

# The Impact of Using Alternative Measures of Past Vote to Adjust Estimates of Voting Intentions: A Case Study Using Australian Polling Data



Darren Pennay, Sebastian Mission, Dina Neiger and Paul J. Lavrakas  
77<sup>th</sup> Annual Conference of the American Association for Public Opinion Research,  
Chicago, May 11-13, 2022

A subsidiary of:

# Authors/Presenter

## Darren Pennay

- Founder and former CEO, The Social Research Centre Pty Ltd
- Honorary Professor in the Practice of Survey Methodology
  - ANU Centre for Social Research and Methods
- Adjunct Professor, Institute for Social Science Research
  - University of Queensland
- Chair and lead author of the Inquiry into the 2019 pre-election polls in Australia

## Sebastian Misson

- Senior Survey Statistician with the Social Research Centre

## Dr Dina Neiger

- Chief Statistician with the Social Research Centre
- Visitor at the ANU Centre for Social Research and Methods.

## Dr Paul J Lavrakas (Presenter)

- Independent Consultant
- Senior Methodological Adviser with the Social Research Centre
- Adjunct Professor with the University of Illinois at Chicago
- International Adviser to the Inquiry into the 2019 pre-election polls in Australia

# Background (1)

- In 2019 Australia joined the rank of countries with a high-profile “polling failure” (as defined by Durand and Blais, 2020, pp. 134-135)
- All 16 national pre-election polls published during the campaign estimated that Labor (the Opposition party) had the support of a majority of voters (ave. of polls showed Labor at 51.5% vs. Coalition at 48.5%)
  - Most commentators declared there would be a change of government
- The result was the exact statistical opposite
  - The Coalition government was returned to office with 51.5% of the vote compared to 48.5% for Labor
- This was the biggest polling miss in Australia in decades!

## Background (2)

- The Association of Market and Social Research Organisations and the Statistical Society of Australia launched a wide-ranging inquiry into the performance of the pre-election polls ([link](#))
- One of the many recommendations from the Inquiry was that pollsters develop more sophisticated sampling balancing and/or weighting strategies
  - Weighting by educational attainment and past vote were identified as possible candidates
- This presentation summarises findings from a subsequent study that used Australian pre-election polling data to measure the impact of weighting estimates of voting intentions by educational attainment and **past vote** (article available [here](#)).

# Weighting by Past Vote? (1)

- Good benchmarks available, correlated with non-response and voting intentions, and widely used:
  - Improving voting intention estimates by balancing or weighting one's sample by "previous vote" is common in electoral polls, particularly in Europe (Durand et al. 2015, p. 1)
  - YouGov's UK Director of Political and Social Research, Anthony Wells, observed that "almost all [pollsters in the UK] ... use how people voted at the last election as a target when designing or weighting ... polling samples" (Wells 2019, p. 2)
  - Among the 438 state-level USA presidential polls conducted in the last two weeks [of the campaign], 46 reported weighting on partisanship, 22 reported weighting on past 2016 vote, 249 reported a weighting scheme that used neither, and it was unclear as to how the sample was weighted for 121 polls (AAPOR Taskforce Report on 2020 Pre-Election Polling, p. 69).

# Weighting by Past Vote? (2)

However ...

- Recall of past vote may be inaccurate for various reasons:
  - (1) memory failure; (2) the tendency of voters to 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)
  - The jury is out as to its usefulness
    - In an examination of 12 election results Durand et al. found that a past vote weighting correction resulted in “little difference between corrected and uncorrected estimates of voting intentions” (Durand et al. 2015, p.12)
    - AAPOR's 2020 Taskforce Report found that 'Weighting on partisanship and past vote may have reduced polling error, but it did not solve the issue ' (AAPOR 2021, p. 69)
    - By way of contrast, a 2013 report by Cooper on behalf of the British Polling Council found that “past vote weighting can and often does make a significant difference to voting intention numbers” (Cooper 2013, p. 2).

# Hypotheses

If past voting behaviour is associated with non-response and the variable of interest (voting intentions) then ...

- **Hypothesis 1** – Weighting to benchmarks for past vote, using a short-term recall measure of past vote, will reduce bias and add an acceptable amount of variance compared to weighting solutions that do not include any adjustment for past vote

If a short-term recall measure of past vote is less affected by measurement error than a long-term recall measure then ...

- **Hypothesis 2** - Weighting to past vote benchmarks, using a short-term recall measure of past vote, will reduce bias and add an acceptable amount of variance *compared to weighting adjustments using a long-term recall measure of past vote.*

# Data for this Study

- A poll conducted by the Australian National University in April 2019
- The sample for the survey was drawn from the Social Research Centre's *Life in Australia*<sup>TM</sup> national probability-based online panel
- Sample size n= 2,205; Interview length = 12mins; Within-panel Completion rate= 76.5%; Cumulative response rate (CUMRR2)=8.6%
- Question of interest: *If a federal election for the House of Representatives was held today, which one of the following parties would you vote for ... ?*
- Data collection 6-26 April, 2019; Election 18 May (with early voting allowed starting 29 April).



## Measures of Past Vote

- **Short-term recall of past vote:** When recruited in August/September 2016 *Life in Australia*<sup>TM</sup>, panellists were asked about their vote choice in the preceding federal election held on 2 July 2016. These responses were appended to our survey dataset
- **Long-term recall of past vote:** A Comparative Study of Election Systems survey conducted on *Life in Australia*<sup>TM</sup> in June 2019 and asked respondents ... "In the last Federal election in July 2016, when the Liberals were led by Malcolm Turnbull and Labor by Bill Shorten, which party got your first preference then in the House of Representatives election?"; the responses were also appended to our survey dataset.
- Our analysis is limited to the 1,684 panellists for whom we had both a short-term and long-term recall measure of past vote (1,684 out of 2,205).

# Methods (1)

- **Weighting** - 15 weighting solutions created
  - Age x sex x geography - baseline weight (1)
  - Age x sex x geography in combination with **educational attainment** (2 and 3)
  - Weights 1 to 3 with a **short-term recall measure** of past vote (4, 6 and 8)
  - Weights 1 to 3 with a **long-term recall measure** of past vote (5, 7 and 9)
  - Short-term recall and long-term recall on their own (10 and 11)
  - Weights 1 – 3 with a **blended recall measure** (50:50 short-term : long-term) (12, 13 and 14)
  - Blended recall measure on its own (15)

Weights were calculated using the *rake* procedure from the survey package in R (Lumley 2020; 2010; 2004).

# Methods (2)

## Error metrics

### ➤ Measures of bias:

- Weighted average absolute error on the primary vote
- Average absolute error on the two-party preferred vote (2PP)

### ➤ **Measure of variance:** The variance introduced by the weights was measured using the design effect (*deff*) calculated by Taylor series linearisation by the *svymean* procedure in the survey package in R (Lumley 2020).

### ➤ **Overall measure:** Mean square error (MSE) (Korn & Graubard 1999) is a measure which combines bias and variance to assess the impact of weighting on the total survey error.

# Selected Results (1)

Wt	Description	Two-party preferred Ave Absolute error (pp)	Design effect	Root Mean square error (pp)
1	Weighted by age, sex, geography (Baseline weight)	5.15 (5.15, 2.67 – 7.64)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)

## Notes:

- Results outside brackets represent the observed estimate based on the original data
- Results in brackets the average estimate from the 10,000 simulations used to produce standard errors (not shown) and confidence intervals
- Significance testing based on simulations: The proportion that one weighting scheme produces superior results to another based on two-tailed probabilities.

## Selected Results (2)

Wt	Description	Two-party preferred Ave. Absolute error (pp)	Design effect	Root Mean square error (pp)	RMSE Sig testing
1	Weighted by age, sex, geography	5.15 (5.15, 2.67 – 7.64)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)	
3	Age by education, sex, geography	4.08 (4.08, 1.46 – 6.71)	1.41 (1.42, 1.24 – 1.59)	4.29 (4.33, 1.90 – 6.76)	NS However, in 97% of the simulations weight 3 produced a less biased estimate than weight 1

Results outside brackets represent the observed estimate based on the original data, results in brackets the average estimate from the simulated samples and the upper and lower confidence intervals.

## Selected Results (3)

Wt	Description	Two-party preferred Ave. Absolute error (pp)	Design effect	Root Mean square error (pp)	RMSE Sig testing
1	Weighted by age, sex, geography	5.15 (5.15, 2.67 – 7.64)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)	
3	Age by education, sex, geography	4.08 (4.08, 1.46 – 6.71)	1.41 (1.42, 1.24 – 1.59)	4.29 (4.33, 1.90 – 6.76)	
8	Age by education, sex, geography and short-term recall	2.41 (2.39, 0.12 – 4.67)	1.14 (1.14, 0.88 – 1.41)	2.70 (2.76, 0.84 – 4.68)	<b>Wt 1</b>

- Significantly less bias
- Significantly more bias

Results outside brackets represent the observed estimate based on the original data, results in brackets the average estimate from the simulated samples and the upper and lower confidence intervals.

## Selected Results (4)

Wt	Description	Two-party preferred Ave. Absolute error (pp)	Design effect	Root Mean square error (pp)	RMSE Sig testing
1	Weighted by age, sex, geography	5.15 (5.15, 2.67 – 7.64)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)	
3	Age by education, sex, geography	4.08 (4.08, 1.46 – 6.71)	1.41 (1.42, 1.24 – 1.59)	4.29 (4.33, 1.90 – 6.76)	
8	Age by education, sex, geography and short-term recall	2.41 (2.39, 0.12 – 4.67)	1.14 (1.14, 0.88 – 1.41)	2.70 (2.76, 0.84 – 4.68)	<b>Wt1, Wt9</b>
9	Age by education, sex, geography and long-term recall	4.45 (4.44, 2.35 – 6.52)	0.90 (0.90, 0.68 – 1.12)	4.57 (4.57, 2.57 – 6.58)	<b>Wt 8</b>

Results outside brackets represent the observed estimate based on the original data, results in brackets the average estimate from the simulated samples and the upper and lower confidence intervals.

# Selected Results (5)

Wt	Description	Two-party preferred Ave. Absolute error (pp)	Design effect	Root Mean square error (pp)	RMSE Sig testing
1	Weighted by age, sex, geography	5.15 (5.15, 2.67 – 7.64)	1.26 (1.26, 1.10 – 1.41)	5.31 (5.32, 2.93 – 7.71)	
3	Age by education, sex, geography	4.08 (4.08, 1.46 – 6.71)	1.41 (1.42, 1.24 – 1.59)	4.29 (4.33, 1.90 – 6.76)	
8	Age by education, sex, geography and short-term recall	2.41 (2.39, 0.12 – 4.67)	1.14 (1.14, 0.88 – 1.41)	2.70 (2.76, 0.84 – 4.68)	<b>Wt1, Wt9, Wt14</b>
9	Age by education, sex, geography and long term-recall	4.45 (4.44, 2.35 – 6.52)	0.90 (0.90, 0.68 – 1.12)	4.57 (4.57, 2.57 – 6.58)	<b>Wt8</b>
14	Age by education, sex, geography and blended estimate of past vote term	3.92 (3.89, 1.64 – 6.15)	1.04 (1.05, 0.79 – 1.30)	4.09 (4.08, 1.96 – 6.20)	<b>Wt8</b>

Results outside brackets represent the observed estimate based on the original data, results in brackets the average estimate from the simulated samples and the upper and lower confidence intervals.



# Findings (1)

1. Adding a short-term recall of past vote adjustment to the various weighting solutions results in better estimates than solutions that do not include any past vote adjustment (Hypothesis 1 is supported)
2. Adding a short-term recall of past vote adjustment to the various weighting solutions results in less biased estimates than when using the long-term recall measure (Hypothesis 2 is supported)
3. Weight 14 which uses a 50% short-term / 50% long-term recall measure of past vote leads to some bias reduction but does not result in a statistically significant reduction in bias compared to the baseline weight.

## Findings (2)

- Our results reflect those of Wells (2019) in that “how or when the (past vote) data was [were] collected makes a difference”
- Weighting by past vote is not a panacea but under the right conditions can result in substantial bias reduction.

# Discussion (1)

- Would commercial pollsters who mainly use non-probability online panels be able to successfully collect and maintain a short-term recall measure of past vote for a sufficient proportion of their panellists for this to be a viable option?
  - A special effort to minimise attrition for a portion of their panel would be required
  - If not already doing so, panel proprietors could routinely collect voting behaviour at the previous election as a profiling variable when recruiting new panellists and update this measure for all panellists following each election.

## Discussion (2)

What impact might weighting by past vote have on other survey estimates, in particular, estimates that could be expected to be strongly aligned with vote choice?

- Peytchev et al. looked at this issue using 15 variables from the 2012 USA General Social Survey. They found the changes were “generally small but three of the fifteen estimates are significantly different. The largest change is [4 pp].” (Peytchev et al. 2019, p. 499–500)
- A USA study by Pew re-balanced the proportion of Republicans and Democrats in the ATP to match 2020 election benchmarks and found most estimates of other attitudes and opinions hardly changed (most < 1 pp, Range 0.5 pp to 3.0 pp) (Pew Research Center, March, 2021).

# Limitations

- This research was conducted on the Life in Australia™ panel which, in retrospect, produced a remarkably unbiased unweighted estimate of voting intentions relative to the election outcome (Ave. Absolute Error on unweighted two-party preferred vote estimate just 0.78 pp)
- The long-term recall of past vote measure that we had available to us for this study left a bit to be desired (i.e. Collected after rather than prior to the subsequent election)
- Even under the right conditions, incorporating a short-term recall measure of past vote into weighting solutions may not overcome one of the problems referred to in the 2020 AAPOR Taskforce Report - that the Republicans represented in the polls may have been less Conservative/more likely to vote Democrat than non-represented Republicans.

# Acknowledgements

- Social Research Centre and ANU Centre for Social Research and Methods for supporting this research
- ANU Centre for Social Research and Methods for allowing their data to be used for this research
- *Life in Australia*<sup>TM</sup> panellists.

# References

- American Association of Public Opinion Research, 2021. AAPOR Taskforce Report on 2020 Pre-Election Polling: 'An Evaluation of 2020 Election Polls in The United States'.
- Durand, C. and Blais, A., 2020. 'Quebec 2018: A Failure of the Polls?' *Canadian Journal of Political Science*, 53(1), pp. 133-150.
- Durand, C., Deslauriers, M. & Valois I., 2015. 'Should Recall of Previous Votes Be Used to Adjust Estimates of Voting Intention? Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach. Retrieved from <https://surveyinsights.org/?p=3543>
- Lumley T (2020). Survey: analysis of complex survey samples. R package version 4.0
- Pennay, D., Misson, S. & Neiger, D. (2021). *The impact of weighting by educational attainment and past vote on estimates of pre-election voting intentions: A case study using Australian polling data*. CSRM and SRC Methods Paper 2/2021, Australian National University
- Pennay, D, Goot, M. Neiger, D. Trewin, D. Lavrakas, P. J., Stirton, J., Hughes, P Sheppard, J. McAllister, I. (2020) *Report of the Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election*, Association of Market and Social Research Organisations and the Statistical Society of Australia. ISBN 978-0-646-82919-7
- Pew Research Center, March, 2021 "What 2020's Election Poll Errors Tell Us About the Accuracy of Issue Polling"
- Peytchev A, Presser S & Zhang M (2018). Improving traditional nonresponse bias adjustments: combining statistical properties with social theory. *Journal of Survey Statistics and Methodolog* 6(4):491–515, <https://doi.org/10.1093/jssam/smx035>
- Wells A (2019). False recall, and how it affects polling, YouGov, <https://yougov.co.uk/topics/politics/articlesreports/2019/07/17/false-recall-and-how-it-affects-polling> Retrieved on 11 January, 2021.

# Thank you

[darren.pennay@anu.edu.au](mailto:darren.pennay@anu.edu.au)



PO Box 13328  
Law Courts Victoria  
Australia 8010



+61 3 9236 8500

A subsidiary of:



Australian  
National  
University