (AAPOR, 2017, p. 52).

The Inquiry into the 2015 British polls explored the possibility of something analogous – a ‘shy Tory’ effect – by comparing the results from pre-election polls with those from re-contact surveys. It concluded that ‘while it is very difficult to definitively rule out the possibility that respondents lied to pollsters before and after the election, circumstantial evidence from contemporaneous surveys suggests that it is unlikely to have been a contributory factor in the polling miss’ (Sturgis et al., 2016, p. 41). It noted that social desirability could work both ways, in some cases actually favouring the Conservatives, and that the net effect would have to be substantial for it to have any bearing on the accuracy of the polls. It also noted that anyone reluctant to declare a party preference could be more likely to offer a don’t know/undecided response than to nominate another party.

In Australia, if ‘shy conservatives’ were abroad in large numbers, one would expect to see a lower primary vote for the conservative parties in the Ipsos and Roy Morgan polls – the only polls that used live interviewers; Ipsos via telephone interviewing and Roy Morgan via face-to-face interviewing. Ipsos recorded a primary vote for the Coalition of 39%, Roy Morgan, 38.5%. But these figures were scarcely lower than the figures recorded by the polls that used online panels or online panels with robopolls: the Essential Report estimated the Coalition’s vote at 38.5%; Newspoll at 38%; and YouGov Galaxy at 39%. Indeed, while Ipsos and Roy Morgan estimated support for Pauline Hanson’s One Nation (a conservative party) at 4%, YouGov Galaxy and Newspoll estimated its support at 3% – precisely the reverse of what one would expect if ‘shy conservatives’ were a factor. The pollster that recorded the highest vote for One Nation (6.6%) was Essential, which relied solely on an online mode of data collection. On its own, this is consistent with a ‘shy conservative’ effect. But it’s also consistent with an overestimate of the One Nation vote: One Nation won just 3.1% of the vote.

On balance, therefore, there is no empirical evidence for a ‘shy conservative’ effect in the 2019 national election polls. As the electoral analyst, Kevin Bonham, argues, ‘unless a respondent is very paranoid, they’re hardly likely to care about admitting they vote Coalition to a robopoll or an online survey’ (Bonham 2019).

6.3.6. Early voting

Before election day, over 40% of electors had cast their ballot by postal voting or by voting early at a polling centre.20 Muller (2019) reports that ‘4.8 million people [used] pre-poll voting, compared to around 3.0 million in 2016—an increase of some 60 per cent.’

Figure 10: reproduced from McAllister and Muller (2018) but updated to include the 2019 election, shows the trend in early voting from 1993 to 2019. Clearly, ‘early voting has increased significantly since the mid-2000s … in the 1998 Australian federal election, just over one in 10 votes were cast early, with the remaining nine out of 10 electors casting a ballot in person on polling day. By 2016, 20 Early votes cast at polling stations include those cast within one’s own electorate, ‘ordinary pre-poll votes’, and ‘declaration pre-poll votes’ cast at a polling station outside of one’s electorate.
almost one voter in every three was casting an early vote’ (McAllister and Muller 2018, p. 104). When we add 2019 to this series, the proportion of voters casting their vote prior to election day increases from 31.2% in 2016 to 43.4%.

**Figure 10: Pre-poll voting over time, 1993 to 2019**

Source: 2019 data sourced from Australian Electoral Commission downloads viewed on 9 July 2020

High levels of early voting should actually make the task of pollsters somewhat easier, reducing the likelihood that an undetected late swing of non-trivial size would affect the reliability of the poll’s estimates. If the pre-election polls are adequately capturing early voters, then as we draw closer to election day the proportion of respondents who have already cast their vote should also increase. Correspondingly, the proportion of potential ‘late deciders’ should decline, since early voters are less likely to be interested in politics and committed to a party than those who vote on election day (McAllister and Muller 2018, p. 107).

Some pollsters did attempt to identify early voters in their pre-election polls. Ipsos asked respondents identified as enrolled voters: ‘Have you already cast your vote, by post or pre-poll, or are you intending to vote in person on Saturday 18th May?’. Essential asked a ‘type of vote’ question with the following options: ‘I have already placed my vote / I will vote at a pre-poll before election day / I will vote by post / I will vote at a polling station on election day / I’m not sure how I will vote / I don’t intend to vote / I intend to vote informal’ (Essential, 2019). The approach taken by YouGov Galaxy, Newspoll and Roy Morgan is not known.

An analysis of early voting at the 2013 election, when early voting accounted for 26.4% of all votes cast, used AEC administrative data combined with Australian Bureau of Statistics census data to identify the individual and area-level characteristics associated with early voting. It showed that ‘the age distributions skew youngest for pre-poll declaration votes, somewhat older for pre-poll ordinary votes and oldest for postal votes (Rojas and Muller 2014, p. 4), that ‘divisions with low levels of education were more likely to have higher numbers of pre-poll ordinary votes (p. 5)’, and that ‘divisions with a very high percentage of people earning around the median income were 3.6 times more likely (2010) or 4 times more likely (2013) to have high percentages of pre-poll ordinary votes’ (p. 6). But they found ‘no particular patterns among rural, outer and inner metro areas’ (p. 3).
The daily uptake of early voting for the 2013, 2016 and 2019 elections, in Figure 11, shows the cumulative increase in early voting from the day pre-polling booths opened until election eve. About two thirds of early votes were cast in the last week of the election campaign; in 2019, pre-poll voting at around 11% on Day 9 of pre-polling had increased to 31% by Day 16, election eve.

Figure 11: Cumulative early voting patterns by day, 2013, 2016 and 2019

With the final polls of the 2019 campaign in the field between 10 May and 17 May, it is reasonable to expect that between about 15% and 35% of those being interviewed would have been early voters. From their poll of 12-15 May, Ipsos report that 25% of respondents had already voted with another 5% intending to cast a postal vote or a pre-poll vote. This aligns well with the take-up of early voting seen in Figure 11. The Essential Research report of their poll from 10-14 May shows that 8% of respondents had already voted, another 20% intended to vote by pre-poll and 12% reported intending to vote by post (Essential, 2019). Again, these figures are well aligned with the early voting trends shown in Figure 11.

The McAllister and Muller analysis of early voting uses the self-reported voting behaviour of respondents from the 2016 Australian Election Study (AES). The AES is a post-election survey of a random sample of voters, conducted after each federal election since 1987 – the 2016 AES had a sample size of 2,818 (Australian Election Study, 2019). ‘The 2016 AES found that 69% of the respondents reported casting a vote in person on polling day, which is exactly the same as the official figure’ (McAllister and Muller, 2018, p. 106).

Based on the material available, it seems that pollsters are accounting adequately for early voters in their polls. It follows that this is unlikely to be a source of measurement error or non-response bias and therefore is not an explanation for the poll ‘miss’. The seemingly adequate representation of early voters in the election polls also reduces the likelihood of an undetected late swing to the Coalition.

6.3.7. Late deciders

A potential problem with translating a measure of voting intentions into an election forecast is that at the time of the last poll some respondents haven’t made up their mind who they are going to vote for, some will change their mind, some are not eligible to vote, some won’t cast a vote and some will vote informally. As the inquiry into the performance of the 2015 British polls observed: ‘If a sufficient number of these types of voters move disproportionately to one party between the final polls and election day, the vote intention estimates of the polls will differ from the election result’ (Sturgis et al.,
2016, p. 34). Late deciders contributed to polling misses in Britain in 1970 and in 1992 (Sturgis et al., p. 34) and provide a partial explanation for the polling miss in the 2016 US election ‘in key battleground states, particularly in the Upper Midwest’ (APPOR 2017, p. 20).

Does an undetected late swing hold the key to the polling miss in Australia? In 2019, the field dates for the final polls were from an unknown start date to 12 May (Roy Morgan), 10-14 May (Essential), 12-15 May (Ipsos and YouGov Galaxy) and 14-17 May (Newspoll). The gap between the pollsters’ last day in the field and election day (18 May) was six days for Roy Morgan, four days for Essential, three days for both Ipsos and YouGov Galaxy, and one day for Newspoll. Benson (2019) reported that for the final Newspoll of the campaign 2,108 of the 3,038 interviews were conducted within the first 24 hours of the poll being released. If there was a late swing to the Coalition, the final Morgan Poll should have been the poll most likely to underestimate the Coalition vote. And it was: the final Roy Morgan poll estimated a two-party-preferred vote for the Coalition of 48%, a half a percentage point lower than the final Essential Report and Newspoll estimates, and one percentage point lower than the final Ipsos and YouGov Galaxy estimates. However, the final Morgan Poll did not have the highest proportion of undecided respondents. The proportion of voters identifying as ‘undecided’ in each of the final polls was: Newspoll, 4%; Roy Morgan, 5%; Ipsos and YouGov Galaxy, 7%; and Essential, 8% (Goot, in press).

Another way of trying to determine if there were a swing towards the Coalition as polling day approached is to see whether there was a swing towards the Coalition across the course of the campaign. The evidence for this is mixed. Three of the pollsters showed no such swing; if anything, they showed the Coalition’s vote in decline. Of the five polls conducted by Newspoll (11 April to 17 May), the Coalition primary vote went from 39% to 38%. In the four Roy Morgan polls (20 April to 12 May), the Coalition’s primary vote went from 37.5 to 38.5%, having peaked at 39.5% some three weeks before the election. The three polls undertaken by Essential (14 April to 14 May) estimated the Coalition primary vote to be 39%, 38% and 38.5% respectively. Two pollsters did show a swing towards the Coalition. According to the two polls by YouGov Galaxy (23 April, 15 May), the Coalition’s primary vote increased from 37% to 39%; and in the two Ipsos polls (1 May, 15 May), the Coalition’s primary vote increased from 36% to 39%.

If there were a late swing, we might expect evidence of an increasing proportion of voters deciding their vote late. But, in the AES, the proportion of late deciders in 2019 hardly stands out. Since 1993, the proportion of respondents who reported making up their mind during the election campaign (as opposed to a long time before) has ranged from 27% (1987) to 42% (1996 and 2016). In 2019, it was 37% (Cameron and McAllister 2019, p. 18).

On the other hand, a post-election poll, conducted by Essential, and reported by The Guardian, purports to show that there was a late swing. According to this study, not made available to Inquiry, 26% of the respondents had not ‘made up their minds as the federal campaign entered its closing weeks’, a proportion that ‘declined to 11% by polling day’ (Murphy; 2019b). Of these respondents, ‘38 per cent ’broke Morrison’s way and 27 per cent Bill Shorten’s way.’ These figures, however, do not support the late swing hypothesis, as Goot shows. One problem is that those who made up their mind either during the last week or on election day weren’t necessarily those categorised as “don’t know” in Essential’s final poll; that group may have included respondents who indicated a party preference but hadn’t made up their minds about whether that was the way they would actually vote.’ Another, is that ‘we are not told what proportion of respondents changed from one party (which?) to another party (which?) rather than simply confirming an earlier decision to vote for one of the parties and not another. Without knowing any of this, there is no way of estimating the impact … [of] the distribution on overall voting-intention figures’ (Goot, 2019).

To see what difference a 38–27 split would have made (38–27–35, allowing for ‘others’) requires us to compare it to the 39–36–25 split when it was assumed that the ‘undecided’ would divide in much the
same way as the rest of the sample. Since the proportion of ‘don’t knows’ in Essential’s final pre-election poll was 8% (not 11%), the new ratios imply that 3% (unchanged) would have voted for the Coalition, 2% (rather than 3%) would have voted Labor, and 3% (instead of 2%) would have voted for some other party. Based on these figures, Goot (2019) calculates that Essential’s two-party-preferred estimate for the Coalition ‘would have risen from 48.5 to 48.8 per cent (49 per cent, if we round up to nearest 0.5 per cent)’. Clearly, this is not sufficient to support the late swing hypothesis.

Two other pieces of research also produced data that run counter to the late swing hypothesis. One, a post-election online poll of 1,000 voters conducted by JWS in the two days after the election, reports that 39% of Coalition respondents decided their vote in the last week of the campaign compared with 37% of Labor voters (JWS, 2019, p. 11). The other was an ‘exit poll conducted by YouGov Galaxy for Channel 9, which purported to show that the government was headed for defeat — and by a similar margin to the pre-election polls’ (Goot 2019).

Re-interviews are another way of detecting whether there has been a late swing. Ideally, respondents from a final pre-election poll are invited to take part in a follow-up poll to see whether the way they said they intended to vote differs from how they say they did vote. The Inquiry is not aware of any pollster having undertaken such a poll.

A different sort of study that did re-contact respondents who had been interviewed some time before the election was conducted by Nick Biddle from the ANU Centre for Social Research and Methods via Australia’s only probability-based online panel.21 His first survey was in the field between 8 and 26 April, with half the responses gathered by 11 April, five weeks ahead of the 18 May election; the follow-up survey was conducted between 3 and 17 June, with half the responses gathered by 6 June, three weeks after the election. The analysis was based on the 1,844 respondents who participated in both surveys.

In the first survey, the Coalition led Labor on first preferences by 3.8 percentage points; in the second survey, when the same respondents were re-interviewed, the Coalition led Labor by 6.4 percentage points. Among those who participated in both waves, the Coalition’s support increased from 38% to 42.2%, Labor’s support increased from 34.1% to 35.4%, while support for ‘Other’ (excluding the Greens), fell from 14.4% to 8.7%. Only for the Greens (13.6% in April and 13.7% in June) was there no shift. Biddle concludes that the disproportionate swing to the Coalition from ‘other’ parties helps explain the election outcome. However, as explanation of the polling miss, this is unpersuasive. As Goot (2019) points out, the election polls did not overestimate support for the minor parties; their estimates, by and large, were reasonably accurate.

Those who didn’t know how they would vote in April (6.5% of respondents) ultimately favoured the Coalition, 42% of this group reporting that they finally settled on the Coalition compared with 21% settling for Labor, 24% for the Greens and 12% for ‘other’. These data certainly show a lopsided result (two to one in favour of the Coalition over Labor). However, Goot (2019) points out, the ‘swing’, even if it had occurred when voters actually turned out to vote, is not sufficient to explain the 8.1 percentage point gap in first preference between the Coalition and Labor at the election; at best, it would explain about 3 percentage points of this gap. That a greater proportion recalled voting for the Coalition after the election than the polls indicated prior to the election is not at all surprising; an inflated ‘recalled’ vote for the election winner is a commonplace of post-election polls (Goot 1973, pp. 266-67).

In most elections, panel surveys and re-interviews with pre-election poll respondents do not support the [late swing] theory’ (Prosser and Mellon 2018, p. 768). Even if all of the respondents in the 2019 AES who reported making up their minds during the election campaign (37%) had made up their minds after the last of the polls, the net two-party swing needed to account for the 3 percentage points’

difference between the polls and the election result would have been 8-9 percentage points. Given an overall swing to the Coalition at the ballot box of 1.1 percentage points, nothing like this is likely to have happened. In short, we find no compelling evidence of a late swing – let alone one sufficient to account for the differences between the polls and the result.

6.4. Possible errors of representation

6.4.1. Coverage

The population of interest

The population of interest for national election polls is Australian citizens and eligible British subjects aged 18 years and older, who have lived at their present address for at least one month (AEC, 2020).

Adopting a TSE perspective (refer to section 6.2 and Appendix 5: ) necessitates giving consideration to how well the population of interest is ‘covered’ (represented) by the various sampling frames (i.e. sample lists) used by pollsters. If the sample frames exclude some elements of the population of interest then there is under-coverage and if the frames contain duplicate units or units beyond the population of interest (e.g. non-eligible or non-enrolled voters) then there is over-coverage.

The potential over-coverage arising from the inclusion of out-of-scope members of the population in these sample frames is considerable (person aged less than 18 years, non-citizens and other ineligible and unenrolled voters), but is effectively mitigated by the use of screening questions (refer to section 6.3.1). Under-coverage is the more serious concern and stems from coverage ‘gaps’ created (variously) by the non-coverage of those without telephone access, without internet access or living in non-private residential dwellings. To understand the extent to which under-coverage error may explain the inaccuracy of the polls requires an examination of each of the sampling frames used by the pollsters.

Non-probability online panels

Essential, Newspoll and YouGov Galaxy used non-probability online panels, either in whole or in part. Such panels suffer from sizeable under-coverage, with somewhere between 9% and 14% of Australian households not having internet access and therefore unable to participate in surveys conducted via online panels (ACMA, 2020, p. 1; ABS, 2018, p. 1). Even proponents of such panels describe them as a combination of ‘idiosyncratic contacts, low response rates, and non-coverage’ (Rivers, 2013, p. 111).

In Australia, recruitment of non-probability online panels is undertaken by various means, including a combination of online and offline methods. Recruitment methods include ‘banner advertising on websites, invitations and messaging, partnerships, print media, online marketing, direct mail, social media, referral programs and piggy backing of CATI/CAPI surveys’ (Pennay et al., 2016, p. 3). This means that non-probability-based panels have no known frame from which they select their samples. The non-coverage inherent in all of these panels is undoubtedly very large and is non-random in nature. It is non-random because those invited to join are different in many ways from those not invited and ‘these differences are often correlated with what is being measured in surveys’ (p. 5). International research cited in section 4.1 shows that polls based on non-probability samples generally yield less accurate election results than those that use probability samples, and there is some support for this finding from the available Australian data (see section 5.7).

An even larger problem faced by non-probability online panels is the opt-in (self-selection) nature of the sampling mechanism. This issue is further explored in section 6.4.3.
Dual-frame RDD (DFRDD)

For their 2019 election polls, Ipsos used two sampling frames – landline phone numbers and mobile phone numbers – in equal proportion. The use of two sampling frames to cover the population of interest is known as a dual-frame sample design or, when used for telephone surveying, Dual-Frame Random Digit Dialling (DFRDD). In 2017-18, only 2% of the Australian adult population were without a telephone service, according to a National Health Survey (Phillips et al., 2019, p. 6).

The Inquiry regards this as an ignorable level of under-coverage for election polling, even if those not covered by the two phone frames have quite different voting intentions from those who are covered. However, the use of DFRRDD does require that the overlapping chance of being selected into a sample of persons who have both a landline and a mobile phone is appropriately accounted for, as should also be the case for those with an increased chance of selection as a result of being contactable via more than one landline or mobile phone number. These chances of selection and over-coverage adjustments ought to be made via the weighting process as discussed in section 6.4.4.

Landline sample frames

Newspoll used a different dual-frame strategy – a non-probability online panel and a frame comprising landline telephone numbers. The number of interviews obtained from each of these frames varied but was generally around 50:50. YouGov’s rationale for using a dual-frame approach for Newspoll ‘was to overcome deficiencies in online panels’. However, what eventuated was a combination of two sampling frames, both with poor coverage properties. As at June 2019, 51% of the adult population were mobile-only while 49% still had a landline connection (see section 3.2). The landline frame disproportionately comprised older people, people with lower levels of educational attainment, and people who spoke no language other than English, and was very different from the rest of the population on a range of household characteristics; for example, household tenure (Phillips et al., 2019). In a nutshell, the landline sampling frame, on its own, had very poor coverage properties and was likely to generate non-ignorable coverage error in relation to election polling.

Area-based sample frames

Roy Morgan Research describes its sampling frame/survey design as follows: ‘face-to-face interviewing with a representative cross-section of ... Australian electors aged 18+.’ It appears that its 2019 election polls were run on the Roy Morgan Single Source Omnibus survey vehicle. The coverage properties of the Single Source are: ‘All States and Territories • 11 major geographic strata • Sydney • Melbourne • Brisbane • Adelaidea • Perth remaining areas of • NSW/ACT • Vic • Qld • SA/NT • WA and Tasmania • 58 specific readership strata • All major community and regional newspaper distribution areas • All major shopping centre catchment areas • All Federal Electorates’ (Roy Morgan, 2019b, p. 5). It is likely that the frame did not cover, or did not adequately cover, remote and very remote regions. Nor would it have covered persons residing in non-residential dwellings or homeless persons. Without knowing more, it is not possible to say definitively whether or not it suffers from non-ignorable coverage bias.

Mitigation strategies of coverage error and concluding remarks

The strategies used by pollsters to mitigate the effects of non-coverage error on their final estimates included stratified sampling, quotas (generally by age, gender and geographic area) and weighting adjustments. While these strategies are likely to have been of some use, research internationally and in Australia shows that they are much more likely to be effective when applied to probability-based samples than non-probability samples (Pennay et al., 2018; Yeager et al., 2011).

The overall adequacy of the mitigation strategies outlined above is further considered in section 6.4.3. Given that such strategies are used to moderate not just the impacts of coverage error but also
sampling error and non-response error, their impact needs to be evaluated with regard to the cumulative effect of errors of representation and not on coverage error alone.

6.4.2. Sample design and sample size

Having decided upon the sampling frame(s), the next stage in the design of a poll is to draw a sample from that frame and to invite members of the ‘designated sample’ to participate in the poll. The way the sample is designed, the size of the resultant sample (i.e. the number of completed interviews/questionnaires), and the form of weighting used all contribute to sampling error.

With a non-probability sample, the most common sampling method is to set quotas for the number of questionnaires/interviews to be completed by respondents with certain characteristics. These quotas are most commonly specified by age group, gender and location – state/territory and possibly region. The Inquiry is not privy to the specific sampling strategies used by the pollsters; for example, the size of these quotas, whether they were independent or interlocking (although Lewis, 2019, talks of a ‘system of interlocking quotas’), or whether they were set within a stratum or at the level of the total sample. We know that the Ipsos approach, using DFRDD CATI, required 50% of completed interviews to be obtained from the mobile phone frame and 50% from the landline frame. If Ipsos used location-based quotas, they could have been applied only to the landline sample as RDD mobile phone numbers in Australia do not contain any geographic information. It is also the case that it is difficult to calculate meaningful confidence estimates or margins of sampling error when non-probability sampling frames and quota samples are used.

Both Ipsos and Roy Morgan endeavoured to mitigate the under-representation of young people in their samples by asking to interview the youngest in-scope male/person in the household at home at the time of contact, young respondents being among the respondents most difficult to secure. A probabilistic sampling approach would be to randomly select the household member invited to take part in the interview; for example, by selecting for interview the member of the household with either the most recent or next birthday. As Posser and Mellon note ‘if a relevant attribute such as vote choice is correlated with being available to take the survey within demographic groupings (i.e. being at home when the interviewer calls), then quota samples are likely to give incorrect answers (Prosser and Mellon 2018, p. 761).

The sample sizes for the final polls of the campaign ranged from 1,004 (YouGov Galaxy) to 3,038 (Newspoll). The maximum hypothetical margin of sampling error (refer to section 5.4 for the reasons why we consider this a hypothetical margin of sampling error) for samples of these sizes ranged from ±3.1% for the YouGov Galaxy sample to ±1.8% for the Newspoll sample. The gap in the two-party-preferred vote between Labor and the Coalition over the ten elections from 1993 to 2019 ranges from 0.2 percentage points in 2010 to 7.2 percentage points 1996 and averaged 3.6 percentage points. Six of the ten elections over this period were decided by a margin of 3 percentage points or less.

The MoE calculations for a 3 percentage points gap between Labor and the Coalition are shown in Table 14 for the sample sizes used by the polls in 2019 election (refer to Appendix 3 Technical Notes for the details of MoE calculations), adjusted for a design effect of 1.522. Based on these calculations,

---

22Polls are not simple random samples and as a result have additional sampling errors (expressed as design effects) due to the complex sample designs and complex weighting solutions they use (see section 6.4.4 and Appendix 5: ). For our calculations, a design effect of 1.5 has been used for illustrative purposes because we don’t have access to the data necessary to calculate the actual design effects for individual polls. This seems a reasonable assumption; for the Comparative Survey of Election Systems (refer to Table 23), the design effect when the sample was weighted by Age, Sex and Location was 1.56.
none of the polls had sample sizes sufficiently large to detect a 3 percentage point winning margin on a two-party-preferred basis with 95% confidence.

**Table 14: Estimated Margin of Error given a 2PP gap of 3 percentage points.**

<table>
<thead>
<tr>
<th>Poll</th>
<th>Sample Size</th>
<th>Effective Sample Size (assumes design effect of 1.5)</th>
<th>MoE (alpha=.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YouGov Galaxy</td>
<td>1,004</td>
<td>669</td>
<td>7.2</td>
</tr>
<tr>
<td>Essential</td>
<td>1,201</td>
<td>801</td>
<td>6.6</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>1,265</td>
<td>843</td>
<td>6.4</td>
</tr>
<tr>
<td>Ipsos</td>
<td>1,842</td>
<td>1,228</td>
<td>5.3</td>
</tr>
<tr>
<td>Newspoll</td>
<td>3,038</td>
<td>2,025</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations

Another way of assessing the adequacy of the sample sizes used by the polls is by estimating the Margin of Error (MoE) on the primary vote. The results, assuming a 50% population proportion and a design effect of 1.5, are shown in Table 15. Using these calculations, only the Ipsos and Newspoll samples were sufficiently large to support a confidence interval width of less than 3%, (refer to calculations in Appendix 3).

We note that none of the polls’ advertised ‘margins of error’ allowed for design effects. Invariably, published polls base their ‘margins of error’ on the total number of interviews.

Another important consideration when determining the optimal sample size for a poll is the ability to detect change over time (as also shown in Table 15 in the right-hand column). As can be seen, to detect a significant difference from one poll to the next, a larger sample size is required.

**Table 15: Estimated margins of error for a single poll and two independent polls.**

<table>
<thead>
<tr>
<th>Poll</th>
<th>Sample Size</th>
<th>Effective Sample Size (assumes design effect of 1.5)</th>
<th>MoE for 50% share of the vote (alpha=0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single Poll</td>
<td>Difference btw two independent polls of the same size</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>1,004</td>
<td>669</td>
<td>3.8</td>
</tr>
<tr>
<td>Essential</td>
<td>1,201</td>
<td>801</td>
<td>3.5</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>1,265</td>
<td>843</td>
<td>3.4</td>
</tr>
<tr>
<td>Ipsos</td>
<td>1,842</td>
<td>1,228</td>
<td>2.8</td>
</tr>
<tr>
<td>Newspoll</td>
<td>3,038</td>
<td>2,025</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations
Ultimately the size of the sample affects the variability of the result not the bias, so the size of the samples used by the pollsters in 2019 was not a reason for the discrepancy between the polls and the election outcome in 2019. Goot’s analysis of poll sample sizes from 1993 to 2019 finds that neither larger sample sizes nor poll aggregation necessarily made the polls any more ‘reliable’ (Goot, in press).

### 6.4.3. Non-response

Another important representation issue that arises in survey research, including election polling, is non-response and with it the possibility of non-response bias. Non-response occurs when not all of those approached to participate in a poll actually do so and non-response error occurs when the variables associated with non-response are correlated with the variables of interest. For example, young males (and young people generally) are much more likely to be non-responders, and if age or gender correlates with voting for a particular party, and if the under-representation of young respondents is not corrected by some means such as over-sampling or weighting, the survey estimates may be biased.

The reasons for non-responses relate to the difficulty of contacting respondents, their willingness to cooperate, and their ability to cooperate (e.g. frail aged, ill-health, language barriers, literacy and digital affinity). Non-response rates for survey research have soared in recent years as contact rates and cooperation rates have declined (see section 3.2), with typical response rates for high-standard DFRRDD CATI surveys in Australia having almost halved from around 20% in 2016 to 11% in 2019, the time between the two most-recent federal elections.

The level of effort made to obtain questionnaires/interviews from the designated sample, the way the approach is made, and the length of time the poll is in the field are all factors that can contribute to non-response and the possibility of non-response bias.

In the discussion that follows, we look at the characteristics of non-response in respect of the two main approaches used by pollsters: probability sampling and non-probability sampling.

#### Non-response bias in probability surveys

The decline in response rates for probability surveys, as well as the greater time, effort and expense required to achieve even these greatly diminished response rates, has led ‘many to question whether [probability] surveys are still providing accurate and unbiased information’ (Kohut et al., 2012). Paradoxically, the decline in response rates for probability surveys has been one of the main reasons for moving to non-probability online panels, which have even lower response rates.

However, the Pew Research Center, among others, has observed that ‘response rates are an unreliable metric of accuracy’ (Kennedy and Hartig, 2019). This is reinforced by the AAPOR Taskforce Report, *Evaluating Survey Quality in Today’s Complex Environment* (Baker et al., 2019, pp. 6-7) and by Prosser and Mellon (2018) who note that ‘declining response rates have not necessarily translated into increasing non-response bias’ (p. 761). If non-response undermines representativeness, the quality of the sample then depends on whether weighting or other means can be used to adjust for this.

---

23 There are two forms of non-response. Unit non-response occurs when a member of the designated sample does not participate in a survey. This is distinct from item non-response, which occurs when a survey respondent dies not provide a response to a particular item or items in the questionnaire. Some items, such as income, are known to have higher level of item non-response. The focus in this section is on unit non-response.
Another study by the Pew Research Center (Keeter et al., 2017), which assesses the weighted estimates from a range of contemporary telephone surveys against a series of high-quality benchmarks, found little relationship between response rates and bias when it came to measuring the political affiliation of respondents. They did find, however, that telephone polls tended to overstate civic engagement and, to a lesser extent, political engagement (discussed below). Importantly, the telephone surveys used in this study were weighted by education, gender, race/ethnicity, region, age and more [emphasis added]. In Australia, telephone surveys tend to have higher response rates than in the US (Social Research Centre, 2019) and the weighted estimates from probability-based surveys in Australia, including telephone surveys, adhere fairly well to independent population benchmarks, as they do in the Pew study (Pennay et al., 2018).

**Non-response bias (and selection error) in non-probability online panel**

Non-probability online panels are not only subject to coverage error (section 6.4.1), but also, they are more likely than probability samples to be subject to self-selection bias. This is attributable, in part, to the channels and inducement strategies these panels sometimes use; e.g. ‘Earn Cash With Quick Paid Surveys! (www.quickpaysurvey.com), ‘Make Money Online With Paid Surveys’ (www.cashcrate.com), or ‘Take Surveys for Cash’ (www.takesurveysforcash.com) (Cornesse and Blom, 2020, p. 2). While the recruitment methods used by non-probability online panels may differ in some respects, the typical methods used have led researchers to conclude that non-probability online panels suffer from non-ignorable non-response bias for many of the statistics they produce (Cornesse and Blom, 2020, pp. 6-8; Pennay et al., 2016, pp. 3-5).

Response rates for non-probability online panels cannot be determined in the same way as probability surveys because the chance of being included in a non-probability online panel is not known or, if it is known, is not reported. Based on the ‘click through’ rates for the online banner advertising that panel recruiters largely rely on, the starting point for estimating the real response rates for non-probability panels that use this method of recruiting is ‘less than 1 per cent’ (Tourangeau, Conrad and Couper, 2013, p. 42). Of the 1% who see an online advertisement to join a panel and ‘click through’, only a subset will enrol in a panel; of these, some will remain active panel participants and of these, only some will agree to participate in particular survey or poll. So, the likely true response rates for non-probability panels are likely to be a fraction of 1%. For the pollsters using non-probability panels, this matters only if the nature of their sample distorts their estimates of vote choice and if their mitigation strategies to overcome any bias don’t work.

Given their very different recruitment methods, it is no surprise that studies have identified several significant differences between members of non-probability panels and members of probability-recruited panels, the latter being much closer than the former to the profile of the general population they are trying to represent. One American study found that members of non-probability panels are more likely than members of probability samples to closely follow the news, to regard themselves as having an influence on politics, and to want to express views about the government’s effectiveness (Fahimi, 2015, p. 5). Another US study found evidence ‘that online non-probability samples … have disproportionately high shares of adults who do not have children, live alone, collect unemployment benefits and are low-income. In some respects, this squares with a stereotype one might imagine for people who find time to participate in online survey panels, perhaps akin to a part-time job’ (Kennedy et al., 2016, p. 4).

**Differential non-response in the Australian election polls**

For election polling, the under (or over) -representation of particular kinds of voters is only a problem if the characteristics of these voters are related to party choice and if this relationship cannot be – or is not – mitigated by weighting.
Over-representation of the more politically engaged

There is evidence from Britain, as well as the US, that polls over-represent more politically engaged individuals. The House of Lords Select Committee Report on Political Polling and the Digital Media notes that ‘There is probably a consensus that … any [political poll] sample that you obtain, by whatever method, contains disproportionately those who are interested in politics’ (House of Lords, 2018, p. 21; see also Sturgis, et al., p. 56; and Kennedy et al., 2016, p. 4, for the US).

One way of checking if the Australian election polls over-represented more politically engaged individuals is to compare the proportion of poll respondents that reported intending to vote in the election with the actual electoral turnout. According to the Australian Electoral Commission, the turnout for the 2019 election was 92%, with 95% of voters casting a formal (i.e. valid) vote for the House of Representatives (AES, 2019a). This equates to 87% of enrolled voters casting a formal vote.

The only data the Inquiry found that addressed this issue came from the final pre-election Essential Report. When asked how, rather than for whom, they would vote, only 4% of respondents selected from the response list that they would not vote at the election and none selected the ‘informal vote’ option (Essential Research, 2019); so, 96% said they would cast a formal vote, a much higher proportion of the electorate than actually cast a formal vote. On this evidence, respondents are more engaged in politics (i.e. more likely to vote and more likely to cast a valid vote) than the electorate as a whole. If this is true of Australian polls more generally, then Australian polls are like their US and UK counterparts in this regard (Kennedy et al., 2016, p. 4; House of Lords, 2018, p. 21; Sturgis, et al., p. 56).

Data from the Australian component of the Comparative Study of Electoral Systems survey (CSES Australia, 2019) sheds some light on this issue. This post-election survey of 2,000 Australian adults, conducted via the Life in Australia™ probability panel, asked respondents about their level of interest in politics and for which party they voted at the 2019 election. The data presented in Figure 12 show that the difference in support for Labor and the Greens on the one side, and the Coalition on the other was: 14 percentage points among those ‘very interested in politics’ (n = 501); 4 points among those ‘somewhat interested’ (n = 857); 3 points among those ‘not very interested’ (n = 330); but 25 points among those ‘not at all interested in politics’ (n = 76). These figures show no linear relationship between interest in politics and ‘left’ or ‘right’ voting; if anything, the relationship looks U-shaped.

However, taking into account the relative size of these groups and the magnitude of the differential in vote choice, the most consequential group in terms of overall bias is those ‘very interested in politics’. The relationship between interest in politics and vote choice would likely play out as an upward bias in the estimate of the Labor vote, if not confounded by other factors.

24 Since voting is compulsory in Australia, the proportion reported as ‘intending to vote’ could be affected by social desirability bias.
In the same CSES Australia survey, people with higher levels of educational attainment were over-represented – as they are in almost all surveys. A 2015 study of three probability samples and five non-probability samples, all conducted in Australia, found that on average respondents had a higher level of educational attainment than the general population. According to the 2016 Census, 20% of the adult population had a Bachelor Degree or higher. However, in the three probability samples, the proportion with a Bachelor Degree or higher was 41-44%; in the five non-probability samples, 31-42% (Pennay et al., 2018, pp. 11-12).

The CSES Australia data also show that education has a linear relationship with vote choice. Of those with tertiary education, 54% said they had voted for Labor or the Greens, and 33% for the Coalition; of those with non-tertiary qualifications, the corresponding figures were 45 and 42%; and of those with no post-school qualifications, the corresponding figures were 40% and 48% (Figure 13). If something like the oversampling of the better educated evident in the CSES Australia were also evident in the election polls, and there is no reason to think otherwise, and if this bias were not mitigated by sampling strategies and/or weighting adjustments, then the over-representation of persons with higher levels of educational attainment would, clearly, be a source of non-response bias (see section 6.4.4).

Source: 2019 CSES Australia, 2019\(^{25}\)

---

Inquiry into the performance of the opinion polls at the 2019 Australian federal election

Figure 13: Education and vote choice at the 2019 election, unweighted data (%)

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Labor/Greens</th>
<th>Coalition</th>
</tr>
</thead>
<tbody>
<tr>
<td>No qualification</td>
<td>48</td>
<td>40</td>
</tr>
<tr>
<td>Non-tertiary qualification</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>Tertiary qualification</td>
<td>33</td>
<td>54</td>
</tr>
</tbody>
</table>

Source: 2019 CSES Australia, 2019

Australian pollsters were hardly alone in failing to ascertain the educational status of their respondents – at a time when the relationship between vote choice and educational attainment was being increasingly remarked upon – let alone adjusting their data for it. In the 2016 US election, as the post-election review reports, ‘Education was strongly correlated with presidential vote in key states: that is, voters with higher education levels were more likely to vote for Clinton rather than Trump. Yet some pollsters – especially state-level pollsters – did not adjust for education in their weighting even though college graduates were over-represented in their surveys. This led to an under-estimation of support for Trump’ (AAPOR, 2017, p. 52). What is equally remarkable is that Australian pollsters appear to have taken no heed of this.

Under-representation of those who speak English poorly or not at all, those with low levels of literacy, and those with little digital affinity

In addition, the Australian polls made no effort to accommodate respondents who were unable or unwilling to complete a questionnaire, or be interviewed, in English. According to the 2016 Census, 2.4% of Australian citizens aged 18 years and over speak English either not well or not at all. As a result, they are unlikely to be represented in the polls. It is possible, of course, that they are less likely to be on the Electoral Roll – or, if enrolled, to have voted.

Others, such as those with low levels of literacy (a particular problem with self-administered polls, including online polls) or low digital affinity, will also have been excluded or under-represented. According to the 2011-12 ABS Adult Skills Survey of Australian adults, 12.6% had poor literacy (measured at Level 1 or below), 7.5% ‘indicated that they had no prior experience with computers or lacked very basic computer skills’, and 38.1% ‘scored at or below Level 1 (the lowest level) in problem solving in technology-rich environments’ (OECD, 2013, p. 3).

Census TableBuilder, 2016 Census, ENGP Proficiency in Spoken English by CITP Australian Citizenship filtered to age group 18 years and over
If any of these characteristics are associated with vote choice and remained unmitigated, as they almost certainly were, this lack of mitigation represents a potential additional source of bias.

6.4.4. Weighting

Survey data are weighted (statistically adjusted) to mitigate the effects of coverage error, account for any disproportionality in the sample design, adjust for non-response and, finally, to balance (calibrate) the sample to key population parameters known to be correlated with the measures of interest – voting intentions, in the case of election polls. The ultimate goal of weighting is to reduce bias in the reported survey estimates but when weighting is used to correct for poor sampling it will generally come at the cost of increased variance.28

A robust weighting solution usually consists of two or three of the following stages, subject to the availability of auxiliary data about non-respondents:

1. Design weights are calculated to account for different selection probabilities; for example, stratified sample designs where the chance of selection into the sample is not proportional to the population; or where the chance of selection is dependent upon a person’s characteristics, such as in a DFRDD survey (refer to section 4.1) where the probability of selection will depend on whether the respondent is contactable via a landline telephone only, a mobile telephone only, or both.

2. Where auxiliary information is available about every selected person, a separate non-response adjustment (response propensity weight) is made to take account of differences between responding and non-responding persons. When no auxiliary information is available, this step is assumed to be implicit in Steps 1 and 3.

3. Design weights are adjusted (calibrated) so that they align with known distributions for key socio-demographic characteristics. These parameters are generally derived from official statistics published by the Australian Bureau of Statistics (ABS). These adjustments can also be used for coverage error provided the information from the sample can be extrapolated to those who are excluded from the sampling frame. However, accuracy depends on certain assumptions. For example, in a phone survey, we assume that people without a phone are similar in their opinions or behaviour to people with a phone who have a similar socio-demographic profile.

The first two stages in the multistage weighting process outlined above are rarely used in polling due to a lack of information about the probabilities of selection into the sample, the absence of an informative reference base needed for calculating design weights (or quasi-design weights in the case of non-probability panels), and little or no information about non-response and coverage error.

From the information available to this Inquiry, it seems that the pollsters calibrated their data to a very minimal set of population parameters. It is our understanding that three of the five polls were adjusted so that their weighted samples matched the distribution of the Australian adult population for age, sex

---

28 Weighting complex samples with a view to reducing bias usually introduces a sample design effect (deff), the result of which is to reduce the effective sample size and increase the variability of the resultant estimates (i.e. widen the confidence intervals). As discussed in section 5.4, some weighting schemes can reduce sampling variability and therefore increase the effective sample size. However, based on our (limited) knowledge of the weighting methods used by the pollsters, in our judgement it is safe to assume that the effective sample sizes are less than the actual sample sizes reported by the polls.
and geography; one pollster, Ipsos, added household income to their weighting solution. Where any of these parameters were included in the specifying of quotas, balancing the samples to these parameters would have been of minimal additional value.

The effectiveness of these weighting approaches

So, how effective were these weighting approaches? Without access to the raw data, we cannot directly answer this question. The best we can do is to try to replicate the weighting techniques used by pollsters on other data sets containing the variables of interest to see how effective these techniques are in overcoming observed biases in these other data sets. This will provide some indirect evidence as to their possible effectiveness when used in election polls.

To consider whether parameters other than sex, age and geography should be included in weighting schemes, it is necessary to have high-quality population benchmarks for any potentially new weighting variables and to select the combination of parameters (and potential interactions between them) that are most strongly associated with the outcome of interest (i.e. voting intentions). The questionnaires used for the election polls then need to include questions to measure these potential new weighting variables, with these questions harmonised with high-quality population benchmarks. For example, in order to use ABS educational attainment benchmarks, poll measurement of education qualifications needs to be consistent with how education qualifications are measured by the Census. For some concepts, this can be achieved by aligning the questions, but for others, differences in mode, survey context and other factors may mean that population benchmarks are not applicable to the data collected by a poll and therefore these measures are not good candidates as far as weighting is concerned.

Deciding what additional variables might warrant inclusion requires empirical and analytical investigations that might include benchmarking studies, modelling of associations and mode effects. These investigations require substantially more time and effort than is available to the Inquiry and are rightly the domain of the pollsters and survey statisticians. However, to illustrate the potential for error reduction through adjustments to weighting approaches, we have broadly replicated the weighting techniques used by the pollsters on two post-election surveys – the CSES Australia and the AES – both of which include a retrospective vote choice question. This has enabled us to undertake three case studies, two looking at the impact of adding educational attainment to a typical weighting solution and a third looking at adding other variables / combinations of variables in addition to education. Given that voting behaviour in these surveys was measured after the election, knowledge of election outcome could have affected how this question was answered.

The details of the weighting approaches used for these case studies are described in their respective technical reports, available from the Australian Data Archive. For the purposes of this exercise, we retained the base weights for each record and recalculated the final weights using raking. These

---

29 YouGov Galaxy described their weighting methodology for a pre-election poll in Queensland in August 2019 as comprising the following … (1) age interlocked with gender and region (2) education (3) previous election voting intention. Weighting targets are YouGov estimates derived for voters from ABS statistics’ [https://d25d2506sfb94.cloudfront.net/r74/Courier%20Mail%20YouGov%20Poll%20Results%20August%202019%5b1%5d.pdf?utm_medium=Social&utm_source=Twitter&utm_campaign=APAC_2019_9_AUCourierMailResults_Ratings](https://d25d2506sfb94.cloudfront.net/r74/Courier%20Mail%20YouGov%20Poll%20Results%20August%202019%5b1%5d.pdf?utm_medium=Social&utm_source=Twitter&utm_campaign=APAC_2019_9_AUCourierMailResults_Ratings). It is not known if this was the method adopted for their national pre-election polls.


31 Raking (also known as ‘rim weighting’ or ‘iterative proportional fitting’) is a weighting method commonly used in the polling and market research industry. It involves a process of iteration to ensure that weights simultaneously match the population proportions of several demographic variables; see Lumley 2004, 2014.
case studies illustrate the impact of adding educational attainment to the standard weighting schemes used by pollsters, as not weighting by education was seen as a consequential factor in the performance of the 2016 US election polls (AAPOR, 2017, p. 52). This is not to suggest that educational attainment on its own is sufficient or the best or the only parameter that should be considered in this context.

To provide comparability with the weighting parameters used by pollsters, our base weighting includes three variables: age (18-24, 25-64 in 10-year age groups and 65 plus), gender (male/female) and geography (capital city/rest of state). These may not correspond exactly to the weighting parameters used by the pollsters but are likely to be sufficiently similar to allow comparison. The case studies and the weighting benchmarks used are written up in Appendix 6.

The two case studies that examine the impact of adding educational attainment to the weighting solution produce contrasting results. In Case Study 1 (based on the CSES Australia post-election survey), the bias reduction achieved by weighting by education was considerable (ranging from a 21-33% reduction in bias depending on the metric used) whereas for Case Study 2 (using the AES post-election survey), adding educational attainment to the weighting solution had only a modest impact on bias reduction.

Would including other weighting variables further improve the accuracy of voting behaviour estimates? To answer this question we undertook a third case study, again using CSES Australia data. Other socio-demographic and socio-economic characteristics, not currently collected by the pollsters, were available to the CSES Australia. These included: general health (ABS National Health Survey), life satisfaction (ABS General Social Survey, 2014), country of birth (Census 2016), employment status (Census 2016), and concession card holders (ABS National Health Survey, 2015. What we found from this case study (refer to Appendix 6) is that adding ‘self-assessed health’ and/or ‘life satisfaction’ as weighting variables to standard weighting schemes has the potential to achieve a non-trivial bias reduction across a range of outcome metrics and, as such, these variables may warrant further investigation by pollsters.

Other possible weighting variables

Weighting by past vote

Correcting voting intention estimates by balancing or weighting one’s sample by ‘previous vote’ is frequent in electoral polls, particularly in Europe (Durand et al., 2015, p. 1). YouGov’s UK Director of Political and Social Research, Anthony Wells, makes a similar observation with regard to pollsters in the UK: ‘Almost all … use how people voted at the last election as a target when designing or weighting … polling samples’ (Wells, 2019). In Australia, however, it is unclear whether any of the pollsters used this measure in 2019.

The prospect of balancing or weighting an election poll sample so that it is representative of the population based on the voting distribution at the last election is, at face value, very appealing. However, recall of past vote may be inaccurate for three reasons: (1) memory failure, especially since ‘voting is not a salient, memorable behaviour among all voters’; (2) the tendency of voters to misremember (or misreport) a previous vote in order to reconcile it with how they currently wish to vote; and (3) social desirability (Durand et al., 2015, pp. 3-12). In addition, some respondents will not have voted in the previous election.

The implications of weighting based on an inaccurate past vote measure can be substantial. Wells conducted an experiment in which he re-weighted YouGov polling data for the 2017 Brexit election using past vote data collected immediately after that election (at which time 41% reported voting for Labour) and past vote data collected from the same respondents in 2019, at which time only 33% reported voting for Labour. The difference these two readings made to the estimate of current voting
intentions was substantial. When the 2017 measure of ‘past vote’ was used, estimated support for Labour was 21%; when the 2019 measure of ‘past vote’ was used, estimated support for Labour increased to 24%. Nonetheless, the substantial correlation between vote choice at successive elections mark ‘past vote’ as an important variable for further consideration if more advanced weighting solutions are to be adopted. Our analysis of 2019 AES data shows that approximately 77% of those respondents who reported voting for the ALP, Liberal Party, National Party or Greens in 2016 reported voting the same way in 2019.

Political ideology, political interest and internet usage

A US study (Kennedy et al., 2016) compared a variety of estimates generated from the first iteration of their probability-based American Trends Panel (ATP) with estimates generated from several non-probability opt-in panels against high-quality population benchmarks. One of the non-probability panels (Panel I, subsequently identified as YouGov) ‘consistently outperformed the others including the probability-based ATP, ranking first in terms of bias reduction across a wide-range of variables for which there were high-quality benchmarks.\(^{32}\) This top-performing sample was notable in that it employed a relatively elaborate set of adjustments at both the sample selection and weighting stages. The adjustments involved conditioning on several variables that researchers often study as survey outcomes, such as political ideology, political interest and internet usage.’ The authors concluded that much of this panel’s success ‘stemmed from the fact that it was designed (before and/or during fielding) to align with the population benchmarks on this broader array of dimensions (Kennedy, et al., 2016, p. 1).

Concluding remarks on weighting

While we are not able to directly extrapolate the benefits of using additional variables from the CSES Australia or the AES to the election polls, it is possible that the accuracy of the polls would be improved by the addition of one or two other variables to the three – sex, geography, and age – that all pollsters appeared to use in 2019. The two primary candidates that warrant further exploration are ‘educational attainment’ and ‘health status’. If we had to recommend just one of these two – and adding any would add to the cost of conducting the polls – we would add education. It was the failure to weight by education at the 2016 presidential election that the AAPOR review identified as crucial to the poll misses in several states.

‘When it comes to accuracy, choosing the right variables for weighting is more important than choosing the right statistical method.’ (Mercer et al., (2018, p. 4).

While the Inquiry Panel recommends that pollsters review their current weighting approaches, it is important to point out that, clearly, there is no silver bullet when it comes to weighting election polls. As noted in the UK Polling Review, ‘The search for the ‘missing’ [weighting] variable is likely … to prove elusive …; if a variable possessing the required characteristics existed the pollsters would likely already be using it. Additionally, it is unlikely that bias in the vote intention estimate can be accounted for by a single direct cause. It is more plausible that the underlying causal model linking survey participation to vote choice is a complex multi-variable system, comprising interacting direct and indirect effects’ (Sturgis et al., 2016, p. 53).

What is the best that can be hoped for?

For the non-probability samples, a marker has been laid down as to the sort of bias reduction that might be achievable given the use of emerging weighting solutions. The Pew Research Center study by Mercer and colleagues into weighting online samples found that ‘even the most effective

---

\(^{32}\) Civic engagement, family, financial, personal, political engagement and technology
adjustment strategy was only able to remove about 30% of the original bias (Mercer et al., 2018, p. 3). Using data from the Online Panels Benchmarking Study (Pennay et al., 2018), similar reductions in bias have been achieved using Australian survey data (Neiger, 2020). But it is important to remember that weighting can only do so much and cannot be depended upon to fully correct for poor sampling; accordingly, weighting should be viewed as an adjunct to, not a substitute for, good sampling.

6.5. The convergence of the polls

The most striking aspect of the final polls (indeed, of all the polls during the 2019 campaign) was how little difference there was in their two-party-preferred figures. This occurred despite their considerable differences in terms of mode of data collection, sampling frames, fieldwork dates, sample sizes and the ways in which the final two-party-preferred vote was estimated. Of the five final polls (as shown in Table 3), one poll (Roy Morgan) estimated the Coalition’s two-party-preferred at 48%; two (Essential and Newspoll) estimated it at 48.5%; and two (Ipsos and YouGov Galaxy) estimated it at 49%. Yet only two polls out of the five (Essential and YouGov Galaxy) used the same survey mode; only two polls (Ipsos and YouGov Galaxy) were conducted on the same days; and only three polls (Essential, Roy Morgan, and YouGov Galaxy) used roughly the same size samples. The spread was wider in terms of the first preference vote for Labor, which ranged from 33% (Ipsos) to 37% (Newspoll and YouGov Galaxy).

The similarity of the estimates for the two-party-preferred vote was not the result of a late convergence. Over the previous four weeks, 16 separate estimates of the Coalition’s share of the two-party-preferred had varied hardly at all: a low of 48%, a high of 49%, whether measured by Newspoll (five polls), Roy Morgan (four), Essential (three), Ipsos (two) or YouGov Galaxy (two).

The odds of a sequence of 16 almost identical polls of this kind occurring by chance are extremely low. In a widely quoted article, the ANU’s Vice Chancellor and Nobel Laureate in astrophysics, Brian Schmidt, estimated the odds as ‘greater than 100,000 to 1’ (Schmidt, 2019). The Inquiry Panel’s own calculations, using the standard error estimates provided by the pollsters, find that the chances of the final five polls being all this similar and wrong are something in the order of 1 in 1.5 million. The odds of 16 consecutive polls being this similar and wrong is many magnitudes greater (refer Appendix 7). 33

The ‘wrongness’ and similarity of the poll results are distinct issues. Had the majority of the final polls been right; that is, if Labor (not the Coalition) had won the election with approximately 51.5% of the two-party-preferred vote, it is unlikely that this Inquiry would have been instituted. A casual observer might then think there was nothing to be investigated. However – the polls could have been right but their similarity would still have signalled that something unusual was going on. As noted above, the similarity between the polls is striking.

There are three possible explanations for all five polls missing and by similar amounts and in the same direction: (1) herding; (2) systematic bias; or (3) some combination of (1) and (2). It is unlikely to be by chance.

6.5.1. Herding

The American Association for Public Opinion Research (AAPOR) defines herding as a range of behaviours ‘from making statistical adjustments to ensure that the released results appear similar to existing polls to deciding whether or not to release the poll depending on how the results compare to

33 The calculations of the odds by Schmidt and the Inquiry Panel and discussion of variability by Mansillo and Jackman (to follow) are provided as a context for framing the discussion and not intended to be treated as the actual odds, as none of these calculations take into account complex sample design, weighting adjustments, non-random elements of the polls such as self-selection bias, coverage error, non-response error and other survey errors discussed elsewhere in the report.
existing polls.’ (AAPOR, n.d.). In addition to being unethical, a ‘troublesome potential consequence of “herding” is that … [it] produce[s] artificially consistent results … that may not accurately reflect public attitudes’ (AAPOR, n.d.).

Herding may occur when a pollster lacks confidence in their findings: (a) because they are at odds with findings they have published previously, and they can think of nothing that has affected the population that is likely to account for the change; (b) because they are at odds with the findings of other polls – or the average of other polls, where there is a spread; or (c) because they are at odds with the results reported by the most prestigious of the polls – the market setter, as it were.

Herding may also occur because, whatever confidence they have in their own poll, the boost to a pollster’s reputation that might follow from being the only one that gets it right would not be as great as the damage to their reputation by being the only one that gets it wrong. For pollsters, ‘it is safer to be in the herd than alone in the field’ (Mansillo and Jackman 2020, p. 143). This is where commercial considerations loom large.

The effect of herding is to make the range of poll results look more robust (Sturgis et al., 2016, p. 10). When the polls herd towards the correct result, ex post, the polls win the critics’ praise. When they herd towards the wrong result, they leave their readers disappointed – or, worse, in a state of ‘collective shock’ (Sturgis et al., 2016, p. 62). One consequence of herding towards the correct set of figures is to boost the reputation of the industry.

Analyses of the polls in the United States show evidence of herding (Linzer, 2012; Clinton and Rogers, 2013; Silver, 2014). The British inquiry into the 2015 polls also produced evidence of herding (Sturgis et al., 2016, pp. 62-8).

A comprehensive analysis of the Australian polling conducted during the lead-up to each of the last five federal elections, from 2007 to 2019, shows that in 2019 the polls’ estimates of the two-party-preferred vote bore ‘the signature of herding’ (Mansillo and Jackman, 2020, p. 146). They were ‘under-dispersed’; that is, they showed less variability than one would expect under the Central Limit Theorem for a series of unbiased random samples following a normal distribution (Mansillo and Jackman, 2020, p. 143) – not just over the last four weeks of the campaign but for the three months leading up to the election, and perhaps beyond. This compares with the 2007 and 2010 elections when they were ‘over-dispersed’, and the 2013 and 2016 elections when their estimates mostly fell within the upper and lower bands of the expected distribution (Mansillo and Jackman 2020, pp. 144-45).

One possible explanation for the ‘clustering of poll results’ in 2019 was ‘the shared consensus that Labor would win the election’ (Mansillo and Jackman, 2020, p. 146). Another could be the widely remarked feature of the Australian polling industry – the existence of a market setter, Newspoll. This is not something observed in America or Britain. In contributions to this Inquiry, some pollsters conceded that, for their clients, the Newspoll results loomed large.

The general expectation of a Labor win – an expectation the polls helped generate – meant that pollsters may not have looked at their methods as closely as they might have. This is an example of ‘confirmation bias’, the tendency to uncritically accept findings that fit with prior expectations.

6.5.2. Herding by suppression

Herding may not only involve the deliberate adjustment of poll results, or the processes that affect those results, in order to produce a result that is similar to others. A decision about whether to release a poll can also constitute herding when it is based on whether its findings fit with other recently released polls. A pollster is herding when a decision is made not to publish because the poll seems out of step.
The only admitted instance of this occurring during the 2019 election is that of Chris Lonergan of Lonergan Research. Lonergan claimed that he was involved in a decision not to publish a poll that showed Labor doing poorly in Queensland and is quoted as saying: ‘No one wants to release a poll that is wildly out of step ... we didn't want to be seen as having an inaccurate poll’ (Koziol, 2019).

According to a report in The Australian (15 July 2019), Lonergan’s decision to ditch the poll was made in conjunction with the client who commissioned it; the client was “not a political party” and “not GetUp”. When interviewed by this Inquiry, Lonergan said that his client has declined to allow these data to be shared with this Inquiry even on a confidential basis.

6.5.3. Was herding a factor in the performance of the polls in 2019?

Unfortunately, since all of the polling organisations declined our request to access their raw data and weighting documentation, we cannot be sure. We are left in the dark as to how they weighted their data and whether any ad hoc decisions were made in producing their final estimates.

The disclosure standards recommended by this Inquiry (see Appendix 10) provide ways by which cases of herding could be detected and deterred. Recommendation 15 requires pollsters to ‘divulge whether their standard weighting methodology was adjusted in any way at all for the current poll.’

---

34 This lack of access to raw data and weighting documentation sets this Inquiry apart from similar inquiries in the US and Britain.
7. Setting standards for election polling

This section draws heavily on the Discussion Paper: Disclosure standards for election and political polling in Australia released by the Inquiry Panel in May 2020 (see https://www.amsro.com.au/amsro-polling-inquiry). While many of the findings and recommendations contained in the Discussion Paper are unchanged, others have been added, deleted or modified to take account of reactions to the Discussion Paper and further considerations by the Inquiry Panel.

This section and Appendix 8 set out the existing codes of conduct/guidelines for election polling in Australia, provide an overview of the international election polling guidelines and examples of the codes that operate in the USA, UK, Canada and New Zealand, and consider the implications of these overseas models for the development of election polling standards in Australia. Recommendations and issues for consideration by AMSRO and other stakeholders are provided in section 10 and Appendix 10.

Standards for Australian market and social research practitioners, including pollsters, are provided by two organisations: The Research Society; and the Association of Market and Social Research Organisations (AMSRO). There are no mandated standards set by governments. A third organisation, the Australian Press Council (APC), provides guidelines to journalists and editors for the reporting of opinion polls with the aim of ensuring that enough information is provided ‘to prevent polling results from being misconstrued’ (Australian Press Council, 2018).

Pollsters who are members of The Research Society are bound to uphold The Research Society’s Code of Professional Behaviour (‘the Code’) that sets out the ethical requirements and standard conditions for conducting and reporting market and social research. Rules 32 to 38 of the Code cover the obligations on members when providing data and reporting findings:

32. Members must ensure that an individual’s identity cannot be inferred via deductive disclosure (for example, through cross-analysis, small samples or combination with other data such as a client’s records or secondary data in the public domain).

33. Members must ensure that findings and any interpretation of them are clearly and adequately supported by the data.

34. When reporting on a project, Members must make a clear distinction between the findings, the Member’s interpretation of those findings and any conclusions drawn or recommendations made.

35. Members must provide their clients with appropriate methodological details to enable them to assess the validity of the results and any conclusions drawn.

36. Members must take reasonable steps to ensure that findings from a project, published by themselves or in their company name, are not incorrectly or misleadingly presented.

37. Members must take reasonable steps to check and where necessary amend any client prepared materials prior to publication to ensure that the published findings will not be incorrectly or misleadingly reported.

38. Members must take reasonable steps to ensure that their name and/or company name are not associated with the dissemination of findings from a project unless they are adequately supported by the data (Research Society, 2020a, p. 8).

These broad provisions of the Code are expanded upon in a separately issued Guideline on Reporting Research Findings (2020) that was made available to this Inquiry but is otherwise only available to members of The Research Society. The Guideline details the professional standards for disclosure for published results. The general disclosure requirements covered in the Guideline are:

- project background, including the declarations of any interest or any conflicts
• a detailed description of data collection methods used including issues such as the method of recruitment, payment of incentives, fieldwork dates, the languages in which interviews were conducted and other and indications of fieldwork effort
• the sample source, sample size and sample coverage and a detailed description of the sampling method
• a statement of response rates
• a statement on the probable statistical margins of error and statistical differences between key figures
• a copy of the questionnaire
• a statement of any significant limitations, and
• an obligation that if members are aware that their research findings are likely to be published, they make their client aware of their obligations under the Code and this guideline.

The Guidelines summarises a Member’s obligations when reporting research findings as ensuring ‘the results of a project are reported accurately, transparently and objectively. Findings must be adequately supported by the data and methodological details must be provided to allow readers to assess the validity of the results. This is of particular importance if results are published’ (The Research Society, 2020b, emphasis added).

Based on information provided by The Research Society, as of July 2020 the Society has over 2,000 individual members and 96 Company and Client Partners. All four of the polling organisations that released national election polls in 2019 (Essential, Ipsos, Roy Morgan and YouGov) have employees who are members of The Research Society, but only Ipsos and Roy Morgan are Company Partners. The Research Society also has a professional accreditation scheme enabling individual members to gain accreditation as a Qualified Professional Researcher (QPR) (The Research Society, 2020c). Ipsos, Roy Morgan and YouGov have employees with QPR accreditation but, based on a viewing of the QPR list on 20 March 2020, not all of the pollsters publicly associated with the polling firms that released national election polls in 2019 have industry accreditation as Qualified Professional Researchers.

The Statistical Society of Australia (SSA) also has a professional accreditation scheme enabling members to gain accreditation as an Accredited Statistician (AStat) or a Chartered Statistician (CStat). None of the individuals publicly associated with the polling firms that carried out publicly released national election polls in 2019 have AStat or CStat accreditation.

Employers in the market and social research industry who are members of AMSRO also agree to adhere to specific disclosure standards. The top tier of AMSRO membership, ‘AMSRO Trust Mark’ members, must adhere to the Privacy Code (Market & Social Research), and The Research Society’s Code of Professional Behaviour, and hold the International Standard for Market, Opinion and Social Research certification (ISO 20252) (AMSRO, n.d). Of the four companies that publicly released national election polling results ahead of the 2019 election, only Ipsos was an AMSRO Trust Mark member; YouGov only joined AMSRO in July 2020. Along with Ipsos (and YouGov), but outside of the AMSRO structure, Roy Morgan Research is ISO 20252 accredited.

While the ISO standards are not designed specifically for election polling, some are broadly applicable:

35 AMSRO claims to represent 75% of the industry by revenue’ (AMSRO, 2020b).
• Disclose the name of client and research provider.
• Disclose and identify any subcontracted services.
• Report, where applicable: the target population; methods of data collection; fieldwork dates; incentive types; methods of statistical analysis; and margin of sampling error; and provide a statement of substantial limitations affecting the validity of the findings.

For published results, there is the additional ISO 20252 requirement to ensure that the conclusions are supported adequately by the data, and that the results and the interpretation of the results are clearly distinguishable.

The Australian Press Council (APC) guidelines for the benefit of journalists and editors are as follows:
‘Editors should take reasonable steps to ensure that reports about previously unpublished opinion poll results include, or have been written taking into account, at least the following matters:
• the name of the organisation that carried out the poll;
• the identity of any sponsor or funder;
• the exact wording of the questions asked;
• a definition of the population from which the sample was drawn;
• the sample size and method of sampling; and
• the dates interviewing was carried out.

Publishers are also encouraged to consider including the following matters where possible:
• how the interviews were carried out (in person, by telephone, by mail, online, etc.); and
• the margin of error.’

This Inquiry is not privy to the extent to which the APC polling guidelines are monitored or enforced. The APC received 554 complaints in 2017-18 (Australian Press Council Annual Report, 2019, p. 20) but did not provide a breakdown of complaints relating specifically to public opinion polls. This is something the APC could consider.

The view of the Inquiry Panel is that the current disclosure standards as they apply to the publicly released election polls in Australia are in need of consolidation, clarification and greater specificity. This would best be achieved by putting in place a new code and new governance structure. The existing codes, including The Research Society Guideline on Reporting Research Findings, and the revised disclosure recommendations of the Inquiry Panel, provided in Appendix 10, provide a good starting point.

In addition to their specific disclosure standards, all the codes examined place an overarching obligation on members along the following lines: ‘researchers/pollsters must not make claims that exceed the limits of the scientific principles on which opinion polls are based and be sure that interpretations and statements are fully consistent with the data’ (ISO 20252, 2019, p. 23). The extent to which Australian election pollsters and journalists achieve this overarching obligation is among the issues explored in the next section – Reporting the polls.
8. Reporting the polls

8.1. The polls are big news

Election polls are big news and the reporting of election polls matters for all the reasons canvassed in section 2. A search of all Australian newspaper outlets (metropolitan and rural/regional) over the period of the campaign, 11 April to 18 May 2019, found that the 16 national election polls released during the campaign generated 613 newspaper articles.\(^{36}\) A separate analysis by Carson and Zion of the 377 Monday to Saturday edition front pages produced by the 12 major daily newspapers throughout the campaign showed that 69% of the front pages (n=258) contained a story about the election and that, of these, 25% were about the polls (Carson and Zion, 2020, pp. 433-438). In this context, the reporting of the polls is important with potentially far-reaching consequences.

8.2. Disclosure standards and the reporting of the polls

When it comes to the reporting of the polls, three distinct levels can be distinguished:

1. **Original sources**: pollster’s website / press release / report.
2. **Primary media**: the place where the poll gets its first and most comprehensive airing and that has a direct relationship or arrangement with the pollster.
3. **Secondary media**: coverage of the poll that is based largely on reports in primary media and original sources.

In setting out for editors and journalists the features of the polls that should be disclosed to readers, the APC guidelines note that ‘space considerations may restrict the amount of background information that can be provided about a poll, but background information on at least a number of important details is desirable. This can be placed in the main body of an article, a footnote, another section that may be read separately or, if online, via a hyperlink’ (Australian Press Council, 2018).

To what extent did the reporting of the polls during the 2019 campaign meet the APC disclosure requirements? To determine this, we examined a selection of the reports (original or primary) that accompanied the release of each poll. If a required poll descriptor (according to the APC guidelines) could be found in the original or primary report, or indirectly via a link to further information, such as to the pollster’s own report, this was regarded as meeting the standard. The 16 original or primary reports used for this analysis were drawn from our search of the Factiva database, as they occurred (i.e. a convenience sampling approach was used). It should be noted that while a single poll sometimes had more than one primary report associated with it, because some polls had contractual relationships with companies (News and Nine) that owned several outlets, we also found that identical reports were shared across company mastheads. Given these limitations, the analysis that follows should be regarded as indicative only. Links to the 16 source articles are provided at Appendix 9.

The minimum disclosure standards coded for each poll were: name of pollster; who (if anyone) commissioned the poll; the fieldwork period; sample size; mode of data collection; references to a margin of error; and whether the question used to measure voting intentions was provided. We did not assess whether the reports included a separate ‘definition of the population from which the sample

---

\(^{36}\) The Factiva database was used for this search. The search terms used for this analysis attempted to capture any reference to the six polling ‘brands’ responsible for publicly releasing national polls during the campaign: Essential, Galaxy, Newspoll, Ipsos, Roy Morgan and YouGov (separate from its Galaxy brand). The search terms specified the named polling company plus ‘poll’, to minimise irrelevant search results. Manual cleaning further removed blog posts (such as The Guardian’s ‘As it Happened’ webpage) and news summaries without detailed reporting.
was drawn’ due to our uncertainty as to what this clause actually means. The results of this analysis are summarised in Table 16.

All 16 original or primary reports fulfilled the basic requirements of mentioning the name of the pollster and the funder (assumed though not explicitly mentioned in the Essential and Roy Morgan polls). For the most part (15/16), they mentioned the sample size, the fieldwork period (14/16), and provided the voting intention question (13/16) or at least some approximation of it. Nine of the 16 reports mentioned the mode of data collection. A minority (7/16) noted a ‘margin of error’.

Table 16: Primary reports that met basic disclosure requirements

<table>
<thead>
<tr>
<th>Disclosure</th>
<th>APC Guideline</th>
<th>Reports that contained this element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pollster identified</td>
<td>Y</td>
<td>16/16</td>
</tr>
<tr>
<td>Funder identified</td>
<td>Y</td>
<td>16/16</td>
</tr>
<tr>
<td>Sample size</td>
<td>Y</td>
<td>15/16</td>
</tr>
<tr>
<td>Fieldwork period</td>
<td>Y</td>
<td>14/16</td>
</tr>
<tr>
<td>Voting intentions question provided</td>
<td>Y</td>
<td>13/16</td>
</tr>
<tr>
<td>Mode/method of data collection</td>
<td>Y</td>
<td>9/16</td>
</tr>
<tr>
<td>Margin of error</td>
<td>Y</td>
<td>7/16</td>
</tr>
</tbody>
</table>

Source: Authors’ analyses of the 16 original sources/primary media (See Appendix 9)

A breakdown by polling organisation is provided in Table 17: . All three of The Guardian’s primary reports of the Essential Report met the specified APC requirements. This is in part because each of these articles contained a link to the Essential Research website, which included a basic methodological description of each poll. The reporting of the Ipsos, Newspoll and Roy Morgan polls consistently met five of the seven APC disclosure guidelines. The two primary reports of the YouGov Galaxy polls in the News Corp mastheads (see Appendix 9 for details) both identified the pollster and the poll funder and one or the other also mentioned either the sample size, the fieldwork dates or the wording of the voting intentions question.

Based on this (limited) analysis, the reporting of the Australian election polls does not consistently meet the fairly minimal APC disclosure standards. This is an issue the Inquiry Panel believes needs to be addressed.
Table 17: The number of ‘primary reports’ that met basic disclosure requirements by pollster and media/reporting outlet.

<table>
<thead>
<tr>
<th>Disclosure</th>
<th>APC Guideline</th>
<th>Essential The Guardian</th>
<th>Ipsos Nine Media</th>
<th>Newspoll The Australian</th>
<th>Roy Morgan</th>
<th>YouGov Galaxy NewsCorp Metro mastheads</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of primary reports examined</td>
<td>7/7</td>
<td>5/7</td>
<td>5/7</td>
<td>5/7</td>
<td>2/7</td>
<td></td>
</tr>
<tr>
<td>Pollster identified</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Funder identified</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Sample size</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Fieldwork period</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Voting intentions question provided</td>
<td>Y</td>
<td>3</td>
<td>-</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Mode/method of data collection</td>
<td>Y</td>
<td>3</td>
<td>1</td>
<td>-</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Margin of error</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis of the 16 original reports / primary media (See Appendix 9)
9. **Our expectations of election polling**

9.1. **The perils of prediction**

One of the difficulties faced by Australian election pollsters, journalists and commentators when trying to forecast an election outcome from poll findings is that election outcomes are not decided by who wins the national vote but who wins the majority of seats or, more accurately, which party is able to form a majority in the House of Representatives.

Table 18 shows the election results from 1993 according to the party (or parties) that won the two-party-preferred vote, their vote share and the proportion of seats this resulted in, and who formed government. This demonstrates the uneven relationship between the national two-party-preferred vote and seats won. In 1993, 51.4% of the two-party vote won Labor 80 seats in the 147-seat House of Representatives whereas 51% in 1998 (just 0.4% less) won Labor just 67 of a possible 148 seats. The Coalition had a two-party-preferred vote of 53.6% in 1996, delivering 94 seats, whereas an almost identical popular vote in 2013 (53.7%) resulted in 90 seats. In 2004, 52.7% of the vote delivered 87 seats to the Coalition whereas the same vote for Labor in 2007 delivered 4 seats fewer – 83. In 1998, the Coalition won 87 seats and government with 49% of the two-party-preferred vote; in 2010 Labor could only form a minority government when 50.1% of the two-party vote produced just 72 seats.

The uneven distribution between the popular vote and the resultant level of parliamentary representation is not a uniquely Australian phenomenon; in the 2016 US presidential election, Clinton won the popular vote by 2.1% but Trump won the electoral college votes, 306 to 232 (AAPOR, 2017).

**Table 18: Two-party-preferred vote by seats won and election winner, 1993 – 2019.**

<table>
<thead>
<tr>
<th>Year</th>
<th>Party with the highest 2PP vote share</th>
<th>2PP Vote %</th>
<th>Seats n/N</th>
<th>Seats %</th>
<th>Election winner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>ALP</td>
<td>51.4</td>
<td>80/147</td>
<td>54.4</td>
<td>ALP</td>
</tr>
<tr>
<td>1996</td>
<td>LNP</td>
<td>53.6</td>
<td>94/148</td>
<td>63.5</td>
<td>LNP</td>
</tr>
<tr>
<td>1998*</td>
<td>ALP</td>
<td>51.0</td>
<td>67/148</td>
<td>45.2</td>
<td>LNP</td>
</tr>
<tr>
<td>2001</td>
<td>LNP</td>
<td>50.9</td>
<td>82/150</td>
<td>54.7</td>
<td>LNP</td>
</tr>
<tr>
<td>2004</td>
<td>LNP</td>
<td>52.7</td>
<td>87/150</td>
<td>58.0</td>
<td>LNP</td>
</tr>
<tr>
<td>2007</td>
<td>ALP</td>
<td>52.7</td>
<td>83/150</td>
<td>55.3</td>
<td>ALP</td>
</tr>
<tr>
<td>2010**</td>
<td>ALP</td>
<td>50.1</td>
<td>72/150</td>
<td>48.0</td>
<td>ALP</td>
</tr>
<tr>
<td>2013</td>
<td>LNP</td>
<td>53.5</td>
<td>90/150</td>
<td>60.0</td>
<td>LNP</td>
</tr>
<tr>
<td>2016</td>
<td>LNP</td>
<td>50.4</td>
<td>76/150</td>
<td>50.7</td>
<td>LNP</td>
</tr>
<tr>
<td>2019</td>
<td>LNP</td>
<td>51.5</td>
<td>77/150</td>
<td>51.3</td>
<td>LNP</td>
</tr>
</tbody>
</table>

Source: Australian Electoral Commission
* Coalition formed government with 49% of the vote and 80 seats
**Labor formed a minority government

There are divergent views as to whether it is appropriate to use the polls, especially the final polls of an election campaign, to predict election outcomes. Courtney Kennedy, the lead author of the AAPOR
Inquiry into the 2016 US polls, was quoted as saying, ‘There was actually quite a bit of diversity of opinion on the committee on [this] issue: Some leaned toward being more aggressive in emphasizing the distinction between the predictors and the pollsters; others less so’ (DeSilver, 2016). This same dilemma was noted in the House of Lords Inquiry into UK polling: ‘several witnesses told us [the Inquiry] that polls are not necessarily intended to be predictors of election outcomes. Instead, they represent a “snapshot” in time, and public opinion can and does shift between the date of a poll and the election’ (House of Lords, 2018, p. 16). Our Inquiry Panel could not reach a consensus on this issue and, clearly, we have not been the first to grapple with this dilemma. However, by highlighting the total survey error properties of the polls, we hope to have demonstrated the need for caution in making predictions based on the polls, an enterprise that Sol Lebovic (founding Managing Director of Newspoll, 1985 – 2005) was known to describe as a “Mug’s game”.

9.2. Developing more reasonable expectations of the polls

Statisticians, Barnett and Sisson, point out that to deliver the level of accuracy needed to call a close election, ‘we don’t just need good opinion polls we need excellent opinion polls, because even a small error could mean predicting the wrong winner’ (Barnett and Sisson, 2019). As we have documented in this report, election polls are not designed to deliver this level of precision. This is the view also reached by Courtney Kennedy who was quoted as saying ‘polls aren’t designed to produce precision on the order of “so-and-so has X.X% chance of winning’ (Kennedy quoted in DeSilver, 2016).

Of course, pollsters and commentators do not just rely on the point-in-time two-party-preferred estimate from one final election poll to make their predictions. They also look to discern trends, are aware of other poll results and may have connections who can brief them on what the parties’ internal polls are showing, or may be doing some of this marginal seat polling themselves, as was the case for YouGov in 2019.

So what are our reasonable expectations of election polling and the reporting thereof? If the election polls aren’t designed to provide a precise basis for predicting the winner of close election contests, what can they be relied on for? The view of the Inquiry Panel is that well-funded, well-designed, well-executed and carefully interpreted national election polls are capable of providing a guide to:

- which of the political parties is ahead and which is behind in terms of the national vote and whether the gap between the government and opposition is widening or narrowing
- the strength of minor parties and of independents
- the likely flow of preferences from minor parties and independents to the major parties – as expressed in the two-party-preferred result – as some pollsters directly measure the second preference of respondents who express an intention to give their first preference vote to one of the two major parties.

The above are all functions for which well-designed election polls are fit for purpose. The polls are, however, not fit for some of the other purposes for which they are used, including very precisely and very reliably forecasting the winner of close election contests.

The Director of Essential Research, Peter Lewis, and Lenore Taylor, Editor of The Guardian, seemed to have reached a similar conclusion from the ‘dual vantage points of editor and pollster’. According to Lewis (2020) they asked themselves ‘whether horse-race polling had done its dash and whether we need to imagine a better way of reflecting public attitudes’. The response by Essential Research and The Guardian has been to deliberately reduce the emphasis they will give to voting intentions and to instead focus more on ‘releasing broader attitudes on leadership, on climate and more recently charting our reaction to this unprecedented (COVID-19) crisis’, choosing to only release two-party-preferred estimates quarterly in arrears.
10. Findings and recommendations

10.1. Polling ‘miss’ or polling ‘failure’?

All of the national election polls published throughout the 2019 election campaign purported to show that Labor had the support of the majority of Australian voters in terms of the two-party-preferred vote. The Coalition went on to win the election with 51.5% of the vote compared to Labor with 48.5%, almost the mirror opposite of what the final polls found.

Despite this ‘miss’, the track record of Australian pollsters since 1993 in terms of ‘calling the election outcome’ has been pretty reasonable. Across the 55 final-week polls conducted before each election since 1993, the polls have a 73% success rate. This is comparable with the performance of the US polls over a similar period (79%). More recently, for the four elections prior to 2019 – 2007, 2010, 2013, and 2016 – the published national election polls enjoyed a phenomenal 96% success rate with 25 correct calls out of 26 polls (section 5.1).

The supplemented Jennings and Wlezien time series data used in this report to track the performance of Australian election polls over time also shows the strong track record of the polls from 2007 to 2016. However, the mean absolute error of the polls in 2019 (measured across all parties with ≥3% vote share) of 1.7 percentage points was the largest error on this metric since 1998 (Figure 6) and the average error on the margin (the primary vote ‘gap’ between the Coalition and Labor) of 5.2 percentage points was the least accurate result since 1987 (Figure 7).

The supplemented Jennings and Wlezien data also show that the Australian polls have performed quite well by international standards. Compared to the international average, Australian polls produced slightly higher errors in estimating the primary vote gap between the winner and the runner up from the mid-1980s to the late 1990s, but since then they have been, on average, more accurate than the international average (Figures 8 and 9).

In the context of the performance of the Australian election polls between 2007 and 2016, the 2019 came as something of a shock.

The analysis presented in this report shows that only one of the polls (YouGov Galaxy) produced a two-party-preferred vote for the winning parties (the Coalition) within its own stated margin of error, that all of the parties called the wrong result and by roughly the same margin, and that the contribution of a late swing in mitigating the error was at best small. In short, the inaccuracy of the 2019 election polls comes down to survey methodology and statistical techniques.

Was the performance of the polls a ‘polling miss’ or a ‘polling failure’? According to Durand and Blais, a polling failure occurs, ‘when the polls: (1) significantly—statistically—err in their estimate of the vote; (2) err in the same direction and at a similar level; and (3) the source of error lie in the polls themselves—that is, their samples, methodology, weighting, likely voter models, question order and so on—not a last-minute shift among voters’ (Durand and Blais, 2020, pp. 134-135).

On this basis, what we witnessed in 2019 was not a polling miss but a polling failure.37

37 Our categorisation of the 2019 Australian polls as a ‘polling failure’ differs from the conclusion reached by Durand and Blais who, relying on a quote from Peter Lewis in The Guardian (21 May, 2019), reported that the 2019 Australian election outcome was, by and large, within polls’ margins of error. This is not supported by our analysis in section 5.4.
10.2. Factors that contributed to the failure of the published national election polls in 2019

The lack of access to data sets and detailed descriptions of the survey methods and statistical techniques used by the pollsters materially affected the ability of this Inquiry to conclusively identify the specific factors that contributed to the relative inaccuracy of the published national election polls in 2019 and identify improvements. Instead, for the most part, we have had to rely on incomplete publicly available methodological descriptions, compare against best practice and draw inferences based on our analyses of overseas inquiries and other data sources. Bearing this in mind, our findings and recommendations now follow.

10.2.1. Contextual factors

The Inquiry identified three broad contextual factors that contributed to the poor performance of the polls.

1) **Conditions for polling have become more difficult** over the last decade, and increasingly so between 2016 and 2019. Section 3.2 tracks the changes in the external environment that led to a rapid decline in contact, cooperation and response rates for polls, driving up the costs of interviewer-administered and high-quality probability-based polls.

2) **Media outlets have a reduced ability to meet these increased costs.** Polling budgets are stretched and it is now more difficult to fund the polls.

3) **The pollsters did not adjust their methods to take account of the lessons from polling failures overseas.** It is possible that the pollsters had a false sense of security given their strong performance in recent years or held the mistaken notion that our compulsory voting system somehow inoculated the Australian polls from the problems that had emerged overseas.

10.2.2. Statistical and methodological factors

Our investigations identified a number of factors that we consider to be of no or very little consequence in explaining the polls’ failure.

- **An undetected late swing:** Support for the late swing hypothesis is weak. There was no consistent evidence of a swing to the Coalition occurring during the campaign, nor did the proportion of ‘undecided’ respondents in the final polls increase as the election approached. A post-election recontact study also produced no convincing evidence of a late swing. Our analysis shows that even if there were a late swing it cannot have been anywhere near the magnitude required to account for the differences between the polls and the election result (section 6.3.7).

- **Estimating the two-party-preferred vote** relative to the primary vote (section 6.3.1). The two-party-preferred depends on: (a) the first preference votes for Labor and the Coalition; (b) the first preference votes for each of the minor parties, independents and others; and (c) the allocation of preferences, either to Labor or the Coalition, of those who voted for the minor parties, independents or others. Due to circumstances peculiar to the 2019 election, this combination of factors led to a slight increase in overall poll error.
  - The average error for the Labor vote increased from 2.6 percentage points on the primary vote to 2.9 percentage points on the two-party-preferred vote. In some polls it was more marked than in others.
- The Ipsos decision to allocate Greens preferences based on the distribution of Greens preferences at the previous election proved to be a good one, but their overestimate of support for the Greens meant they ended up with a two-party-preferred miss for Labor of 2.5 percentage points. The final Essential Report substantially overestimated support for One Nation (by 3.5 percentage points) but if Essential had been able to accurately estimate the One Nation preference flow (65% LNP : 35% ALP), this would have increased its estimate of the Coalition’s two-party-preferred vote from 48.5% to 49.5%, with a corresponding decrease for Labor, from 51.5% to 50.5%.

- **Shy conservatives**: While some respondents may have been reluctant to say that they intended to vote for the Liberal or National parties, the Panel found no evidence that this happened in sufficient numbers to affect the polls (section 6.3.5).

- **Respondents deliberately misleading pollsters**: Proponents of this theory suggest that respondents ‘game’ the polls. There is no evidence to support this as a contributing factor to the 2019 miss (section 0).

- **Early voting**: If anything, the increased take-up of early voting made the task of pollsters easier as approximately 25% of respondents to the final polls had already voted (section 0).

- **Ballot order effects**: Neither major party gained an advantage by being first or higher up the ballot. The order of the candidates on the ballot papers did not have an impact on the outcome of the election or produce a result that might somehow have rendered the polls less accurate than usual (section 6.3.2).

Our evidence, though in the absence of the pollsters’ co-operation, largely indirect, is that the all-important first preference votes were either underestimated (LNP) or overestimated (ALP) because of inadequately adjusted, unrepresentative samples.

- Very likely the polls were skewed towards over-representing more politically engaged and better educated voters with this bias not corrected.

- As a result, the polls over-represented Labor voters (section 6.4.4).

This finding stands independent of methodology because even though the methods used by the pollsters differ they share a common difficulty in struggling to establish contact with and gain the cooperation of a representative sample of voters. This conclusion is broadly similar to that reached by the reviews into the performance of the 2015 UK polls and the 2016 US polls.

The convergence of the polls:

- One of the most striking aspects of the 2019 polls was how little difference there was in the two-party-preferred figures produced by the polls. While the Panel found no direct evidence of herding, this cannot be ruled out, especially given the presence of a market influencer (Newspoll), and the widespread expectation that Labor would win – an expectation generated in no small part by a long lines of polls.

In the absence of any direct evidence of herding, the most likely explanation is a **systematic skew in the polls coupled with a confirmation bias**. The respondent pool for polling is shrinking (sections 3.2 and 6.4.3) with fewer people now able to be contacted and willing to participate in polls. Moreover, the kinds of people who are willing to participate in polls skew towards the more highly educated/more engaged in politics (section 6.4.3). Over recent election cycles there is some indication that Labor
voters are over-represented (Coalition voters under-represented) among respondents, with 17 of the 25 final poll results since 2010 (68%) overestimating 2PP support for Labor (section 5.8).

10.3. Reporting the polls
The reporting of the polls did not consistently meet the basic APC guidelines (section 8.2).

10.4. Recommendations
Establishing an appropriate level of confidence in the polls and the reporting of the polls requires realistic expectations about the level of precision that can be delivered, adequate disclosure of the methods and statistical techniques used, and appropriately nuanced reporting.

The current disclosure standards as they apply to the publicly released election polls need consolidation, clarification and greater specificity. This would best be achieved by putting in place a new code of practice and new governance structure for election polling.

The Inquiry Panel agrees with the conclusion reached by Mansillo and Jackman in their review of the 2019 polls: ‘there is a strong scientific case—and perhaps a commercial case as well—for greater transparency in how poll data are collected and weighted to ensure representativeness … Given the errors in the 2019 election, there would seem to be a strong normative case, too, for Australia’s pollsters pulling back the curtain, even a little, to help us better understand why and how they generate the numbers they do. The quality of political discussion, the salience and content of the nation’s policy agenda—indeed, the health of Australian democracy—would gain from a commitment to this transparency around the polling of Australian public opinion’ (Mansillo and Jackman, 2020, p. 145).

Governance
Recommendation 1 – Establish a Code of Conduct for Election Polling: AMSRO, as the initiators of this Inquiry, help facilitate the establishment of a Code of Conduct for Election Polling in order to provide an oversight, regulatory and disclosure regime for election polling in Australia. This could be achieved by working in tandem with the recently announced Australian Polling Council (YouGov, 2019). This Code should be in place before the next federal election polling cycle.

Recommendation 2 – Consult the experts: The development of this Code could be led by pollsters but should be informed by the views of experts from AMSRO, the Statistical Society of Australia, The Research Society, political scientists and the Australian Press Council and/or interested media outlets.

Recommendation 3 – Ensure compliance: Fundamental to the integrity and reputation of any such Code are the disclosure requirements it establishes, how these are monitored, and how compliance is ensured. We recommend a similar approach to that of the British Polling Council (BPC) in ensuring that a broad constituency sign up to the Code and that an appropriate arbitration and sanctions process is in place. BPC membership comprises almost every market research organisation in the UK that publishes political polls. Its management committee and officers are drawn from the member organisations, but its Sub-Committee on Disclosure (which provides the technical advice) comprises representatives from research organisations, academia and the media as well as member organisations. The rules of the BPC include a complaints-handling mechanism and enable sanctions to be determined including a period of probation, suspension or expulsion from the BPC (Appendix 8).

Methodology

38 16 of the 30 final polls from 1993 to 2007 (53%) underestimated the LNP vote compared to 17 out of 25 (68%) from 2010 onwards.
Recommendation 4 – Develop more effective sample balancing and/or weighting strategies:
Pollsters need to identify and better understand the biases in their samples and to develop more effective sample balancing and/or weighting mitigation strategies to improve representativeness. Weighting or balancing by education seems promising and this report suggests several other variables that may warrant further consideration (section 6.4.4 and Appendix 6 – Case Studies).

Recommendation 5 – Trial new calculation method for the two-party-preferred vote: Pollsters should not rely solely on the preference flow from the previous election to arrive at a two-party-preferred estimate, not least because some parties may be new, and for very small parties and independents preference flows from the previous election may be difficult to trace. Models that involve stated preferences with imputation of missing data or the use of stated preferences in combination with preferences from previous election could be experimentally trialled by pollsters.

Recommendation 6 – Clarify measures of uncertainty: Pollsters could use more robust methods of estimating the variability associated with their results beyond the currently inadequately calculated and inadequately reported ‘Margin of Error’ heuristic (see section 5.4). In addition, pollsters should routinely report the proportion of respondents who are ‘undecided’ about their vote choice and identify those who are only ‘leaning’ towards a particular party.

Disclosure

Recommendation 7 – Establish disclosure standards: The 24 disclosure standards set out in Appendix 10 act as a starting point for the development of a coordinated set of publicly available disclosure standards for election polls in Australia.

Recommendation 8 – Extend beyond election polling: AMSRO and The Research Society endorse the disclosure standards so that they apply to all Members and all publicly released research, not just election polling.

Recommendation 9 – Get the media outlets onside: The support of major news outlets and others who commission or publish polls should be sought to help ensure compliance with any new disclosure standards. Australian media organisations should play an active role in supporting these standards.

Resources

Recommendation 10 – Provide educational resources: Educational resources about polling methods and standards should be made available to journalists and other interested parties. This could be along similar lines to election polling resources provided by AAPOR in the US, and by the British Polling Council in Britain (AAPOR, 2020; BPC, 2019a). The Research Society, in its submission to this Inquiry, stated that it was willing to make such resources available. We recommend that AMSRO work with The Research Society and the Australian Press Council to make educational resources and training available.
Appendix 1: Terms of reference

Revised 6 July 2019

1. To assess the accuracy of the published opinion polls (both national and sub-national) at the 2019 Australian federal election.

2. To evaluate whether any inaccuracies identified might be part of a pattern evident at any previous elections or during polling since the last (2016) election.

3. To investigate the causes of any inaccuracies that are identified. Potential causes to be considered will include (but not necessarily be limited to): the possible impact of late changes in vote preferences; the extent to which sample frames (sample sources) provided adequate population coverage; sampling methods; interview methods; data weighting or other data adjustments undertaken; differential availability and willingness to participate; question order and question wording; and data item refusal.

4. To assess whether the analysis or reporting of polls was influenced by a reluctance to be out of line with the published figures of other polls.

5. To consult and seek relevant evidence from appropriate stakeholders, including but not exclusively, polling organisations, news organisations, journalists, and other polling experts.

6. To assess whether adequate information was communicated to, and communicated by, the pollsters, the media and other interested commentators to inform the public about how polls were conducted and what their results mean.

7. To make, as it sees fit, recommendations for improving how opinion polls are conducted and how their results are conveyed to the general public.
Figure 14: The evolution of polling methods in Australia, 1993 – 2019

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to Face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATI to landline only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CATI to landline and mobile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVR: landline only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVR: mobile only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MultiFrame: SMS: online panel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MultiFrame: SMS: face to face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** To be included in this table, the polling company must have conducted and publicly released a poll between the announcement of the election and the start of voting on election day (i.e. prior to 8:00 am on the east coast).
Appendix 3: Technical notes on the calculation of error metrics

Notation

*Parties* - Candidate parties in the election, e.g. LNP, LABOR, GRN, PHON, Other (combined)

*Polls* - Election polls included in the assessment, e.g. Newspoll, YouGov Galaxy, Ipsos, Essential, Roy Morgan final polls

\[ R_{ij} \] - Result of poll \( i \) for party \( j \) (%)

\[ n_i \] - Sample size for poll \( i \)

\[ V_j \] - Election result for party \( j \) (%)

\( z \) - Standardised z score for two-sided 95% confidence interval (=1.96)

\( Y \) - Election year

\( C \) - Countries in which elections are held

**Absolute Error**

**Primary vote absolute error**

For poll \( i \) and party \( j \) absolute error is calculated as:

\[ |R_{ij} - V_j| \]

For example, for Ipsos poll and GRN party the absolute error is:

\[ ABS_{GRN\text{Ipsos}} = |10.4\% - 13\%| = 2.6\% \]

**Primary vote absolute error on margin**

For poll \( i \) absolute error on the margin is calculated as:

\[ |(R_{\text{winner}}^{i} - R_{\text{runner-up}}^{i}) - (V_{\text{winner}} - V_{\text{runner-up}})| \]

For example, for Ipsos poll the absolute error on the margin is:

\[ ABS_{M^{\text{Ipsos}}} = |(39\% - 33\%) - (41.4\% - 33.3\%)| = 2.1\% \]

**2PP absolute error**

For poll \( i \) absolute error on for 2PP is calculated as:

\[ |R_{2PP}^{i} - V_{2PP}| \]

For example, for Ipsos poll the absolute error on 2PP is:

\[ ABS_{2PP^{\text{Ipsos}}} = |49\% - 51.5\%| = 2.5\% \]
Average Absolute Error

Unweighted

For poll $i$ unweighted average absolute error for parties $j \in Parties$ is calculated as:

$$ AAE^i = \frac{\sum_{j \in Parties} ABS^i_j}{\sum_{j \in Parties} k_j} $$

where $k_j = \begin{cases} 1 & \text{if } j \in Parties \\ 0 & \text{otherwise} \end{cases}$

For example, for Ipsos poll unweighted average absolute error for LNP, LABOR, GRN, PHON and OTH (all other parties combined) is:

$$ AAE^{Ipsos}_{LNP,LABOR,GRN,PHON,OTH} = \frac{(2.4 + 0.3 + 2.6 + 0.9 + 0.8)}{5} = 1.4 $$

Weighted by share of primary vote

For poll $i$ weighted average absolute error by share of primary vote for parties $j \in Parties$ is calculated as:

$$ WAAE^i = \frac{\sum_{j \in Parties} ABS^i_j \times V_j}{\sum_{j \in Parties} V_j} $$

For example, for Ipsos poll weighted average absolute error by share of primary vote for LNP, LABOR, GRN, PHON and OTH (all other parties combined) is:

$$ WAAE^{Ipsos}_{LNP,LABOR,GRN,PHON,OTH} = \frac{(2.4 \times 4.1 + 0.3 \times 3.3 + 2.6 \times 10.4 + 0.9 \times 3.1 + 0.8 \times 12.8)}{100} = 1.5 $$

Margin of Sampling Error (MoSE or MoE)

The phrase ‘margin of error’, which should be more accurately be referred to as the ‘margin of sampling error’ is currently (and imprecisely) used by the pollsters to convey the degree of uncertainty surrounding their results.

This uncertainty arises from the fact that poll results are estimates based on a sample of the voting population and not a census of the entire voting population and are therefore subject to a margin of ‘sampling’ error. The way that the ‘margin of error’ is calculated and applied by the pollsters rests on the assumption that the survey estimates are based on a simple random probability sample of eligible and enrolled voters. In fact, none of the polls use simple random samples, most did not use probability samples, and none factor into their margin of error calculations the complex nature of their sample design, or the impact that weighting has on sample variance. As a result, there are good reasons to suspect that all of the pollsters understate the sampling error associated with their polls.

For this note we will refer to MoSE as MoE to avoid confusion as this is the commonly used and understood label within the polling community.

MoE for proportion

MoE is traditionally calculated for all polls using a simple random probability sample assumption and the actual poll result (as opposed to 50%, which is a more conservative way of calculating MoE when the actual proportion is unknown and is sometimes used by pollsters when communicating their uncertainty).

$$ MoE^i_j = z \times \sqrt{\frac{\hat{n}_j}{100} \times \left(1 - \frac{\hat{n}_j}{100}\right)} $$

(Cochran 1977)

For example, for Ipsos LNP, MoE is:
MoE for difference between two proportions (same poll)

Although MoE for the party share of the vote is the one that is usually referred to when poll results are published, when winning margin or gap between the parties or 2PP is discussed, different MoE calculations based on the formula for the variance of the difference of two multinomial proportions (adapted by Franklin, 2007 from Kish, Survey Sampling, 1965, pp. 498-501) are as follows:

\[
\text{MoE}^{ij}_{j_1-j_2} = z \times \sqrt{\frac{R_{j_1}^{i1} \times (100 - R_{j_1}^{i1}) + R_{j_2}^{i2} \times (100 - R_{j_2}^{i2})}{n_i}}
\]

Continuing with the Ipsos example, the MoE for 2PP ALP (51%) and 2PP LNP (49%) difference of 2% is:

\[
\text{MoE}^{Ipsos}_{\text{LNP-ALP}} = 1.96 \times \frac{(0.49 + 0.51) - (0.49 - 0.51)^2}{1842 - 1} = 0.045 \text{ or } 4.5%
\]

MoE for difference between polls

As shown by Franklin, 2007, if the difference of interest is change in polling share for the same party in to separate polls that use independent samples (e.g. difference over time) then the MoE is calculated as follows:

\[
\text{MoE}^{i1,i2} = z \times \sqrt{\frac{R_{j_1}^{i1} \times (100 - R_{j_1}^{i1}) \times n_{i1} + R_{j_2}^{i2} \times (100 - R_{j_2}^{i2}) \times n_{i2}}{n_i}}
\]

For a poll of the same size (e.g. Ipsos n = 1842) and the most conservative 95% confidence interval (estimated primary vote of 50%), MoE is calculated as

\[
\text{MoE}^{Ipsos}_{\text{LNP}} = 1.96 \times \sqrt{\frac{0.50 \times (1 - 0.50) + 0.50 \times (1 - .50)}{1842}} = 0.032 \text{ or } 3.2%
\]

Having calculated MoE, 95% Confidence Intervals are calculated as estimated vote plus/minus MoE.

Mean absolute error (MAE)

When making comparisons across time and space, the results of many elections and polls need to be summarised. For this purpose, Jennings and Wlezien (2017) introduce the concept of Mean Absolute Error (MAE), which is the average of absolute error or average of absolute error on the margin for each party across all polls.

For a specific country and year, the MAE is calculated as:

\[
\text{MAE}_C^Y = \frac{\sum_{j \in \text{Parties}} \frac{\sum_{i \in \text{Polls}} \text{ABS}_{i}^{j}}{\sum_{i \in \text{Polls}} l_i}}{\sum_{j \in \text{Parties}} \sum_{i \in \text{Polls}} l_i}
\]

where

\[
l_i = \begin{cases} 1 & \text{if } i \in \text{Poll in country } C \text{ and year } Y \text{ for party } j \\ 0 & \text{otherwise} \end{cases}
\]

This is frequently calculated as

\[
2 \times \sqrt{\frac{R_i^{j1} \times (1 - R_i^{j1})}{n_i}}
\]
\[ k_j = \begin{cases} 1 & \text{if } j \in \text{Parties for parties which have at least 3\% of the primary vote} \\ 0 & \text{otherwise} \end{cases} \]

For example, for the 2019 Australian election, the corresponding MAEs for parties with at least 3\% of the primary vote are (refer to Table 19: MAE calculations example, Australia 2019 for party averages and to Table 4 for errors on the margin):

\[
\text{MAE}_{\text{Australia}}^{2019} = \frac{2.84 + 2.54 + 1.42 + 1.10 + 0.63}{5} = 1.71
\]

\[
\text{MAE}_{M}^{\text{Australia}}^{2019} = \frac{7.1 + 6.1 + 2.1 + 5.8 + 5.1}{5} = 5.24
\]

Table 19: MAE calculations example, Australia 2019

<table>
<thead>
<tr>
<th>Poll</th>
<th>Primary vote</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LNP</td>
</tr>
<tr>
<td>Newspoll</td>
<td>3.4</td>
</tr>
<tr>
<td>YouGov Galaxy</td>
<td>2.4</td>
</tr>
<tr>
<td>Ipsos</td>
<td>2.4</td>
</tr>
<tr>
<td>Essential</td>
<td>2.9</td>
</tr>
<tr>
<td>Roy Morgan</td>
<td>2.9</td>
</tr>
<tr>
<td>Party Average</td>
<td>2.8</td>
</tr>
</tbody>
</table>
## Appendix 4: Error estimates in the final polls 1993-2019

### Table 20: Error in estimates of the Liberal-National Party share of the two-party-preferred vote, final polls, 1993-2019 (percentage points)

| Election | AGB McNair/ACNielsen | Roy Morgan | Newspoll | Saulwick | Quadrant | ARS | ANU | Galaxy/YouGov Galaxy | Essential | JWS | AMR | Lonergan | ReachTEL | Ipsos | |MA| |
|----------|----------------------|------------|-----------|----------|----------|-----|-----|-----------------------|-----------|-----|-----|---------|----------|-------|-----|-----|
|          | CATI†                | CATI†      | CATI†     | CATI†    | CATI†    | Online | CATI† | Online                | CATI†     | CATI† | IVR¥ | IVR¥     | CATI‡    |       |     |     |
| 1993     | +1.9*                | +0.9*      | +1.9      | +0.4*    |          |        |        |          |           |       |       |       |          | 1.3      |       |     |     |
| 1996     | -2.6*                | -8.6       | -0.1*     | -2.6     |          |        |        |          |           |       |       |       |          | 3.5      |       |     |     |
| 1998     | +1.0*                | +1.0*      | -2.0*     | +1.0*    | +1.0*    |        |        |          |           |       |       |       |          | 1.7      |       |     |     |
| 2001     | +1.1*                | -5.4 Ftf   | +2.1      | +1.1*    |          |        |        |          |           |       |       |       |          | 2.5      |       |     |     |
| 2004     | +1.3*                | -3.7       | -2.7      | -6.7     | -0.7*    |        |        |          |           |       |       |       |          | 3.2      |       |     |     |
| 2007     | -2.3                 | -0.8*      | +0.7*     | +0.7*    |          |        |        |          |           |       |       |       |          | 1.9      |       |     |     |
| 2010     | -1.9*                | -0.9*      | -0.1*     | -1.9*    | -0.9*    | -1.9   |        |          |           |       |       |       |          | 1.4      |       |     |     |
| 2013     | +0.5*                | 0.0* mixed | +0.5*     | -0.5*    | -1.5*    | +0.5*  | -2.7* | -0.5*     |           |       |       |       |          | 0.8      |       |     |     |
| 2016     |                      |            |          |          |          |        | +0.6*| +0.1*     | +0.6*     |       |       | -1.4* |          | 0.6      |       |     |     |
| 2019     | -3.5 FtF             | -3.0 #     |          |          |          |        | -2.5*| online   | -2.5*     |       |       |       |          | 2.9      |       |     |     |

*Within the margin of error, assuming a 95% confidence interval

Shading: N > 2,500
+ overestimate; - underestimate
† landline only; ‡ landline + mobile; ¥ mobile only
# IVR + online
Ftf: face-to-face
Mixed: SMS + CATI† + online
|M| Absolute mean

Source: Goot (in press, Table 2)
Appendix 5: Total Survey Error framework

The framework used by the Inquiry Panel to try and determine the extent to which any methodological features or statistical techniques may have contributed to the inaccuracy of the public domain national election polls in 2019 is the Total Survey Error (TSE) framework (Groves, 2004; Groves and Lyberg, 2010, Lavrakas, 2013). TSE refers to the ‘accumulation of all errors that may arise in the design, collection, processing and analysis of survey data’ (Biemer, 2010). Adopting a TSE perspective informs survey design decisions, and sometimes trade-offs, so that resources are allocated in such a way as to reduce total survey error for key estimates. The TSE paradigm is an extension of the much broader concept of Total Survey Quality used by many national statistical agencies, including the Australian Bureau of Statistics.

The TSE framework was used to look for potential sources of error (bias and variance) in the techniques utilised by the pollsters in the hope of identifying improvements.

The representation side of the framework is where errors of non-observation occur. These include:

- **Coverage error** – ‘gaps’ in the sampling frame compared to the population of interest (e.g. the exclusion of persons without a landline from landline surveys or the exclusion of the offline population from online panels). This type of error is typically associated with bias.

- **Sampling error** – the uncertainty surrounding survey estimates arising from the fact that these estimates are based on a sample and not a census of the entire voting population. Inefficient or inappropriate sample designs can lead to additional sampling error. This type of error is typically associated with variance.

- **Non-response errors** – at both the unit-level (a function of non-contacts, refusals and being unable to participate; e.g., language barriers) and at the item level (when a respondent may be unwilling or unable to answer a particular question). For unit-level non-response, this type of error is typically associated with bias. For item-level non-response (aka ‘missing data’), this type of error can be associated with bias and/or variance.

- **Adjustment errors** – made where the final sample needs to be adjusted to account for the design effects introduced by the sample design and non-response. This is accomplished by weighting that tries to reduce bias, but generally adds error in the form of variance (imprecision) to the study’s findings. A well-designed weighting solution balances variance and bias with a view to reducing overall survey error.
The measurement side of the model is where errors of observation occur. These include:

- **Specification error** (sometimes called construct validity) – arises when the survey questions or response scales do not adequately capture the construct or domain they are intended to measure (e.g. voting intention). This type of error is typically associated with bias.

- **Measurement error** – arises from many sources, especially from poor questionnaire design (wording, ordering of items, formatting), mode effects, interviewer errors and respondent errors. These types of error may be associated with bias and/or variance.

- **Processing errors** – arise from how the raw data is transformed and can be attributed to the coding of free text or verbatim responses, the treatment of outliers, the imputation of missing data, data derivations, etc. These types of error may be associated with bias and/or variance.

- **Inferential errors** – can be introduced to the survey process at the stage of interpreting the survey findings. These types of error may be associated with bias and/or variance.

From a TSE perspective, ‘sampling error’ is just one of many types of error. Adopting a TSE perspective helped the Inquiry Panel identify all potential sources of error.
Appendix 6: Weighting case studies

Case studies 1 and 2: Including educational attainment in weighting solutions

Case Study 1 – The Comparative Study of Electoral Systems survey (CSES Australia, 2019)

The CSES Australia was conducted in June 2019, one month after the election, and included a measure of reported voting behaviour: ‘In the Federal election for the House of Representatives on Saturday 18 May, which party did you vote for first in the House of Representatives?’ The sample used for the CSES Australia was the Life in Australia™ probability-panel and the cumulative response rate for the survey was 8.4% 40. Of the 2,000 respondents who participated in CSES Australia, 278 respondents were either not eligible (e.g. not citizens) or chose not to answer the voting question. The reported voting behaviour of the remaining 1,722 respondents is the basis for this analysis.

Table 21 shows that adding education to the weighting solution improves the accuracy of the weighted estimates for all parties that won at least 3% of the vote by 21% and reduces both the two-party-preferred error and the absolute error on the margin by 33%. 41 Despite a loss of accuracy for ‘other parties’, adding education to the weighting solution results in a considerable reduction in bias compared to just weighting by sex, age and location.

Since the polls failed principally because they underestimated the Coalition vote, and since the polls didn’t weight for education, the impact of weighting by education on the voting estimates produced by the CSES is of potential importance. If weighting for education were to have had roughly the same impact on the election polls as it did on the CSES Australia sample, then the estimate of the Coalition’s vote would have been about 40.5% rather than 38.6%. This would have increased the gap between the Labor and Coalition primary vote estimates from the poll average of 2.8 percentage points to 4.7 percentage points; a good deal closer to the actual gap at the election of 8.1 percentage points.

Table 21: The impact of weighting by educational attainment on the vote choice estimates produced from the CSES Australia survey, 2019.

| Party          | Election Result | Percentage Difference from Election Outcome | | | | | | | Change (+/-%)
|----------------|-----------------|---------------------------------------------|--|--|--||--|--|--|
| LNP            | 41.4            | 2.2                                         | -2.1| -0.6| -69.0| | | | |
| Labor          | 33.3            | 0.5                                         | 2.3| 2.3| -0.1| | | | |
| Greens         | 10.4            | 3.0                                         | 5.7| 4.6| -19.1| | | | |
| One Nation (PHON) | 3.1           | -1.6                                        | -1.5| -1.2| -21.8| | | | |

40 The cumulative response rate (CUMRR2) takes account of non-response at each point of the recruitment and response cycle. It is the product of the recruitment rate (RECR), the profile rate (PROR), the retention rate (RETR) and the completion rate: CUMRR2 = RECR × PROR × RETR × COMR. The recruitment rate is the rate at which eligible individuals agree to join the panel. The profile rate is the rate at which initially consenting individuals complete the panel profile, thus joining the panel.

41 The error on the margin refers to the ‘gap’ between the ALP and LNP on the primary vote. For example, if the election outcome were ALP 51 and LNP 49, then the gap between the winner and runner up is +2 points. If the poll estimate were ALP 49 and LNP 51, then the ‘gap’ between the winner and runner up is –2 points. In this case, the error on the margin would be 4 points.
The second case study uses data from another post-election probability survey, the Australian Election Study. For this exercise, we limit our analysis to the 2019 post-election cross-sectional survey that used an Address-Based Sampling frame to obtain completed questionnaires from a sample of 1,211 respondents aged 18 years and over with a response rate (AAPOR RR1) of 30.7%. Of the 1,211 respondents, 1,117 were eligible to vote and answered the voting question. The unweighted sample profile for the AES has the usual failings in that males, young people and non-university graduates are all under-represented.

The sample selection and data collection mechanisms for the AES sample are radically different from those used by pollsters:

- The sampling frame used for the AES, the Geocoded National Address File (G-NAF), is not used by the pollsters.
- The sample is probability proportional to size random sample of addresses.
- Quotas for age, gender and geography are not used.
- Approach letters and survey materials are designed specifically for the study. (Respondents are approached via mail.)
- There is extensive follow-up of non-responders.
- The data are collected through multiple modes (hard copy and online).
- The data collection period extends over four months to maximise response.

As a result of these differences, weighting is expected to impact differently on the AES from the samples used by the pollsters. Accordingly, findings from this case study cannot be generalised to election polling, other than to highlight that weighting solutions need to be developed in the context of each sample and with an understanding of the sample/survey’s biases.

Table 22 shows the unweighted AES estimates are quite skewed and particularly so in terms of support for the Coalition, with an overestimate of 6.6 percentage points. For this data set and for this single outcome variable (reported vote choice), weighting by sex, age and geography reduces this overestimate from 6.6 percentage points to 2.8 percentage points. On the other hand, the slight overestimate of support for the Greens (0.8 percentage points) increases to 4.6 percentage points.

---

42 Excluding non-citizens and those with a null response to the voting question.
43 The data set and associated documentation is available from the Australian Data Archive https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/SYYFCS
Adding education to the weighting solution has only a modest impact on the estimates. It reduces the average absolute error (all parties ≥3%) by 6%, two-party average absolute error by 8%, and the absolute error on the margin by 23%. For the AES, adding education to the weighting solution results in only a modest bias reduction; most of the reduction comes from the sex, by age and by location weight. This may not be particularly surprising – in the five elections held between 1993 and 2004, only in 2001 in the AES was there a significant connection between education and the vote, net of other demographic variables (Goot and Watson 2007, p. 273).

Table 22: The impact of weighting by educational attainment on the vote choice estimates produced from the 2019 AES.

<table>
<thead>
<tr>
<th>Party</th>
<th>Election result</th>
<th>Percentage Difference from Election Outcome</th>
<th>Change (±/-%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unweighted</td>
<td>Weighted by sex, geography, age</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighted by sex, geography, age and education</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>LNP</td>
<td>41.4</td>
<td>6.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Labor</td>
<td>33.3</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>Greens</td>
<td>10.4</td>
<td>0.8</td>
<td>4.6</td>
</tr>
<tr>
<td>One Nation (PHON)</td>
<td>3.1</td>
<td>-1.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>United Australia Party (UAP)</td>
<td>3.4</td>
<td>-3.0</td>
<td>-3.1</td>
</tr>
<tr>
<td>Other</td>
<td>8.3</td>
<td>-3.5</td>
<td>-3.5</td>
</tr>
<tr>
<td>Average absolute error on the primary vote (All parties ≥3% of vote)</td>
<td>1.0</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Average absolute error on the primary vote (Two top parties)</td>
<td>2.5</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Absolute error on the primary vote margin (Top two parties)</td>
<td>3.6</td>
<td>2.1</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Source: 2019 Australia Election Study

Case Study 3: Additional bias reduction from alternative weighting solutions

For this illustration we limit the number of additional variables to two – in addition to gender, geography, age and education. While it would be possible to include many more variables in a weighting solution, any solution that attempted to do so would quickly become highly inefficient: it would contain very high and low weighting values that would add a lot of variance, reducing the accuracy of the estimates when the variables being estimated were not strongly related to the weighting variables.

Weighting by life satisfaction and self-reported health status

In the US, Schlenker et al. (2011, p. 127) report that ‘conservatives are happier than liberals’. In surveys of people from across the globe, conservatives report being more satisfied with their lives than liberals (Carroll, 2007; Napier & Jost, 2008; Taylor, Funk, & Craighill, 2006). Gallup reported that between 2005 and 2007, 61% of Republicans but only 47% of Democrats said they were very happy (Carroll, 2007). The ‘ideology–happiness relationship is reliable, small to moderate in size, and potentially quite meaningful given its implications for understanding political ideologies and behaviour’ (Schlenker et al., p. 127). Table 23 shows that weighting the CSES Australia 2019 sample to life
satisfaction benchmarks increases the primary vote estimate for the Coalition (LNP) while reducing it for Labor and the Greens. Since the election polls underestimated support for the Coalition and overestimated support for Labor, this is a notable result.

Huijts et al. (2010, p. 3) use European Social Survey data to examine the relationship between political ideology and self-reported health status and confirm earlier findings from Europe and the US that those with a conservative political ideology are less likely to report being in poor health, net of other factors. The same relationship is seemingly evident in the CSES Australia sample. Weighting the sample by self-reported general health increases the primary vote estimate for the Coalition and reduces it for Labor.

The various weighting scenarios are summarised in Table 23. When added to sex, geography and age by education, the combinations of variables that bring about the greatest additional error reduction are as follows:

- Life satisfaction and employment status, and employment status with self-assessed health status, which remove all error (down from 2.8 percentage points) on the estimate of the margin between Labor and the Coalition as measured by the polls compared to the actual ‘gap’ at the election.

- Self-assessed health in combination with a Concession card adjustment, which brings the greatest reduction in the average error, from 1.4 to 0.8 percentage points.

- Country of birth with Concession card, which results in the greatest (albeit slight) reduction in error across all parties with ≥3% of the vote, reducing it from 2.1 to 1.9 percentage points.

Which weighting combination is ultimately preferred may depend on which of the errors is considered the most important to reduce or which combination brings about the most effective reduction across all metrics. It also depends on the relative weighting efficiency of each solution; this element is explored in Figure 16: Comparison of weighting solutions impact on measures of absolute error.
Table 23: Summary of the impact of various weighting solutions used for the CSES Australia survey on selected error metrics

<table>
<thead>
<tr>
<th>Weighting variables</th>
<th>Weighting efficiency (%)</th>
<th>Absolute error on primary vote (Percentage points)</th>
<th>Average</th>
<th>Margin (Two top parties)</th>
<th>Margin (All parties ≥3% of vote)</th>
<th>Coalition</th>
<th>Labour</th>
<th>Green</th>
<th>Other parties (includes PHON and UAP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, Geography and Age</td>
<td>64.4</td>
<td></td>
<td>4.4</td>
<td>2.2</td>
<td>2.8</td>
<td>2.1</td>
<td>2.3</td>
<td>5.7</td>
<td>5.9</td>
</tr>
<tr>
<td>Sex, Geography, Age and Education</td>
<td>56.0</td>
<td></td>
<td>3.0</td>
<td>1.5</td>
<td>2.2</td>
<td>0.6</td>
<td>2.3</td>
<td>4.6</td>
<td>6.3</td>
</tr>
<tr>
<td>Sex, Geography and Age by Education</td>
<td>57.2</td>
<td></td>
<td>2.8</td>
<td>1.4</td>
<td>2.2</td>
<td>0.6</td>
<td>2.3</td>
<td>4.5</td>
<td>6.2</td>
</tr>
<tr>
<td>Life Satisfaction, Employment</td>
<td>49.7</td>
<td></td>
<td>0.0</td>
<td>1.8</td>
<td>2.1</td>
<td>1.8</td>
<td>1.8</td>
<td>3.1</td>
<td>6.6</td>
</tr>
<tr>
<td>Employment, Self-Assessed Health</td>
<td>49.4</td>
<td></td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
<td>0.9</td>
<td>0.9</td>
<td>4.5</td>
<td>6.3</td>
</tr>
<tr>
<td>Self-Assessed Health</td>
<td>50.7</td>
<td></td>
<td>0.1</td>
<td>0.9</td>
<td>2.0</td>
<td>0.9</td>
<td>0.8</td>
<td>4.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Life Satisfaction</td>
<td>50.9</td>
<td></td>
<td>0.3</td>
<td>1.7</td>
<td>2.0</td>
<td>1.8</td>
<td>1.5</td>
<td>3.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Self-Assessed Health, Country of Birth</td>
<td>48.7</td>
<td></td>
<td>0.4</td>
<td>1.2</td>
<td>2.0</td>
<td>1.3</td>
<td>0.9</td>
<td>4.1</td>
<td>6.4</td>
</tr>
<tr>
<td>Life Satisfaction, Country of Birth</td>
<td>49.3</td>
<td></td>
<td>0.5</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
<td>1.7</td>
<td>2.8</td>
<td>6.7</td>
</tr>
<tr>
<td>Self-Assessed Health, Concession Card</td>
<td>50.5</td>
<td></td>
<td>0.6</td>
<td>0.8</td>
<td>2.0</td>
<td>1.1</td>
<td>0.5</td>
<td>4.8</td>
<td>6.3</td>
</tr>
<tr>
<td>Life Satisfaction, Concession Card</td>
<td>50.8</td>
<td></td>
<td>0.7</td>
<td>1.7</td>
<td>2.0</td>
<td>2.0</td>
<td>1.3</td>
<td>3.4</td>
<td>6.6</td>
</tr>
<tr>
<td>Country of Birth, Concession Card</td>
<td>54.4</td>
<td></td>
<td>1.8</td>
<td>1.2</td>
<td>1.9</td>
<td>0.2</td>
<td>2.0</td>
<td>3.9</td>
<td>6.2</td>
</tr>
<tr>
<td>Life Satisfaction, Self-Assessed Health</td>
<td>47.3</td>
<td></td>
<td>2.1</td>
<td>1.5</td>
<td>2.1</td>
<td>2.5</td>
<td>0.4</td>
<td>3.8</td>
<td>6.6</td>
</tr>
<tr>
<td>Concession Card</td>
<td>56.9</td>
<td></td>
<td>2.1</td>
<td>1.1</td>
<td>2.0</td>
<td>0.2</td>
<td>1.9</td>
<td>4.5</td>
<td>6.2</td>
</tr>
<tr>
<td>Employment, Concession Card</td>
<td>55.3</td>
<td></td>
<td>2.3</td>
<td>1.2</td>
<td>2.0</td>
<td>0.2</td>
<td>2.1</td>
<td>4.3</td>
<td>6.1</td>
</tr>
<tr>
<td>Country of Birth</td>
<td>54.6</td>
<td></td>
<td>2.6</td>
<td>1.3</td>
<td>2.0</td>
<td>0.2</td>
<td>2.4</td>
<td>4.0</td>
<td>6.2</td>
</tr>
<tr>
<td>Employment, Country of Birth</td>
<td>53.7</td>
<td></td>
<td>3.1</td>
<td>1.5</td>
<td>2.1</td>
<td>0.4</td>
<td>2.7</td>
<td>3.8</td>
<td>6.2</td>
</tr>
<tr>
<td>Employment</td>
<td>56.1</td>
<td></td>
<td>3.3</td>
<td>1.7</td>
<td>2.3</td>
<td>0.8</td>
<td>2.5</td>
<td>4.4</td>
<td>6.1</td>
</tr>
</tbody>
</table>

Source: Authors’ analysis
The vertical bars in Figure 16 show the impact of the various weighting solutions on the relevant error metrics in terms of the percentage point reduction in error (the axis on the left side of the figure); the average errors for the polls are also shown as a reference point. The plotted line on the chart shows the relative weighting efficiency of each solution against the right axis of the chart (the higher the weighting efficiency the better; i.e. the impact on variance is less pronounced). Balancing these two dimensions enables us to identify the optimal weighting solution. This is identified as self-assessed health status. This solution has a mid-range weighting efficiency and provides close to zero (0.1 percentage points) bias for the margin on the primary vote between the two major parties and is among the lowest in terms of two-party and all-party average absolute errors: 0.9 and 2.0 percentage points respectively (refer to Table 23).

The addition of either self-assessed health or life satisfaction benchmarks is a common feature of all lower error weighting/ reasonable efficiency solutions.

**Figure 16: Comparison of weighting solutions impact on measures of absolute error**

Source: Authors’ analysis
Table 24: Population benchmarks used for weighting the CSES Australia and AES case studies.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Population %</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>18-24 years</td>
<td>12.2%</td>
<td></td>
</tr>
<tr>
<td>25-34 years</td>
<td>19.3%</td>
<td></td>
</tr>
<tr>
<td>35-44 years</td>
<td>17.1%</td>
<td></td>
</tr>
<tr>
<td>45-54 years</td>
<td>16.5%</td>
<td></td>
</tr>
<tr>
<td>55-64 years</td>
<td>14.9%</td>
<td></td>
</tr>
<tr>
<td>65 or more years</td>
<td>20.1%</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>Female</td>
<td>50.9%</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49.1%</td>
<td></td>
</tr>
<tr>
<td>Age by Education</td>
<td></td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>18-24</td>
<td>12.2%</td>
<td></td>
</tr>
<tr>
<td>25-34 Bachelor and above</td>
<td>7.4%</td>
<td></td>
</tr>
<tr>
<td>25-34 Below Bachelor</td>
<td>11.8%</td>
<td></td>
</tr>
<tr>
<td>35-44 Bachelor and above</td>
<td>6.2%</td>
<td></td>
</tr>
<tr>
<td>35-44 Below Bachelor</td>
<td>10.9%</td>
<td></td>
</tr>
<tr>
<td>45-54 Bachelor and above</td>
<td>4.3%</td>
<td></td>
</tr>
<tr>
<td>45-54 Below Bachelor</td>
<td>12.2%</td>
<td></td>
</tr>
<tr>
<td>55-64 Bachelor and above</td>
<td>3.3%</td>
<td></td>
</tr>
<tr>
<td>55-64 Below Bachelor</td>
<td>11.6%</td>
<td></td>
</tr>
<tr>
<td>65+ Bachelor and above</td>
<td>2.7%</td>
<td></td>
</tr>
<tr>
<td>65+ Below Bachelor</td>
<td>17.4%</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>Greater Sydney</td>
<td>20.7%</td>
<td></td>
</tr>
<tr>
<td>Rest of NSW</td>
<td>11.3%</td>
<td></td>
</tr>
<tr>
<td>Greater Melbourne</td>
<td>19.8%</td>
<td></td>
</tr>
<tr>
<td>Rest of Vic</td>
<td>6.3%</td>
<td></td>
</tr>
<tr>
<td>Greater Brisbane</td>
<td>9.6%</td>
<td></td>
</tr>
<tr>
<td>Rest of Qld</td>
<td>10.2%</td>
<td></td>
</tr>
<tr>
<td>Greater Adelaide</td>
<td>5.5%</td>
<td></td>
</tr>
<tr>
<td>Rest of SA</td>
<td>1.6%</td>
<td></td>
</tr>
<tr>
<td>Greater Perth</td>
<td>8.1%</td>
<td></td>
</tr>
<tr>
<td>Rest of WA</td>
<td>2.2%</td>
<td></td>
</tr>
<tr>
<td>Greater Hobart</td>
<td>0.9%</td>
<td></td>
</tr>
<tr>
<td>Rest of Tas</td>
<td>1.2%</td>
<td></td>
</tr>
<tr>
<td>Greater Darwin</td>
<td>0.6%</td>
<td></td>
</tr>
<tr>
<td>Rest of NT</td>
<td>0.4%</td>
<td></td>
</tr>
<tr>
<td>Australian Capital Territory</td>
<td>1.7%</td>
<td></td>
</tr>
<tr>
<td>Table 24 (cont.): Population benchmarks used for weighting the CSES Australia and AES case studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor and above</td>
<td>25.5%</td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>Below Bachelor</td>
<td>74.5%</td>
<td></td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>61.6%</td>
<td>ABS Census 2016</td>
</tr>
<tr>
<td>Not employed</td>
<td>38.4%</td>
<td></td>
</tr>
<tr>
<td><strong>Country of Birth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian born</td>
<td>66.3%</td>
<td>ABS Census 2016 with ERP March 2019 adjustment</td>
</tr>
<tr>
<td>Mainly ESB background</td>
<td>10.6%</td>
<td></td>
</tr>
<tr>
<td>Mainly NESB background</td>
<td>23.1%</td>
<td></td>
</tr>
<tr>
<td><strong>Self-Assessed Health</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>20.2%</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>35.5%</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>29.1%</td>
<td>NHS 2018</td>
</tr>
<tr>
<td>Fair</td>
<td>11.3%</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>3.9%</td>
<td></td>
</tr>
<tr>
<td><strong>Life Satisfaction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score = 0-4</td>
<td>4.9%</td>
<td></td>
</tr>
<tr>
<td>Score = 5</td>
<td>7.6%</td>
<td></td>
</tr>
<tr>
<td>Score = 6</td>
<td>6.8%</td>
<td></td>
</tr>
<tr>
<td>Score = 7</td>
<td>18.4%</td>
<td></td>
</tr>
<tr>
<td>Score = 8</td>
<td>32.5%</td>
<td></td>
</tr>
<tr>
<td>Score = 9</td>
<td>15.8%</td>
<td></td>
</tr>
<tr>
<td>Score = 10</td>
<td>14.0%</td>
<td></td>
</tr>
<tr>
<td><strong>Concession Card</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has Concession card</td>
<td>31.8%</td>
<td>NHS 2015</td>
</tr>
<tr>
<td>Does not have Concession card</td>
<td>68.2%</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 7: The convergence of the polls – what are the odds?

The odds of getting a series of such similar results were calculated as follows:

The outcome of interest is the two-party-preferred vote for the Coalition with a mean of 51.5% and a standard deviation of 1.16%.

- The mean is the actual election outcome.
- The standard deviation is based on the published margins of errors for the five opinion polls (+/-3.2% for YouGov Galaxy, 2.8% for Roy Morgan, 2.5% for Essential, 2.4% for Ipsos and 1.7% for Newspoll). The average margin of error across the five polls (2.3%) was calculated as reflecting a 95% confidence interval; the standard deviation is half this number.
- A normal distribution for the outcome of independent opinion polls was assumed.

On this measure, the poll that performed best was YouGov Galaxy, which was 1.56 standard deviations from the election result. The probability of being 1.56 standard deviations or more from the mean is 0.059. The probability of all five independent polls being worse than this is approximately 1 in 1.5 million. However, this is generous to the polls as all the others were more than 2 standard deviations away and considerably more in the case of Essential and Roy Morgan. For example, you would get a much lower probability of one poll being at least one standard deviation below the mean and four polls being at least two standard deviations below the mean.

A possible explanation is a systematic bias similarly affecting all these polls. The average of the two-party-preferred estimate for the Coalition across the five polls is 48.6% with two polls showing 49% (0.4% from adjusted mean); two showed 48.5% (0.1% from adjusted mean) and one showed 48% (0.6% from adjusted mean). Making the same assumptions as in the first paragraph but with a mean of 48.6%, the probability of being within 0.6% of the mean is 0.45. The probability of all five independent polls being in this range is approximately 2 chances in 100.
Appendix 8: International and Australian standards

To inform our recommendations as to the appropriate disclosure standards for Australian election polls, we examined some of the international models used to prescribe such standards. This was not an exhaustive examination of such models, but a review of those models that came to our attention in the course of our broader work on this Inquiry.

The ESOMAR/WAPOR Guideline on Opinion Polls and Published Surveys is the main international standard. The minimal requirements set out by the ESOMAR/WAPOR code with respect to opinion polls and published surveys is that any published report should ‘include the name of the research company and sponsor of the survey, the universe, the sample size, the mode of interview and dates of interviewing. If space is limited the other relevant information listed in the guideline should be made available online. Any publication should also make clear how additional details can be obtained and this information should be available within 24 hours of publication (ESOMAR/WAPOR, 2014a).

The general requirements of the ESOMPAR/WAPOR code for opinion polls and published surveys are supported by the ESOMAR/WAPOR Guideline on Opinion Polls and Published Surveys that states ‘the following information must be included in the survey report, or made available online or in other published form’ (ESOMAR/WAPOR, 2014b, pp. 7-8):

a) The name of the organisation which conducted the poll and its sponsor, the organisation(s) or person(s) who paid for the poll
b) The universe effectively represented (i.e. who was interviewed)
c) The actual sample size and its geographical coverage
d) The dates of fieldwork
e) The sampling method used. For quota samples and other non-probability samples, provide the characteristics by which the sample was selected. For probability samples, additional information, including the response rate, must be provided on request
f) The method by which the poll was conducted (face-to-face, telephone interview, internet access panel, mixed mode, etc.).
g) Whether weighting was used to adjust the results and the general demographic or behavioural characteristics used for the weights. The general weighting variables should be described but proprietary algorithms and specific weighting variables do not need to be disclosed.
h) The percentages of respondents who give ‘don't know’ answers (and in the case of voting-intention studies, of those who say they will not vote)
i) The relevant questions asked. In order to avoid possible ambiguity the actual wording of the question should be given unless this is a standard question already familiar to the audience, such as an approval rating of the government or the government’s leaders or has been given in a previously published report to which reference is made.

The other international standard that has elements that can be applied to the reporting of election and other political polling is ISO 20252: the international standard for market, opinion and social research,
including insights and data analytics. Although the ISO standards are not designed specifically for election polling, some of the standards seem broadly applicable:

- Disclose name of client and research provider.
- Disclose and identify any subcontracted services.
- Report where applicable: the target population; methods of data collection; fieldwork dates; incentive types; methods of statistical analysis and margin of sampling error; and provide a statement of substantial limitations affecting the validity of the findings.

For published results, there is an additional ISO 20252 requirement to ensure that the conclusions are adequately supported by the data and that the results and the interpretation of the results are clearly distinguishable.

The governance of published polls in the United States was for a long time overseen by the National Council for Public Polls (NCPP), established in 1969. The NCPP website does not appear to have been updated since issuing its analysis of the final national and state-wide pre-election polls for the 2012 election. According to its president, Evans Witt of Princeton Survey Research Associates, the ‘NCPP has been relatively dormant for a number of years’.

Nonetheless, the NCPP is important both in its own right and for the precedents it established. The mission of the NCPP is to ‘affirm … commitment to standards of disclosure designed to ensure that users of survey results that enter the public domain have an adequate basis for judging the reliability and validity of the results reported’ (National Council for Public Polls, circa 2012). The purpose of the NCCP Code is ‘(not) to pass judgment on the merits of the methods employed in specific surveys. Rather, (its) … purpose is to ensure that pertinent information is disclosed concerning methods that were used so that consumers of surveys may assess studies for themselves’.

NCPP member organisations that undertake polls are required to provide evidence to the Council that they comply with the Code and, if approved, are permitted to state that they comply ‘with the Principles of Disclosure of the National Council on Public Polls’.

The NCPP code specifies the level of disclosure to be met in relation to survey findings publicly released by a member organisation, or subsequently made public; the additional disclosure requirements that arise in response to any specific written requests for additional information; and recommended levels of disclosure in relation to raw data sets and survey questions. There is also a complaints and dispute resolution procedure.

There are three levels of disclosure summarised below:

**Level 1 Disclosure:** All reports of survey findings issued for public release by a member organisation will include the following information:

- Sponsorship of the survey.
- Fieldwork provider (if applicable).
- Dates of interviewing.
- Sampling method employed (for example, random-digit dialled telephone sample, list-based telephone sample, area probability sample, probability mail sample, other probability sample, opt-in internet panel, non-probability convenience sample, use of any oversampling).

---

44 International Standards Organisation
45 Via email correspondence to the Inquiry on 10 February, 2020
- Population that was sampled (for example, general population; registered voters; likely voters; or any specific population group defined by gender, race, age, occupation or any other characteristic).
- Size of the sample that serves as the primary basis of the survey report.
- Size and description of the subsample if the survey report relies primarily on less than the total sample.
- Margin of sampling error (if a probability sample).
- Survey mode (for example, telephone/interviewer, telephone/automated, mail, internet, fax, e-mail).
- Complete wording and ordering of questions mentioned in or upon which the release is based.
- Percentage results of all questions reported.

The NCCP code requires that Members endeavour to have print and broadcast media include the above items in their news stories. The code also makes a distinction between privately commissioned surveys and publicly reported election polls. However, in the event that the results of a privately commissioned poll are made public, the above items should be disclosed.

**Level 2 Disclosure:** In response to any specific written request for additional items pertaining to any survey findings a member organisation has released publicly, the member organisation will additionally release any of the following:

- Estimated coverage of target population.
- Respondent selection procedure (for example, within household), if any.
- Maximum number of attempts to reach respondents.
- Exact wording of introduction (any words preceding the first question).
- Complete wording of questions (per Level 1 disclosure) in any foreign languages in which the survey was conducted.
- Weighted and unweighted size of any subgroup cited in the report.
- Minimum number of completed questions to qualify a completed interview.
- Whether interviewers were paid or unpaid (if an interviewer administered survey mode).
- Details of any incentives or compensation provided to respondents.
- Description of weighting procedures (if any) used to generalise data to the full population.
- Sample dispositions adequate to compute contact, cooperation and response rates.

**Level 3 Disclosure:** The third level of disclosure encourages the following:

- Release raw data sets (ASCII, SPSS, CSV format) for any publicly released survey results (with telephone numbers or other identifying personal information removed).
- Post complete wording, ordering and percentage results of all publicly released survey questions to a publicly available website for a minimum of two weeks.
- Publicly note their compliance with these Principles of Disclosure.
Membership of the NCCP is drawn from news/media outlets, research centres, academia and polling organisations.

The US is also served by the member-based organisation, the American Association of Public Opinion Research (AAPOR). Individual AAPOR members subscribe to the AAPOR Code of Ethics, which sets out the principles and actions expected of all public opinion and survey researchers. Section 3 of the Code deals with Standards for Disclosure (AAPOR, 2015). AAPOR also offers a broad standards regime for research organisations under the auspices of its Transparency Initiative.\(^4\) Officially launched in 2014, the AAPOR Transparency Initiative operates on a voluntary opt-in basis and sets out a series of ‘minimum disclosure’ standards for any survey or poll that is released publicly by an organisation receiving the AAPOR certification of transparency. To date, nearly 100 survey organisations have received this AAPOR certification.

Under the Disclosure Standards, members of the AAPOR Transparency Initiative are obliged to adhere to guidelines set out in the AAPOR Code of Ethics as follows (AAPOR, 2015):

1. ‘Who sponsored the survey and who conducted it. If different from the sponsor, the original sources of funding will also be disclosed.
2. The exact wording and presentation of questions and response options whose results are reported. This includes preceding interviewer or respondent instructions and any preceding questions that might reasonably be expected to influence responses to the reported results.
3. A definition of the population under study and its geographic location.
4. Dates of data collection.
5. A description of the sampling frame(s) and its coverage of the target population, including mention of any segment of the target population that is not covered by the design. This may include, for example, exclusion of Alaska and Hawaii in U.S. surveys; exclusion of specific provinces or rural areas in international surveys; and exclusion of non-panel members in panel surveys. If possible, the estimated size of non-covered segments will be provided. If a size estimate cannot be provided, this will be explained. If no frame or list was utilized, this will be indicated.
6. The name of the sample supplier, if the sampling frame and/or the sample itself was provided by a third party.
7. The methods used to recruit the panel or participants, if the sample was drawn from a pre-recruited panel or pool of respondents.
8. A description of the sample design, giving a clear indication of the method by which the respondents were selected, recruited, intercepted or otherwise contacted or encountered, along with any eligibility requirements and/or oversampling. If quotas were used, the variables defining the quotas will be reported. If a within-household selection procedure was used, this will be described. The description of the sampling frame and sample design will include sufficient detail to determine whether the respondents were selected using probability or non-probability methods.
9. Method(s) and mode(s) used to administer the survey (e.g., CATI, CAPI, ACASI, IVR, mail survey, Web survey) and the language(s) offered.

\(^4\) Survey organisations operating outside of the US are also able to apply for membership of the AAPOR Transparency Initiative.
10. Sample sizes (by sampling frame if more than one was used) and a discussion of the precision of the findings. For probability samples, the estimates of sampling error will be reported, and the discussion will state whether or not the reported margins of sampling error or statistical analyses have been adjusted for the design effect due to weighting, clustering, or other factors. Disclosure requirements for non-probability samples are different because the precision of estimates from such samples is a model-based measure (rather than the average deviation from the population value over all possible samples). Reports of non-probability samples will only provide measures of precision if they are accompanied by a detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. To avoid confusion, it is best to avoid using the term “margin of error” or “margin of sampling error” in conjunction with non-probability samples.

11. A description of how the weights were calculated, including the variables used and the sources of weighting parameters, if weighted estimates are reported.

After survey results are reported, we will make the following items available within 30 days of any request for such materials:

12. Procedures for managing the membership, participation, and attrition of the panel, if a pool, panel, or access panel was used.

13. Methods of interviewer training, supervision, and monitoring, if interviewers were used.

14. Details about screening procedures, including any screening for other surveys that would have made sample members ineligible for the current survey must be disclosed (e.g., in the case of online surveys if a router was used).

15. Any relevant stimuli, such as visual or sensory exhibits or show cards. In the case of surveys conducted via self-administered computer-assisted interviewing, providing the relevant screen shot(s) is optimal, though not required.

16. Details of any strategies used to help gain cooperation (e.g., advance contact, compensation or incentives, refusal conversion contacts) whether for participation in a group, panel or access panel or for participation in a particular research project.

17. Procedures undertaken to ensure data quality, if any. Where applicable, this includes re-contacts to confirm that the interview occurred and/or to verify the respondent’s identity, measures taken to prevent respondents from completing the same survey more than once, and other quality control procedures (e.g., logic checks and tests for speeding and patterning). If no such efforts were undertaken, this will be disclosed.

18. Summaries of the disposition of study-specific sample records so that response rates for probability samples and participation rates for non-probability samples can be computed. If response or cooperation rates are reported, they will be computed according to AAPOR Standard Definitions. If dispositions cannot be provided, the reason(s) will be disclosed and this will be mentioned as a limitation of the study.

19. The unweighted sample size on which one or more reported subgroup estimates are based.

20. Specifications adequate for replication of indices or statistical modelling included in research reports.'

In addition to AAPOR’s review of the performance of the polls in the 2016 US election (AAPOR, 2017), AAPOR has been an active advocate for transparency in election polling as far back as the 2008 pre-election polls in New Hampshire (AAPOR, 2008); as recently as January 2020, AAPOR released a
Inquiry into the performance of the opinion polls at the 2019 Australian federal election

100

AMSRO Inquiry Panel

AAPOR is also active in providing election polling resources for researchers and for media (AAPOR, 2020a).

In the US, media organisations such as CNN, ABC and others detail their own standards in relation to election polling (CNN, 2019, ABC News, 2019). In the case of CNN, they will only publish polls that meet the industry’s best practices.

CNN requires pollsters to answer 16 questions:

1. What survey firm conducted the poll?
2. How were respondents interviewed – by live interviewers on the phone, IVR, online, self-administered questionnaire or another method?
3. Who paid for the survey and why was it done?
4. How many people were interviewed for this survey?
5. In what language(s) were respondents interviewed?
6. Please provide a copy of the full text and interviewer instructions/programming for all questions included in this survey release.
7. When was your survey conducted?
8. What is the source of your sample for this survey, and by what method were respondents selected? Please be as specific as possible, and if via web panel(s), please include a description of how the panellists were recruited. If your study was conducted online and included respondents chosen via routers, approximately what percentage of respondents were directed to the survey via routers?
9. If any quotas were applied to sampling or interviewing, at what stage were they applied, what variables and targets were used, and what is the source of your estimate of the target quota?
10. What is the universe of people you are trying to survey, and what makes you confident that the sample source represents that universe?
11. If surveys were conducted by telephone, what percentage of interviews were conducted via calls to cell phones? If surveys were conducted online, were respondents allowed to complete the survey via mobile browsers, and approximately what share of your respondents did so?
12. If surveys were conducted by telephone, how many callback attempts did a sampled number receive before being retired?
13. If surveys were not conducted by a live interviewer, what do you do to ensure your respondents are real people and are paying attention to the survey?
14. What is your estimate of this survey’s error, how is it calculated, and why is this an appropriate error estimation for your survey? If you are reporting a margin of sampling error, has it been adjusted for design effects?
15. If your survey has been weighted, please list the weighting variables and the source of the weighting parameters. If your survey has not been adjusted for education, please explain why and provide an unweighted frequency for education distribution among your respondents.
16. Is there a minimum unweighted sample size you require before releasing any subset estimates, and if so, what is it?

In addition, CNN announced that they will not report on: polls that are conducted by campaigns or by those who have a financial or advocacy interest in the outcome; polls that are conducted by telephone using robocalls rather than live interviewers; polls that are conducted without any type of sampling, where anyone who chooses to can participate; polls that are conducted solely using unrepresentative sample sources; polls that do not take into account people who take surveys on their cellphones, either over the phone or by web; and polls that do not ensure that respondents of all education levels are adequately reflected.

In Britain, the British Polling Council (BPC), established in 2004, is an association of polling organisations that publish polls. The BPC is modelled on the NCPP. ‘Membership is limited to organisations that can show to the satisfaction of the BPC that the sampling methods and weighting procedures used are designed to accurately represent the views of all people within designated target groups (such as all adults, or voters etc.)’

Through full disclosure, the Council ‘aims to encourage the highest professional standards in public opinion polling and to advance the understanding, among politicians, the media and general public, of how polls are conducted and how to interpret poll results (British Polling Council, 2019). The BPC will also provide interested parties with advice on best practice in the conduct and reporting of polls. As is the case for the NCPP – but in contrast to CNN – the BPC does not pass judgment on the merits of methods employed in specific surveys. Rather, the purpose of the BPC is to ensure that all relevant information is disclosed concerning the methods that were used so that consumers of surveys may assess the studies for themselves.

Nonetheless, the BPC does see fit to ensure that membership is limited to organisations who can show that the sampling methods and weighting procedures used are designed to accurately represent the views of all people within designated target groups (such as all adults, or voters, etc.)

The BPC has an established set of Objects and Rules that member organisations agree to abide by. Failure to do so could see them placed on probation or even expelled. Investigatory powers as they apply to disciplinary hearings are well documented.

In terms of published data, the main elements of disclosure, as they pertain to this Inquiry, are as follows:

- Client commissioning the survey
- Dates of interviewing
- Method of obtaining the interviews (e.g. in-person, telephone, internet)
- The universe effectively represented (all adults, voters etc.)
- The percentages upon which conclusions are based; and
- Size of the sample and geographic coverage.

In addition, whenever it is practical to do so, the following information should also be published:

- Complete wording of questions upon which the release is based
- A web address where full computer tables may be viewed.

Survey organisations reporting results will endeavour to have print and broadcast media include the above items in their news stories, and will in any event make a report containing these items together
with full computer tables of the results available on the survey organisation’s website within two working days of the original release, or make such information available on request.

The research organisation responsible for conducting the survey will place the following information on its own website or provide the information to any interested party on request:

- A full description of the sampling procedures adopted by the organisation
- Computer tables showing the exact questions asked in the order they were asked, all response codes and the weighted and unweighted bases for all demographics and other data that has been published
- A description of the weighting procedures employed, including weighted and unweighted figures for all variables (demographic or otherwise) used to weight the data, whether or not such breakdowns appear in any analysis of sub samples
- An e-mail address for further enquiries. It is assumed that all other reasonable requests for data necessary for readers of the polls to assess the validity of the data will be answered, and a link to the BPC web-site.

In May 2018, and following the Sturgis Inquiry into the 2015 British election polls, the BPC introduced an important new rule that requires its members, when reporting estimates of vote intention, ‘to publish a statement of the level of uncertainty that has historically been associated with polls of voting intention’ (British Polling Council, 2018). Currently, that statement reads: ‘All polls are subject to a wide range of potential sources of error. On the basis of the historical record of the polls at recent general elections, there is a 9 in 10 chance that the true value of a party’s support lies within 4 points of the estimates provided by this poll, and a 2 in 3 chance that they lie within 2 points.’ (http://www.britishpollingcouncil.org/british-polling-council-introduces-new-rule-on-uncertainty-attached-to-polls/, accessed 9 October 2020)

The BPC distinguishes between publicly released polls and privately commissioned surveys, a distinction we think is also useful in the Australian context and it acknowledges that organisations conducting privately commissioned surveys have the right to maintain the confidentiality of survey findings. However, when results of a privately commissioned poll are made public by the organisation that commissioned the survey (including its employees or agents), such results will be deemed to have entered the public domain and procedures outlined above will be followed.

The BPC has also documented a procedure for handling complaints and resolving disputes that member organisations must abide by.

The BPC currently has 23 member organisations, including almost every market research organisation in the UK that publishes political polls. Its management committee and officers are drawn from the member organisations, but its Sub-Committee on Disclosure (which provides the technical advice) comprises representatives from research organisations, academia and the media as well as member organisations. It is worth noting that the BPC accepted and implemented all the recommendations arising from the Report of the Inquiry into the 2015 British general election opinion polls (Sturgis et al., 2016, British Polling Council, 2016).

Also important in the UK is the role of the Market Research Society (MRS). An extract from the report of the House of Lords Select Committee on Political Polling and the Digital Media explains the role of the MRS as follows:

‘(The MRS) promotes, develops, supports and regulates standards and innovation across market, opinion and social research and data analytics. Its standards for the research sector are contained in its Code of Conduct, which covers commissioning and design, client
In terms of transparency, the MRS code requires members to ‘comply with reasonable requests to make available to anyone the information necessary to assess the validity of any published findings from a project’ (Rule 55) and requires that ‘Members must ensure that reports include sufficient information to enable reasonable assessment of the validity of results’ (Rule 60) (Market Research Society, 2019).

There are also a variety of codes in the UK relating to the media’s reporting of polling. The House of Lords Select Committee on Polling and the Digital Media (2018) notes: ‘Together with the BPC and the MRS, the press and broadcaster regulators make up the broad supervisory framework, such as it is, that currently governs polling’ (p. 56).

In Canada, the Canadian Research Insights Council (CRIC) formed in June 2019 out of the Canadian Market Research Association, has 32 members. The CRIC Public Opinion Research and Disclosure standards, as applied to research released into the public domain, are designed to support sound and ethical practices in the disclosure of research:

- Ensure research is unbiased and supports decision-making in the public, private and not-for-profit sectors.
- Enhance public trust and improve the public’s understanding of the use of research.
- Ensure the appropriate transparency and disclosure of research results and methods of studies (Canadian Research Insights Council, 2019).

The CRIC standards adhere to and complement ESOMAR standards:

1. Members will disclose on their website details on how to obtain required disclosures for survey research released into the public domain.
2. In advance of any undertaking, members will inform those who wish to conduct research studies to be released into the public domain, including its use in whole or in part for advertising or other public displays, about the CRIC Public Opinion Research Standards and Disclosure Requirements and our professional obligations.
3. Members will disclose who sponsored the research and who conducted it. If different from the sponsor, the sources of funding will also be disclosed.
4. Members will provide an e-mail address and contact name for further enquiries. All other reasonable requests for data necessary for readers of the research to assess the validity of the research will be answered.
5. Members will embed a clear URL on the release to the CRIC website that links to the CRIC Public Opinion Research Standards and Disclosure Requirements and a statement confirming compliance with the CRIC Standards.
6. Members will disclose the following in a common, readily accessible format on a verifiable corporate or officially sponsored website, and make available to all when a report, results, or both from survey research are released into or appear in the public domain:
   a. The exact wording and presentation of questions and response options. This includes any preceding interviewer or respondent instructions and any preceding questions that might reasonably be expected to influence responses to the reported results. The
percentage of respondents answering who are undecided or refuse to respond on voting questions should be reported.

b. A definition of the population under study.

c. The dates data were collected.

d. A description of the sampling frame(s) and its coverage of the target population.

e. The name of the sample supplier; if the sampling frame and/or the sample itself was provided by a third party, always with an embedded URL to the supplier’s web site home page.

f. Whether the sample was drawn from a pre-recruited panel or pool of respondents and, if applicable, the methods used to recruit the panel or participants.

g. Details of any strategies used to help gain cooperation (e.g., advance contact, compensation or incentives)

h. A description of the sample design, giving a clear indication of the method by which the respondents were selected, recruited, intercepted, or otherwise contacted or encountered, along with any eligibility requirements and/or oversampling. The description of the sampling frame and sample design will include sufficient detail to determine whether the respondents were selected using probability or non-probability methods.

i. The Method(s) and mode(s) used to administer the survey (e.g., CATI, CAPI, ACASI, IVR, mail survey, Web survey) and the language(s) offered.

j. A statistic that expresses the amount of sampling error in a survey’s results—such as a margin of error or a Bayesian credibility interval—if scientifically applicable.

k. Whether weighting/calibration/normalization was used to adjust the results and the impact of the procedure on the data as measured through a weighting efficiency calculation and/or disclosure of the range (i.e., min and max) and variance of weights.

l. For research released on public policy topics or election voting, the detailed tables or representations thereof by standard demographic questions with weighted and unweighted number of respondents so that the public can transparently see the original number of responses by standard demographic category and the weighted number of responses by standard demographic category. This data must be made available upon request for research released on other topics.

m. For research released in to the public domain with a media partner, the required disclosures should be made available within 90 minutes of the release of the results by the media partner (Ibid).

Before the CRIC was established, the Market Research and Intelligence Association, established in 2004, represented opinion pollsters and market researchers in Canada, and had developed a similar code of conduct.

In New Zealand, the Research Association of New Zealand (RANZ), a member of ESOMAR, established the New Zealand Political Polling Code effective January 2014.

The RANZ code closely follows the main elements of the ESOMAR code. The RANZ code details ‘best practice’ for the researcher conducting the poll, ‘best practice’ for the researcher in reporting
Inquiry into the performance of the opinion polls at the 2019 Australian federal election
AMSRO Inquiry Panel

results, and ‘best practice’ for the media in publishing the results. The code, binding on member companies, applies only to ‘political polls’ (Research Association of New Zealand, 2014).

It is intended that the code assist politicians, political scientists, journalists and members of the public to be confident that political polls represent the opinions of the wider public, and are a guide as to likely voting behaviours.

The rationale provided by RANZ for the establishment of the code is summarised as follows:

- The development of the code is in recognition of the fact that reporting of polls can have an impact on how people vote
- Inaccurate polls or polls that are reported inaccurately can impact on voting attitudes and behaviours and thus influence the democratic process
- It behoves all members of the polling and media communities to treat polling responsibly. Reliable polls, rather than informal surveys, require a high degree of rigour. These guidelines are designed to ensure that rigour is understood and applied.

The main elements of this code are as follows:

- The code documents best practice guidelines for the conducting and reporting of political polls in New Zealand
- The code is binding on companies that are members of Research Association New Zealand and on researchers that are members of the Research Association New Zealand
- The code only covers ‘political polls’ that, for the purpose of the code, are polls that related to public votes such as national elections, local body elections and parliamentary referenda. This is in recognition of the fact that reporting of polls may have an impact on how people vote.

The code sets out a number of recommended and mandatory requirements under the headings of sampling, collection method, weighting, margin of error, timing and results. Whereas most of the codes reviewed for this paper focus mainly on prescribing disclosure requirements, the RANZ code is somewhat more prescriptive in terms of survey methodology and statistical techniques.

- **Sampling:**
  - The report must include the sample size, and the sample size of “decided” voters
  - The report must disclose the sampling method
  - The report should disclose that multiple call-backs occurred
  - The report should include the response rate
  - The report should disclose the population the sample represents
  - The report must exclude those unlikely to vote from the analysis of voting behaviour
  - The report should include a definition of how it was determined someone was likely to vote.

- **Collection method:**
  - Phone: The report must disclose how a respondent is selected. The report should disclose if calls were to landlines only, and any impact this may have had on the poll.
Online: The report should disclose panel recruitment and makeup, and that it complies with the ESOMAR guideline for online research. The report should disclose if there were any major platforms that the poll was not accessible on.

Omnibus: The report must disclose if the questions were part of an omnibus survey.

Question order: The report must disclose the order of questions asked and any political questions asked before the principal voting behaviour question.

- Weighting:
  - Weighting method: The report should confirm the sample was weighted.
  - Weighting variables: The report should disclose the variables the poll is weighted on.

- Margin of error: The report must disclose the maximum margin of error. The report should disclose the sample size and maximum margin of error for demographic breakdowns.

- Statistical significance: The report should highlight results that are statistically significant. This includes trend changes, not just from the previous poll.

- Timing:
  - The report must disclose the dates the data collection occurred. The median date of collection should be included in the report.

- Reporting:
  - Undecided: The report must state the number and percentage of undecideds and refusals.
  - The agency should prepare a report suitable for publication with full results and methodology. The report should include the wording of the voting questions.

Table 25 summarises the transparency/disclosure standards that would apply in Australia if the ISO standards and The Research Society Code of Professional Behaviour (including the associated reporting guidelines) and APC guidelines were generally applicable, and also sets out the disclosure standards provided in selected international codes (ISO and ESOMAR/WAPOR) and in the US, Britain, Canada and New Zealand.

Whilst summary in nature and somewhat dependent upon how sometimes general provisions in these codes are interpreted, viewing the various disclosure provisions of the codes in this way provides a reasonable starting point for comparing the transparency and disclosure standards notionally in place for election polling in Australia with other national and international models, as well as for judging the adequacy of the current standards regime in Australia.
## Table 25: Comparison of selected international, national and Australian disclosure standards for election polling

<table>
<thead>
<tr>
<th>Disclosure requirements</th>
<th>ISO</th>
<th>ESOMAR/WAPOR</th>
<th>NCPP</th>
<th>AAPOR (TI)</th>
<th>CNN</th>
<th>BPC(f)</th>
<th>CRIC</th>
<th>RANZ</th>
<th>Australia (RS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research organisation/funder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Name of research organisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Research funder/sponsor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Identification of subcontracted services (including fieldwork provider)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Contact details for further queries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td><strong>Population/Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>In-scope population (may include limiting to voters / likely voters)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Estimated coverage of the target population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Procedures for managing the membership, participation, and attrition of the panel, if a pool, panel, or access panel was used.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Achieved sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Sample size for subset estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Geographic coverage (a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Sampling method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>- Description of sampling frame and its coverage of the target population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>- Name of sample supplier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>- Recruitment/respondent selection methods (e.g. if sample was drawn from a pre-recruited panel). Within household selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>- A description of the sample design (including any quotas)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>- Any screening for respondents. E.g. in the case of an online survey if a router was used.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Six-month prohibition of including the same online panel member in the same poll</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Minimum 72-hour fieldwork period for online panel surveys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td><strong>Fieldwork</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Fieldwork dates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Methods of interviewer training, if applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Paid or unpaid interviewees used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>In what languages were respondents interviewed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Number of attempts to reach a respondent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Use of respondent incentives &amp; other strategies to gain cooperation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
<tr>
<td>Online panels should be managed in accordance with ESOMAR guidelines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APS/Research Society (RS)</td>
</tr>
</tbody>
</table>
Table 25 (cont.): Comparison of selected international, national and Australian disclosure standards for election polling

<table>
<thead>
<tr>
<th>Disclosure requirements</th>
<th>ISO</th>
<th>ESOMAR/WAPOR</th>
<th>NCPP</th>
<th>AAPOR (TI)</th>
<th>CNN</th>
<th>BPC(+)</th>
<th>CRIC</th>
<th>RANZ</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>Methods of data collection (including survey mode)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>- Proportion of telephone interviews completed on a mobile phone / Proportion of online surveys completed on a mobile phone or small screen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>- If not undertaken by a live interviewer, how do you ensure the respondent is a real person and paying attention to the survey?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>Response rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Response rates (and sample dispositions) for probability samples (or a similar statistic for non-probability samples)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Minimum number of completed questions to qualify as a completed interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Survey questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APC/RS</td>
</tr>
<tr>
<td>The exact wording of the relevant questions and the order in which they are asked</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APC/RS</td>
</tr>
<tr>
<td>Full questionnaire including interviewer and programming instructions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APC/RS</td>
</tr>
<tr>
<td>Whether the survey questions were asked as part of an omnibus survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Any non-English language versions of the questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Any visual stimuli, show cards, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Adjustments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Whether weighting was used and the universe used for the weights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>- A full description of sampling and weighting procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>- Has the survey been adjusted for education of respondents? If not, why not?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>- Include unweighted bases for any subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
</tbody>
</table>

Available upon release
Available after release/On website/Upon request
<table>
<thead>
<tr>
<th>Disclosure requirements</th>
<th>ISO</th>
<th>ESOMAR/WAPOR</th>
<th>NCPP</th>
<th>AAPOR (TI)</th>
<th>CNN</th>
<th>BPC</th>
<th>CRIC</th>
<th>RANZ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis and reporting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods of statistical analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/RSS</td>
</tr>
<tr>
<td>Margin of sampling error (or similar). Has margin of error been adjusted for design effects?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>A verifiable claim about the record of accuracy of the poll</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/RSS</td>
</tr>
<tr>
<td>Comment on the statistical significance of the findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/RSS</td>
</tr>
<tr>
<td>Specifications adequate for replication of indices or statistical modelling included in research reports.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>A statement of substantial limitations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/RSS</td>
</tr>
<tr>
<td>Percentage results for all questions reported (incl. refused and undecided for voting intentions)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/RSS</td>
</tr>
<tr>
<td>Complete computer tables showing the exact questions asked in the order they were asked, all response codes and the weighted and unweighted bases for all demographics and other data that has been published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data quality checks (including checks to ensure surveys completed by 'real' people)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ISO/APC/RS</td>
</tr>
<tr>
<td>Statement of adherence to disclosure standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not make claims that exceed the scientific limits of the poll</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) Presumably also covered by describing the in-scope sample

b) There are additional ISO requirements to disclose considerable details about sampling and methodological detail to clients but no such requirements for public disclosure

c) Membership to the BPC is restricted to those organisations that use sampling methods and/or weighting procedures to broadly represent the opinions of all peoples in designated groups
Appendix 9: Source materials – primary reporting of the polls

The table below contains hyperlinks to the primary articles and secondary materials used as a basis for the analysis in section 7, which examined the disclosure of basic information about the polls, in accordance with Australian Press Council guidelines. Some of these articles are behind paywalls.

Table 26: Source materials used to analysis the disclosure standards of the primary reporting of the polls.

<table>
<thead>
<tr>
<th>Polling company and date of release</th>
<th>Source article</th>
<th>Supplementary materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Source</td>
<td>Link</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Appendix 10: Recommended disclosure standards for Australian election polls

The minimum set of disclosure standards put forward for the consideration of AMSRO and other stakeholders are outlined below. Of course, not all of these can be included in a press release, news article or online story. Other national and international codes get around this by requiring that such information be made available via the pollster's and/or media outlet's website, or available upon request.

The Inquiry Panel believes that disclosure of the items listed below would enable consumers of the polls to make an informed decision as to the quality of the methods used to conduct a poll and produce the findings. We appreciate that there are also commercial and practical considerations that need to be borne in mind but we hope that these recommendations can serve as a blueprint for any disclosure regime that might emerge in response to this Inquiry.

In terms of the burden on pollsters and publishers in making these paradata available, the vast majority of the specified items (excluding items 10, 12, 17 and 23) are of a generic/static nature and, once documented, would not have to be updated for each poll.

Our recommendation is that Items 1-19 be regarded as mandatory and be available upon the release of each poll.

1. Declaration of any conflicts of interest (e.g. undertaking polling for a political party, interest group, etc.) (New)
2. Name of the research organisation undertaking the poll/survey (ISO/APC).
3. Name of the organisation funding the poll/survey (ISO/APC).
4. Name of subcontractors used for fieldwork and/or to provide the sample (e.g. sample provider and/or online panel provider) (ISO expanded).
5. The in-scope population (e.g. eligible voters) (ISO/APC).
6. Geographic scope for the poll (e.g. National, State, seat, other) (ISO/APC).
7. Description of the sampling frame and its coverage of the population (RS).
8. In which language/s the poll was conducted/available in (RS).
9. Methods used to recruit respondents: for online, specify whether an opt-in or probability panel; for CATI surveys, specify whether landline and/or mobile telephone numbers were used and/or whether or not such numbers were random-digit dialled telephone sample, list-based telephone sample or both (RS).
10. Achieved sample size (APC/RS).
11. Sample design including quotas and stratification scheme, and including other elements of sample design as applicable/appropriate (e.g. cluster or multi-stage design) (RS).
13. Methods and mode of data collection (e.g. online/CATI/IVR, Mixed mode, other) (ISO/APC/RS).
14. The exact wording of relevant questions (APC/RS).
15. Whether the raw data has been weighted and, if so, the weighting benchmarks used. The general weighting variables should be described but proprietary algorithms and specific weighting variables do not need to be disclosed. Pollsters should ‘divulge whether their standard weighting methodology was adjusted in any way at all for [their current] poll. See item 22 for expanded requirements (RS, expanded).

16. How the two-party-preferred estimate was derived (New requirement – for election polling only).

17. A description of a poll’s margin of error and how it has been calculated (ISO/APC/RS).

18. Quantifying the proportion of unknowns, unstated and undecided respondents for key questions (New for Australia but required by other codes).

19. Membership and accreditation of key professionals with the relevant professional bodies (e.g. QPR or AStat) (The Research Society, 2020b, Statistical Society of Australia, 2019) or any new entity such as the proposed Australian Polling Council. In addition to the above, the Inquiry Panel believes that Items 20-24 should be available to interested parties upon request (perhaps following the agreement of Non-disclosure provisions):

20. Whether respondent incentives were used (ISO/APC/RS).

21. The full questionnaire, or questionnaire module in the case of omnibus surveys (RS).

22. A detailed description of the weighting techniques used (e.g. post-stratification, raking, model-based, etc.) and the variables and benchmarks used for the weighting parameters.

23. Provision of response rates or similar statistics including description of how these metrics have been calculated. The Inquiry’s preference is that a uniform means for calculating response metrics is used with a strong preference for the AAPOR standards to be adopted (RS, expanded).

24. The description of fieldwork effort including the number of attempts made to contact each sample member (NCPP, CNN, CRIC, RS).

---

47Qualified Professional Researcher (QPR) is an accreditation available to researchers, data analysts, or insights practitioners who are members of the AMSRS. AStat – an Accredited Statistician – is the accreditation available to members of the Statistical Society of Australia.
References


Australian Bureau of Statistics, 2019. 4364.0.55.001 – National Health Survey: First Results, 2017-18, ABS, Belconnen, ACT.
Inquiry into the performance of the opinion polls at the 2019 Australian federal election

AMSRO Inquiry Panel

Australian Bureau of Statistics, 2018. 8146.0 – Household Use of Information Technology, Australia, 2016-17, ABS, Belconnen, ACT.

Australian Bureau of Statistics, 2016. Census of Population and Housing, General Community Profile, Catalogue no. 2001.0. G13 Language Spoken at Home (a) by Proficiency in Spoken English/Language by Sex, viewed on 18 July 2020


Australian Election Study, 2019, viewed on 9 July 2020 https://australianelectionstudy.org/


Biddle, N., 2019. ‘Predicting the unpredicted: what longitudinal data can tell us about the 2019 Australian federal election’. Centre for Research and Methods, CSRM Working Paper No. 6/2019,


ESOMAR, 2017. Global market research 2017: an ESOMAR industry report. ESOMAR, Amsterdam, Netherlands

ESOMAR/WAPOR, 2014a. ‘Key Requirements for Opinion Polls and Published Surveys’, viewed 5 August 2020

ESOMAR/WAPOR, ESOMAR /WAPOR, 2014b. ‘Guideline on Opinion Polls and Published Surveys’, viewed 5 August 2020

Fahimi, M., Barlas, F.M., Thomas, R.K. and Buttermore, N., 2015. ‘Scientific surveys based on incomplete sampling frames and high rates of nonresponse’. Survey Practice, 8(60), pp.1-12

Franklin, C., 2007. ‘The ‘Margin of Error’ for Differences in Polls’ viewed on 16 September 2020
https://abcnews.go.com/images/PollingUnit/MOEFranklin.pdf


Goot, M. 2019. ‘Did late deciders confound the polls?’ Inside Story, 19 September, viewed on 30 June


http://hdl.handle.net/1959.14/315551.


House of Lords, 2018. House of Lords Select Committee on Political Polling and the Digital Media, viewed 22 January 2020


Ipsos, 2019a, viewed 6 May 2020

Ipsos, 2019b, viewed on 10 January 2020


Inquiry into the performance of the opinion polls at the 2019 Australian federal election


Inquiry into the performance of the opinion polls at the 2019 Australian federal election
AMSRO Inquiry Panel

Murphy, K., 2019a. ‘Essential poll: majority of voters think Bill Shorten will be the winner on Saturday’, viewed on 6 May 2020 https://www.theguardian.com/australia-news/2019/may/16/essential-poll-majority-of-voters-think-bill-shorten-will-be-the-winner-on-saturday

Murphy, K., 2019b. ‘Post-election research shows 11% of voters made up their mind on polling day’, viewed 13 July 2020 https://www.theguardian.com/australia-news/2019/jun/04/post-election-research-shows-11-of-voters-made-up-their-mind-on-polling-day


https://www.statsoc.org.au/Professional-accreditation


Wells, A. 2019. ‘False recall, and how it affects polling’ YouGov, viewed on 2 September 2020
https://yougov.co.uk/topics/politics/articles-reports/2019/07/17/false-recall-and-how-it-affects-polling

