NCRM National Centre for Research Methods

The Inquiry into the 2015 pre-election polls: preliminary findings and conclusions

Royal Statistical Society, London

19 January 2016

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

UNIVERSITY OF Southampton

MANCHESTER 1824 The University of Manchester

THE UNIVERSITY OF EDINBURGH

Inquiry Panel

Dr. Nick Baker, Group CEO, Quadrangle Research Group Ltd

Dr. Mario Callegaro, Senior Survey Research Scientist, Google UK

Dr. Stephen Fisher, Associate Professor of Political Sociology, University of Oxford

Dr. Jouni Kuha, Associate Professor of Statistics, London School of Economics

Prof. Jane Green, Professor of Political Science, University of Manchester

Prof. Will Jennings, Professor of Political Science and Public Policy, University of Southampton

Dr Ben Lauderdale, Associate Professor in Research Methodology, London School of Economics

Dr. Patten Smith, Research Director, Research Methods Centre, Ipsos MORI and Chair of the Social Research Association

Scope of today's meeting

- This is a methodological inquiry
- Report to be published in March – today we are setting out preliminary findings & conclusions
- Not making specific recommendations today

NCRM National Centre for Research Methods

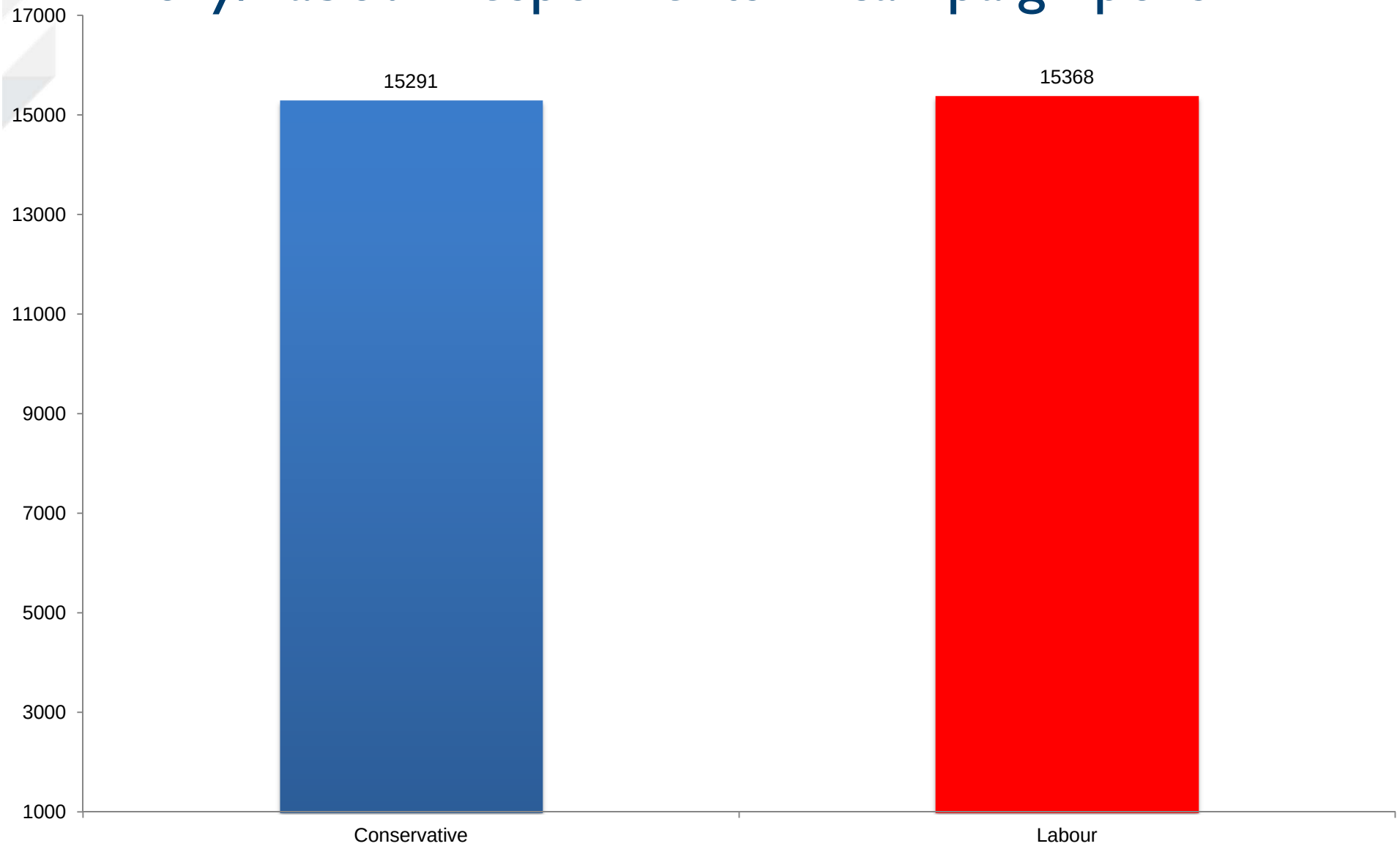
Historical context – how bad was the miss?


7 May
22:0
PARTY BU
ARE CURRENT
CON
316
BBC EXIT POL
CONSERVATI
LARGEST PA
BBC
PHIL COOMES/BBC

The final polls

	Published	Fieldwork	Sample	Con	Lab	Lib	Ukip	Green	Other
BPC Members									
Populus		5–6 May	3917	34	34	9	13	5	6
Ipsos MORI	Evening Standard	5–6 May	1186	36	35	8	11	5	5
YouGov	The Sun	4–6 May	10307	34	34	10	12	4	6
ComRes	Daily Mail, ITV News	5–6 May	1007	35	34	9	12	4	6
Survation	Daily Mirror	4–6 May	4088	31	31	10	16	5	7
ICM	The Guardian	3–6 May	2023	34	35	9	11	4	7
Panelbase		1–6 May	3019	31	33	8	16	5	7
Opinium		4–5 May	2960	35	34	8	12	6	5
TNS		30 Apr–4 May	1185	33	32	8	14	6	6
Non-BPC									
Lord Ashcroft		5–6 May	3028	33	33	10	11	6	8
BMG	May2015.com	3–5 May	1009	33.7	33.7	10.4	12	4	6
Result				37.8	31.2	8.1	12.9	3.8	6.3
Average MAE (=1.8)				*4.2*	*2.4*	*1.0*	*1.5*	*1.1*	*0.7*

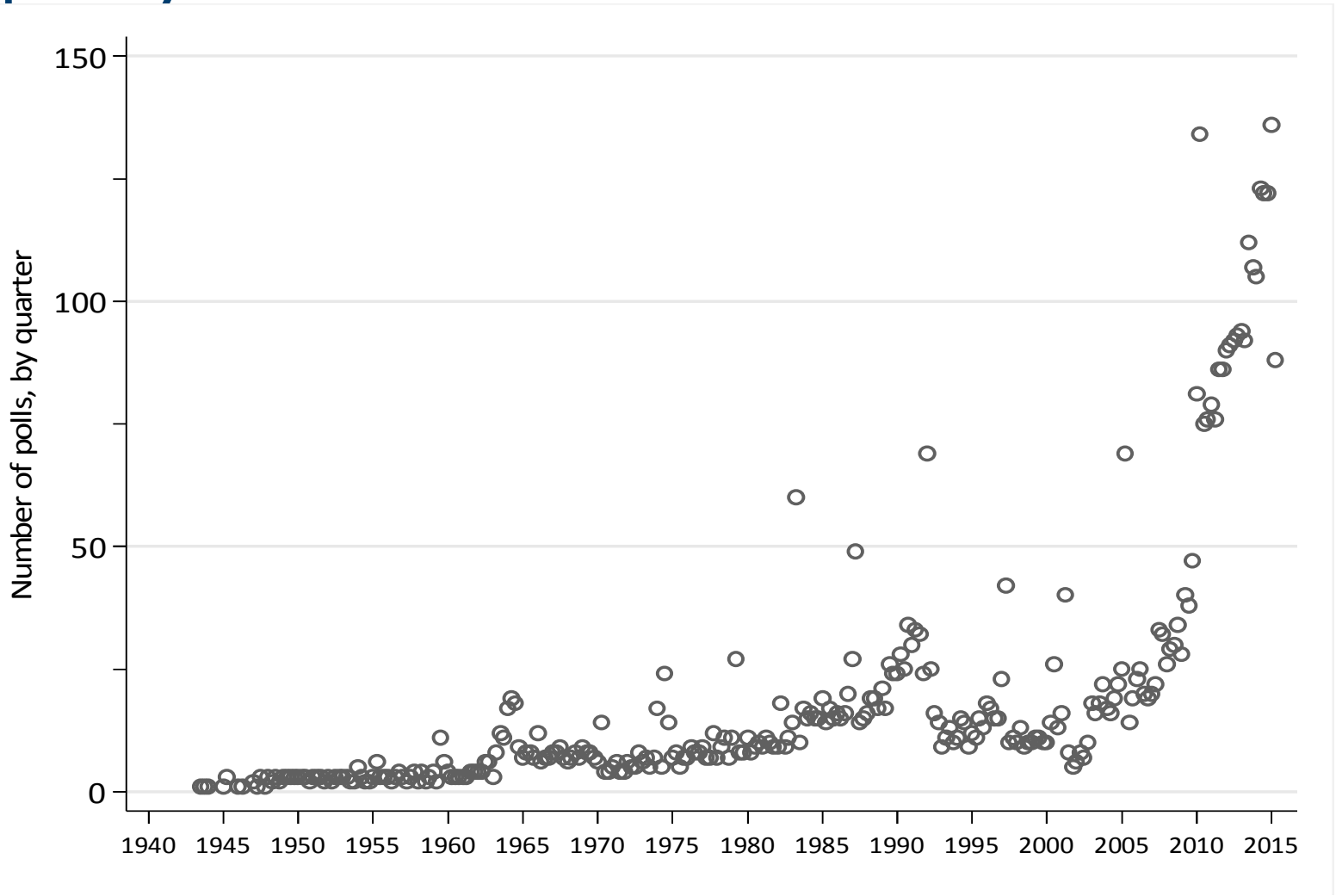
Tory/Labour respondents in campaign polls



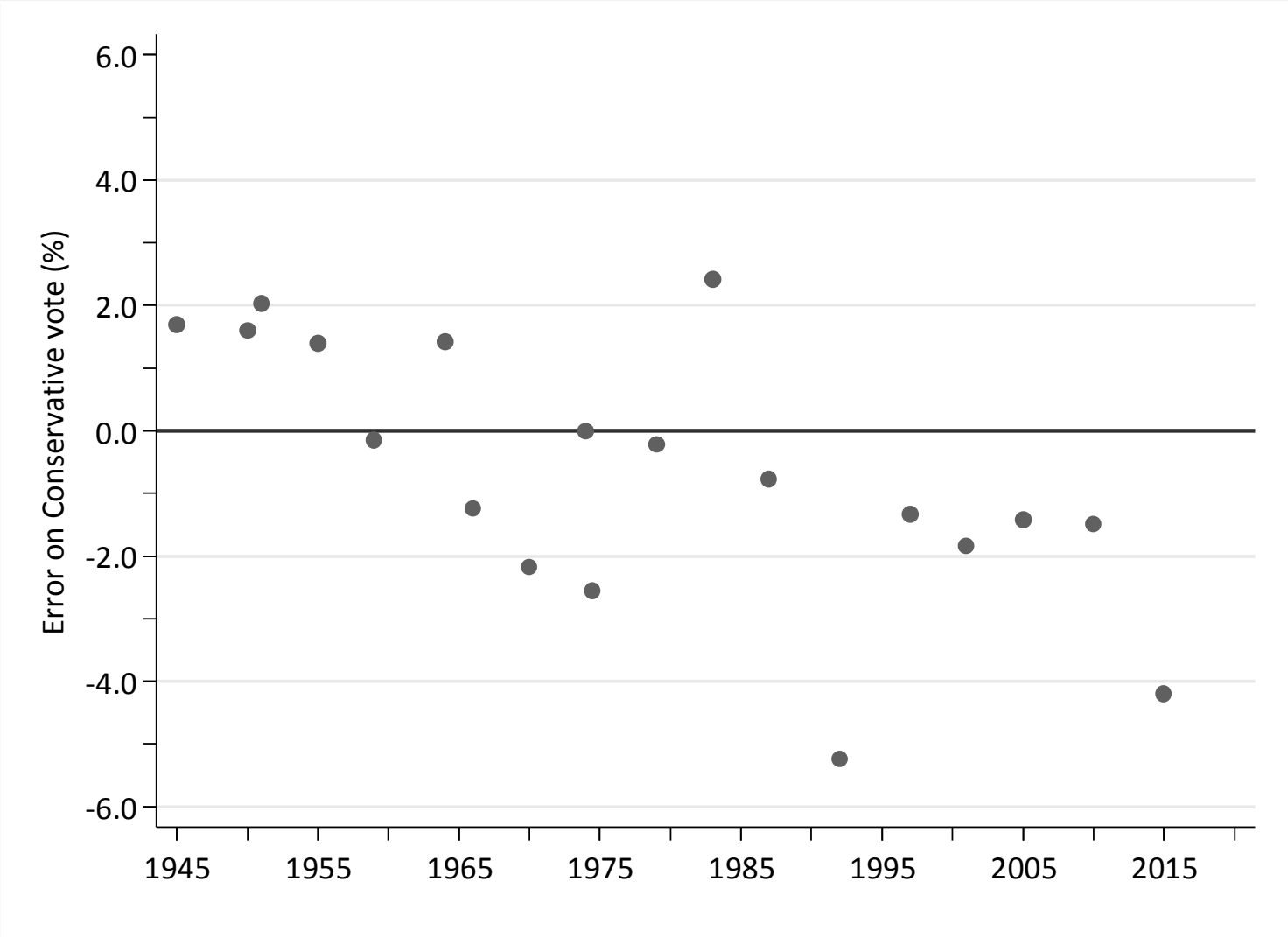
Context

- During the short campaign (March 30th to May 7th)
  - 91 polls for GB.
  - 13 polls for Scotland.
  - 4 polls for Wales.
  - 61 polls for constituencies (Ashcroft).
- During the ‘long campaign’ (May 2010 to 2015)
  - 1,942 polls for GB.
  - Compared to approx. 3,500 polls for GB between 1945 and 2010.

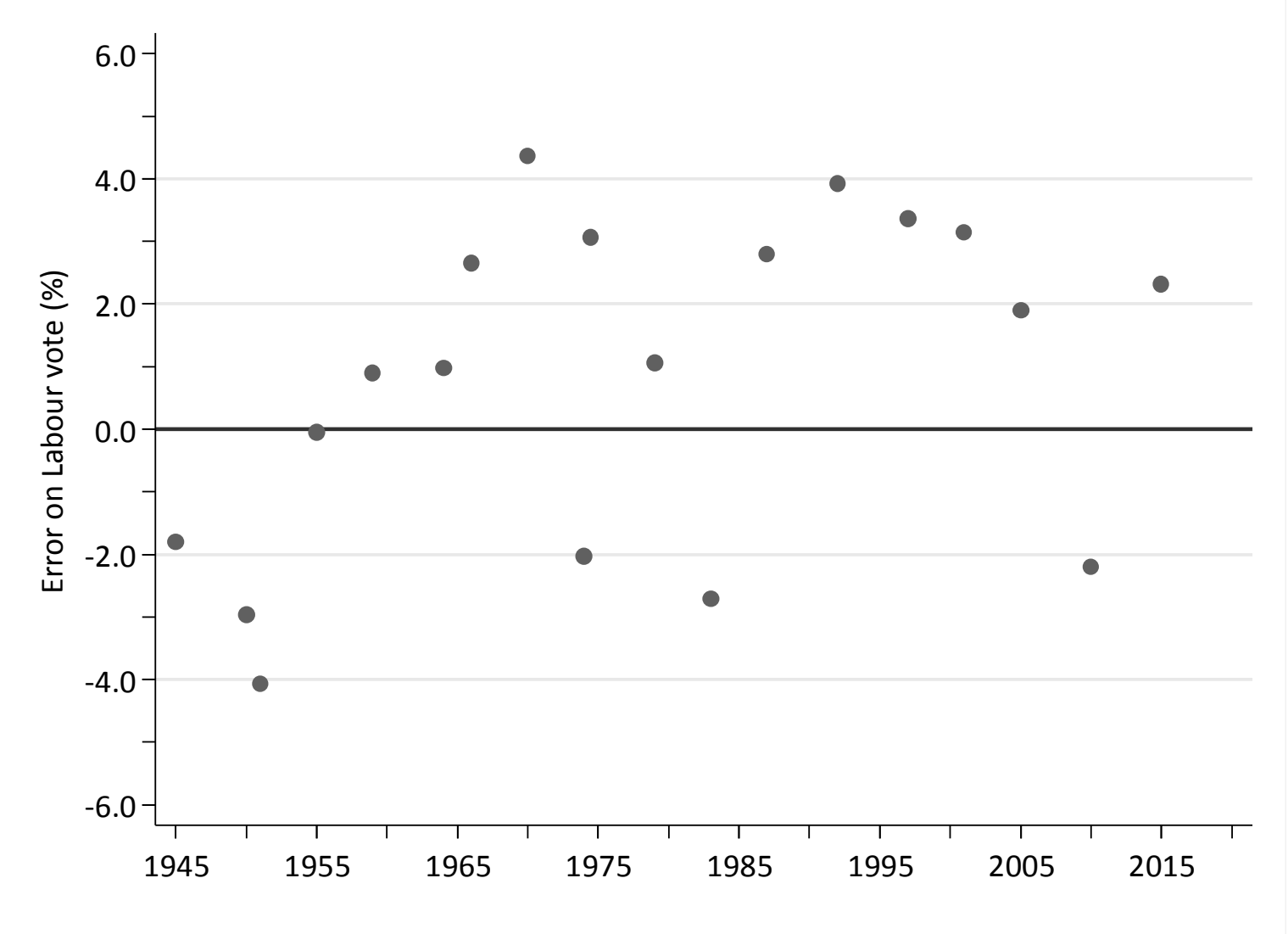
Frequency of GB Polls



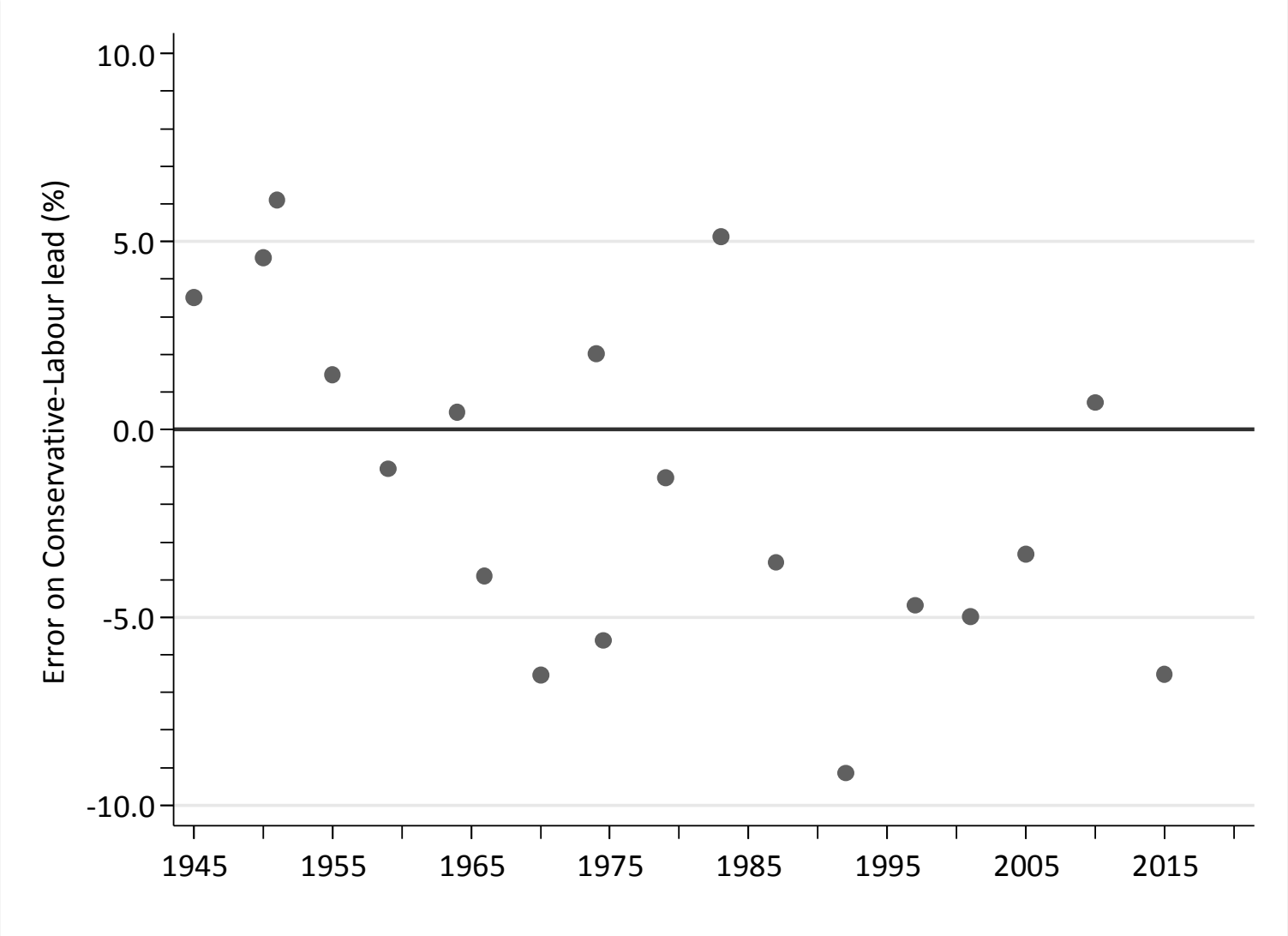
Error on Conservative vote share



Error on Labour vote share



Error on Con/Lab lead



NCRM National Centre for Research Methods

2015 Inquiry

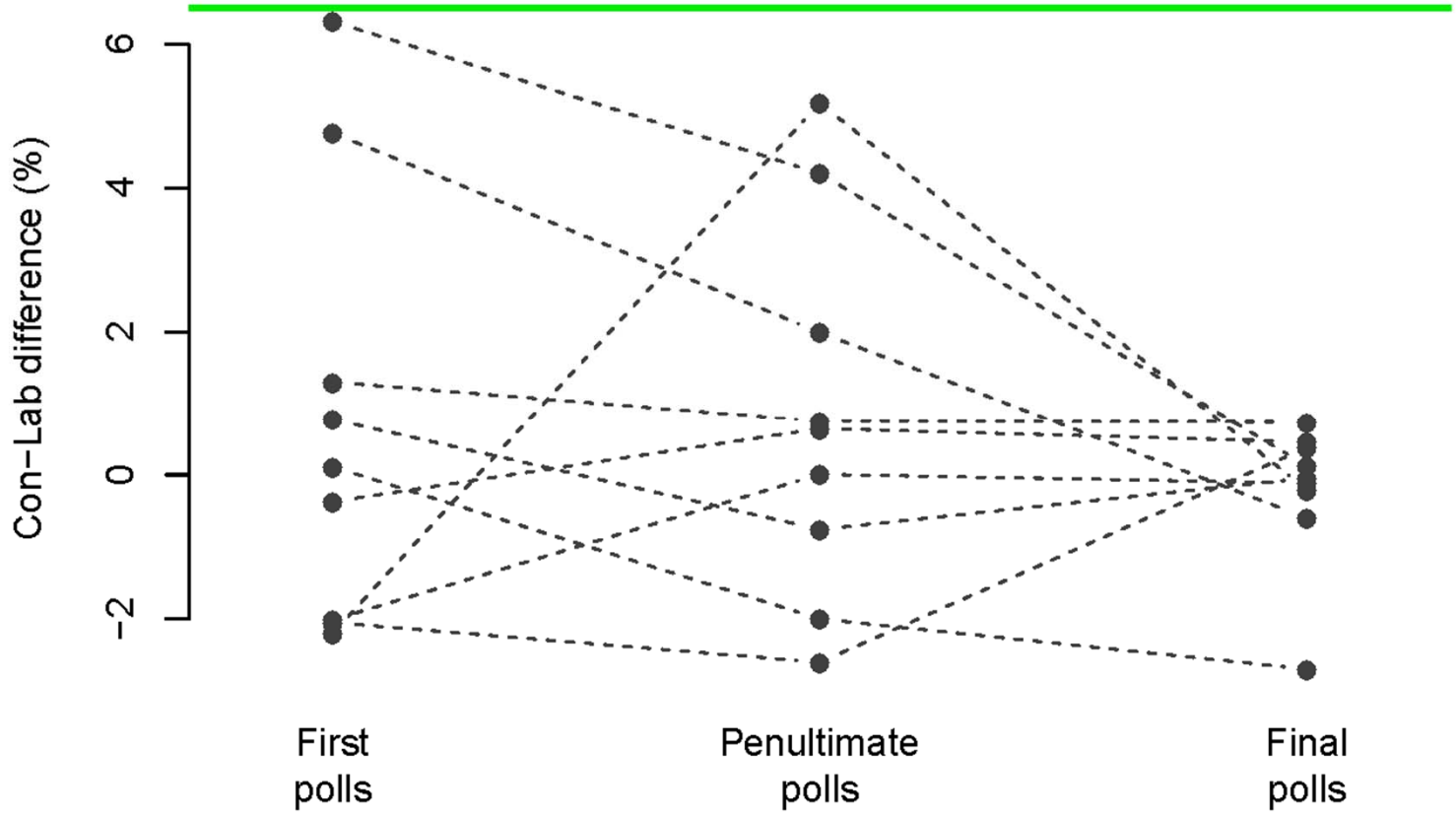
What we have done

- Identify potential causes of the miss and assess the evidence
- Sometimes evidence is not as strong as we would like, or doesn't exist at all
- Conclusions based on balance of probabilities not definitive proof
- We are not attempting a mathematical decomposition of the 6.6% average error

Evidence

- Three polls from each pollster
  - 1st poll of short campaign
  - Penultimate poll
  - Final poll
- Plus re-contact surveys, if undertaken
- Main parties + Ashcroft requested but declined
- Contemporaneous probability surveys
  - British Election Study
  - British Social Attitudes survey
- All published estimates have been replicated using these micro-data

Published estimates: Con-Lab difference



General and Specific causes

- We are interested primarily in *general* causes
- This does not imply that all identified causes apply equally to all pollsters
- Nor that some pollsters were not subject to errors we have not identified

Unlikely to have had an effect

- Postal voting
- Voter registration
- Overseas voters
- Question wording/framing
- Differential turnout misreporting
- Mode of interview

What's left

- Selection of samples
  - Obtain sample of registered voters;
  - Weight to population targets;
  - Weight by likelihood of voting
- Late swing
  - Don't Knows/Refusals
  - Party switching
- Deliberate misreporting

Herding

- Many observers were surprised at the statistical consensus of the campaign polls
- 10/11 final polls had a lead of 0% or 1%
- Did design decisions pull the estimates to an implicit consensus of a dead-heat?

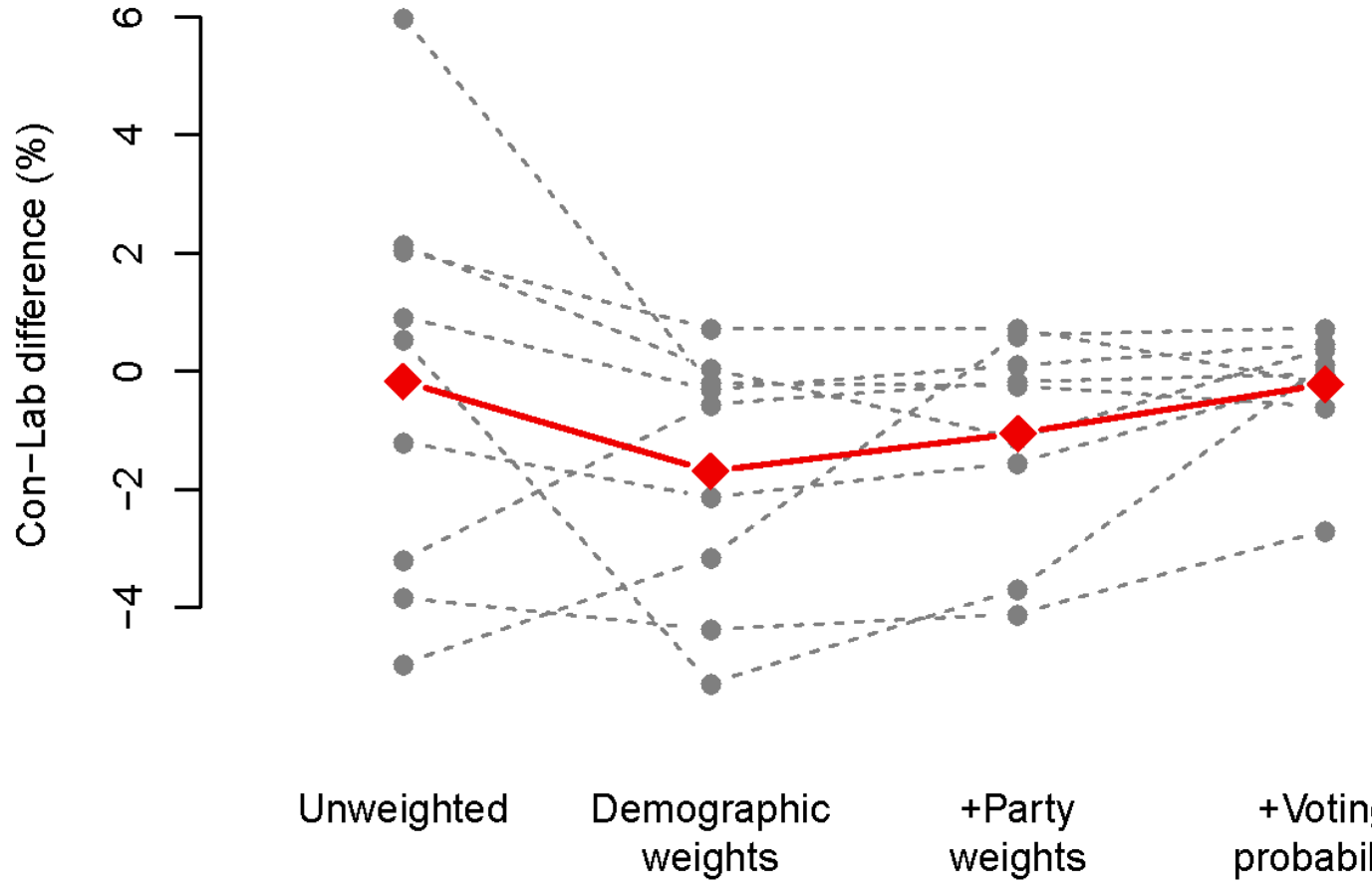
NCRM National Centre for Research Methods

The methodology of polls

Methodology of the polls: Three steps

1. Sample of eligible adults
   - Collect a quota sample of respondents
   - Weight to known population distributions: demographics (age, region, social grade,…) and party leaning/past vote
2. Sample of voters
   - Assign each respondent a turnout weight – probability that they will vote
   - Multiply by weights from Step 1, to give the final weights
3. Predicted election result
   - Respondents' stated vote intention, weighted by the final weights

Final polls: Con-Lab after different weightings

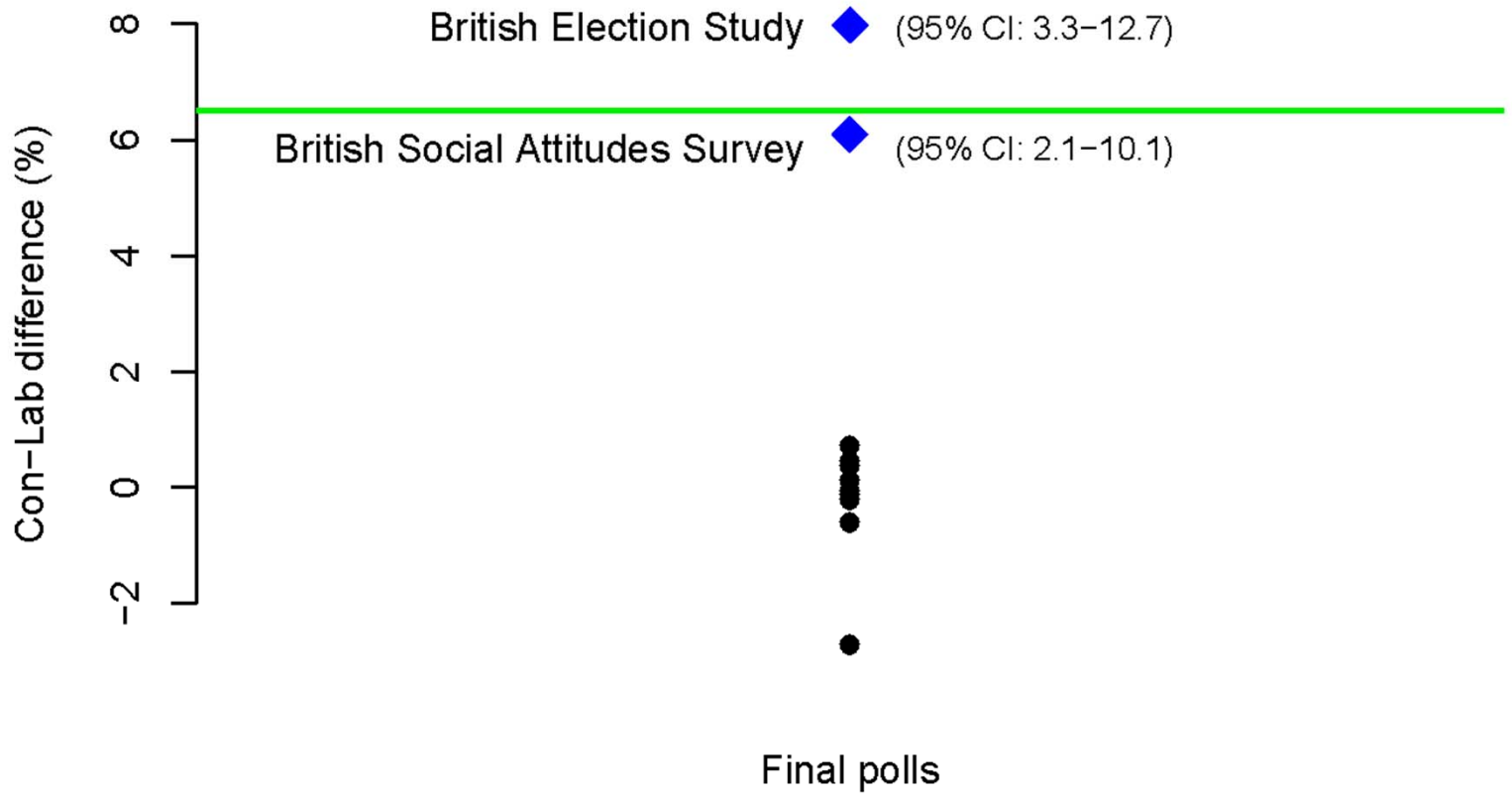


Assumptions of the methodology

For this to work, some key conditions should by met at each step:

1. Within levels of the weighting variables, sample members should have the same voting intention as the same group in the population
2. Assigned probabilities of turnout should be accurate, conditional on weighting variables and voting intention
3. Respondents' stated vote intentions should agree with how they actually voted

Final polls vs. Post-election surveys



Polls vs. Post-election surveys

BES and BSA differ from the polls in all three key steps of the methodology:

1. Random (probability) sampling instead of quota sampling
2. Turnout probabilities not needed, because respondents are known to have voted
3. The question is asked after the election, when respondents know how they had voted

So which one(s) explain the difference?

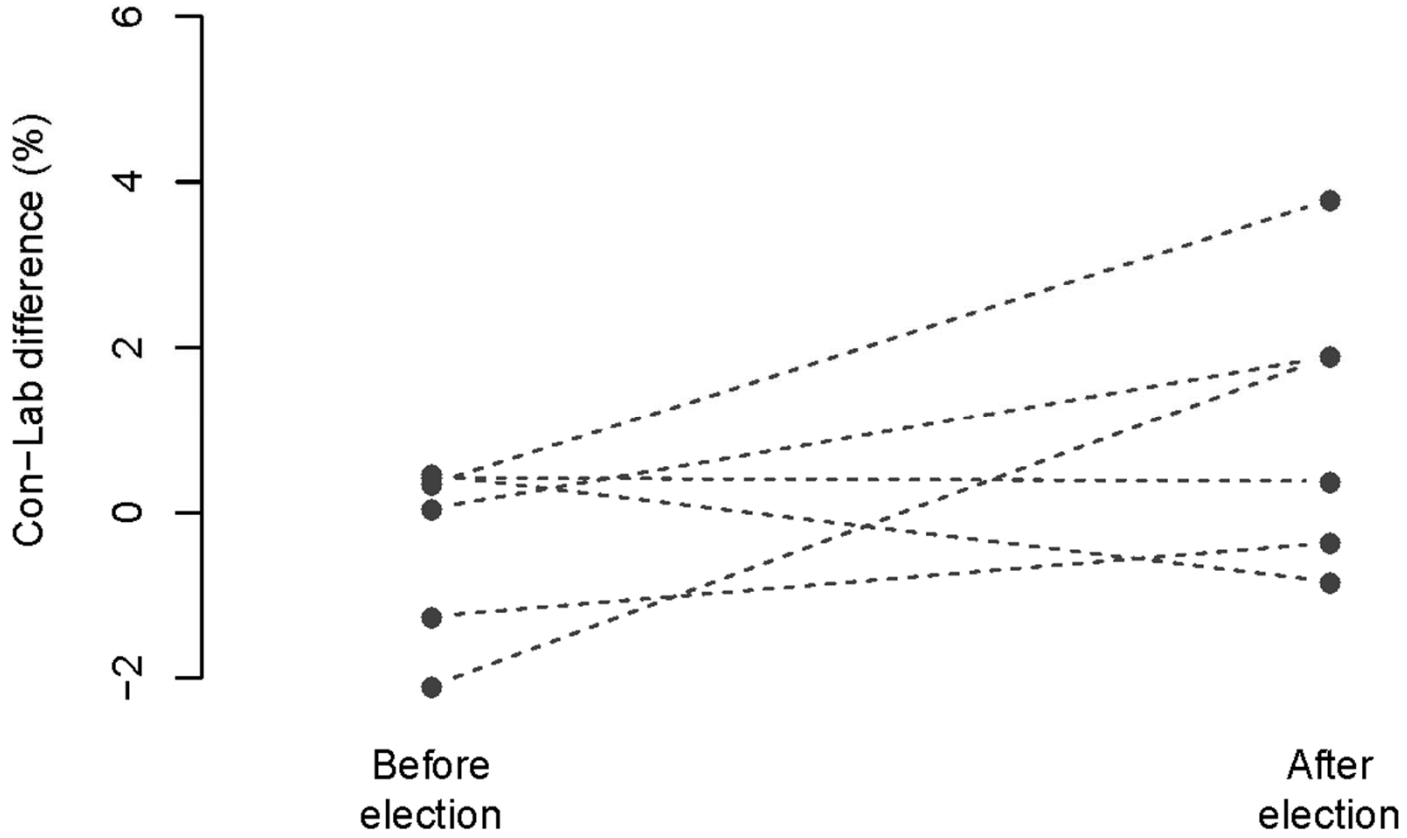
NCRM National Centre for Research Methods

Turnout weighting and Late swing

Late swing

- Main evidence from post-election re-contact polls, where (some of) the respondents of pre-election polls were interviewed again after election
- Compare reported vote after election to what voters among the respondents had said before election
- Evidence is inconsistent, but some sign of small swing toward Conservatives

Reported vote before and after election (known voters)



Turnout weighting

- Assigned turnout probabilities should be good estimates of actual turnout probabilities
  - conditional on demographics and/or answer to a direct question about likelihood of voting
- Estimating these probabilities is difficult
- Rationale for the approaches used not always obvious
- Poll estimates are vulnerable to errors in turnout weighting
- Nevertheless, imperfections of turnout weighting did not contribute much to the error this time

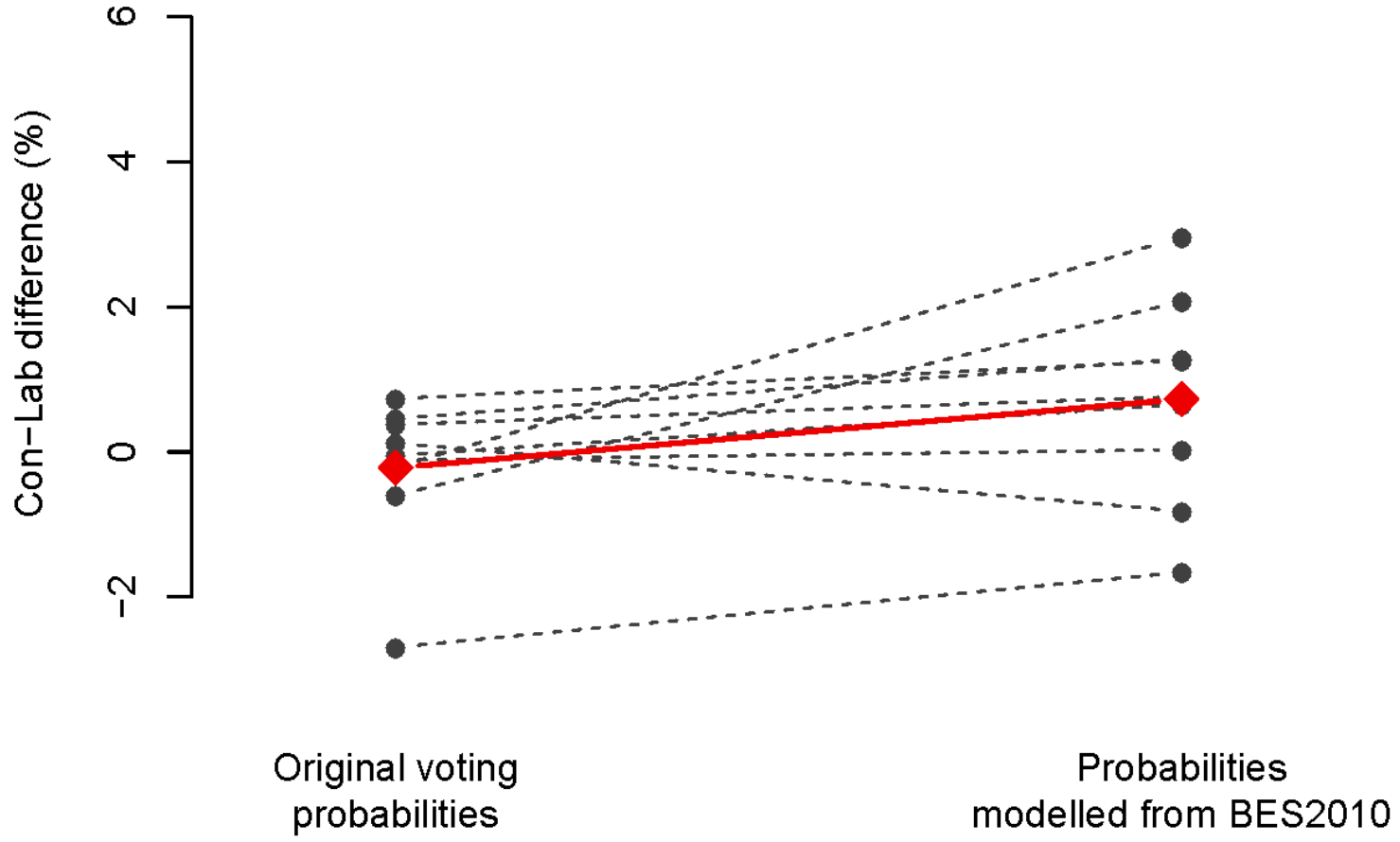
Turnout weighting: Analyses

Different types of evidence on the effects of turnout weighting:

- Pre-election polls vs. Pre-election vote intention of known voters in re-contact polls
- Assessment of the accuracy of the turnout probabilities, based on re-contact polls and validated vote data
- Sensitivity of the estimates to different specifications of the turnout probabilities

None of these show a substantial effect

Final polls: Original vs. Revised turnout weights



NCRM National Centre for Research Methods

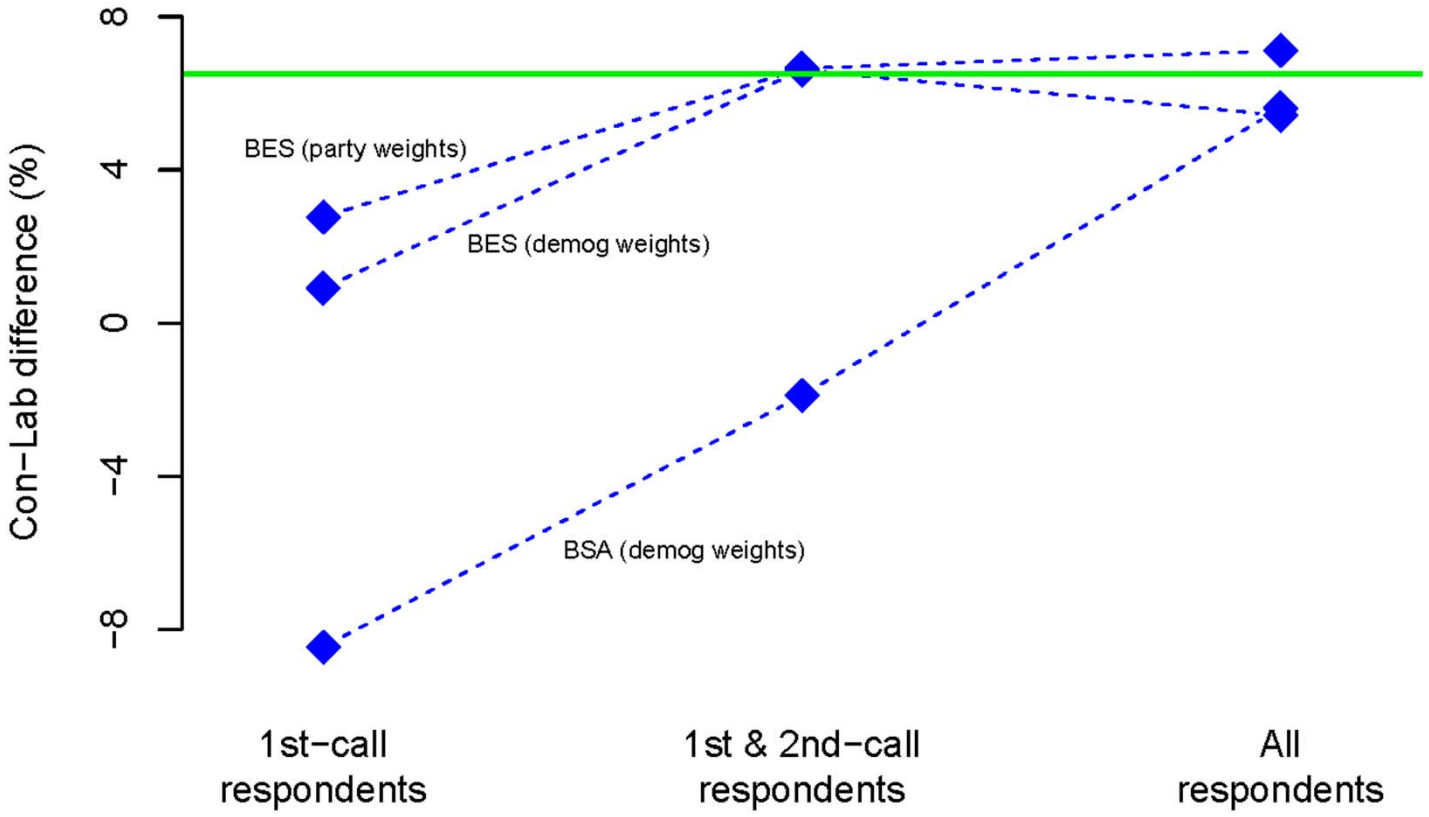
Unrepresentative samples

Unrepresentative samples?

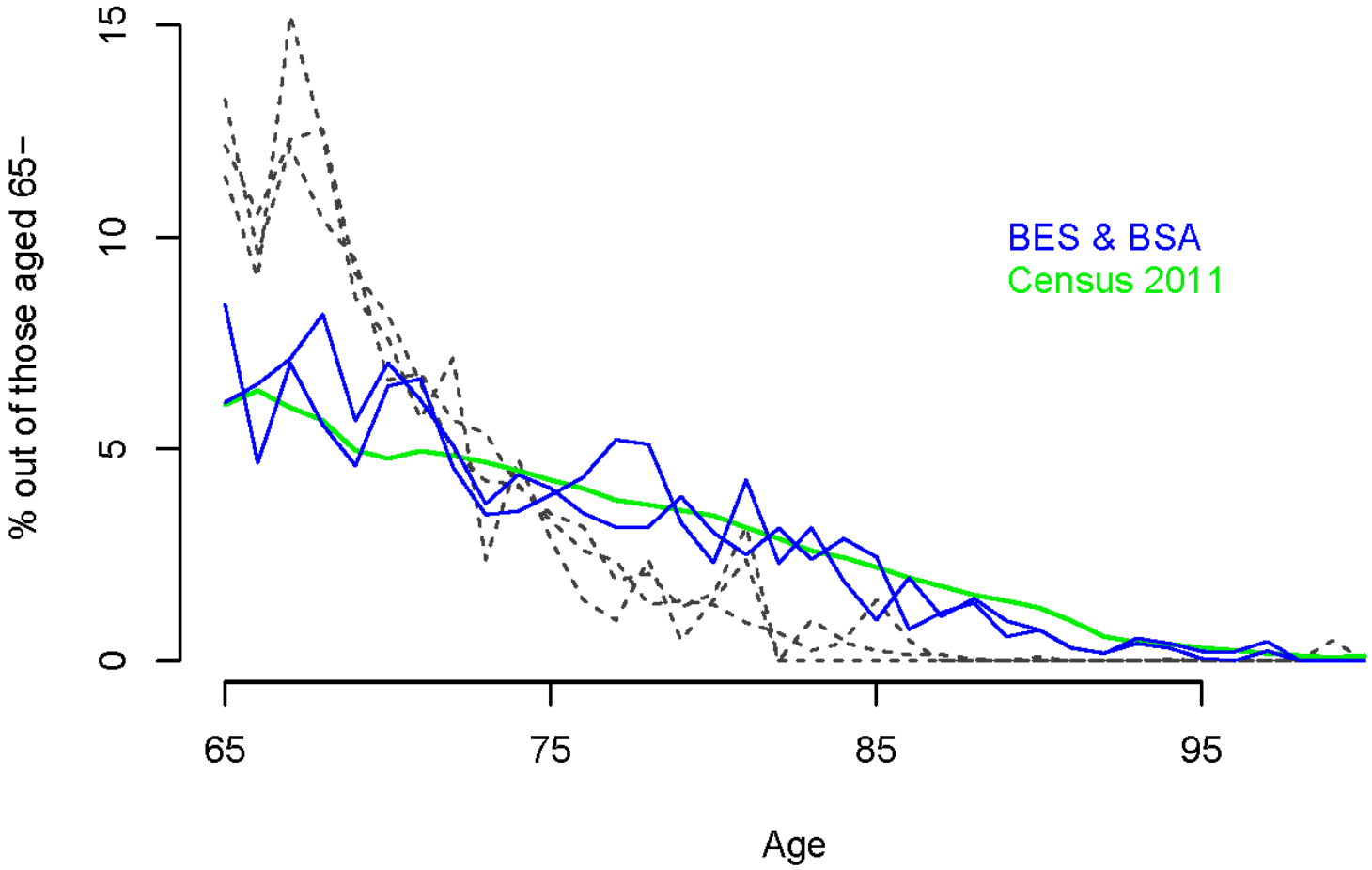
- If everything else now ruled out, unrepresentative samples is left as explanation of the error
- But what direct evidence do we have of this?

- Voters in re-contact polls vs. random samples (BES/BSA)
- Treat random samples as though quota samples
  - Inspired by Jowell et al. (1993)
- Representativeness of other characteristics of respondents in the samples, compared to other sources

Early-call respondents in BES and BSA



Age among those aged 65- (three polls)

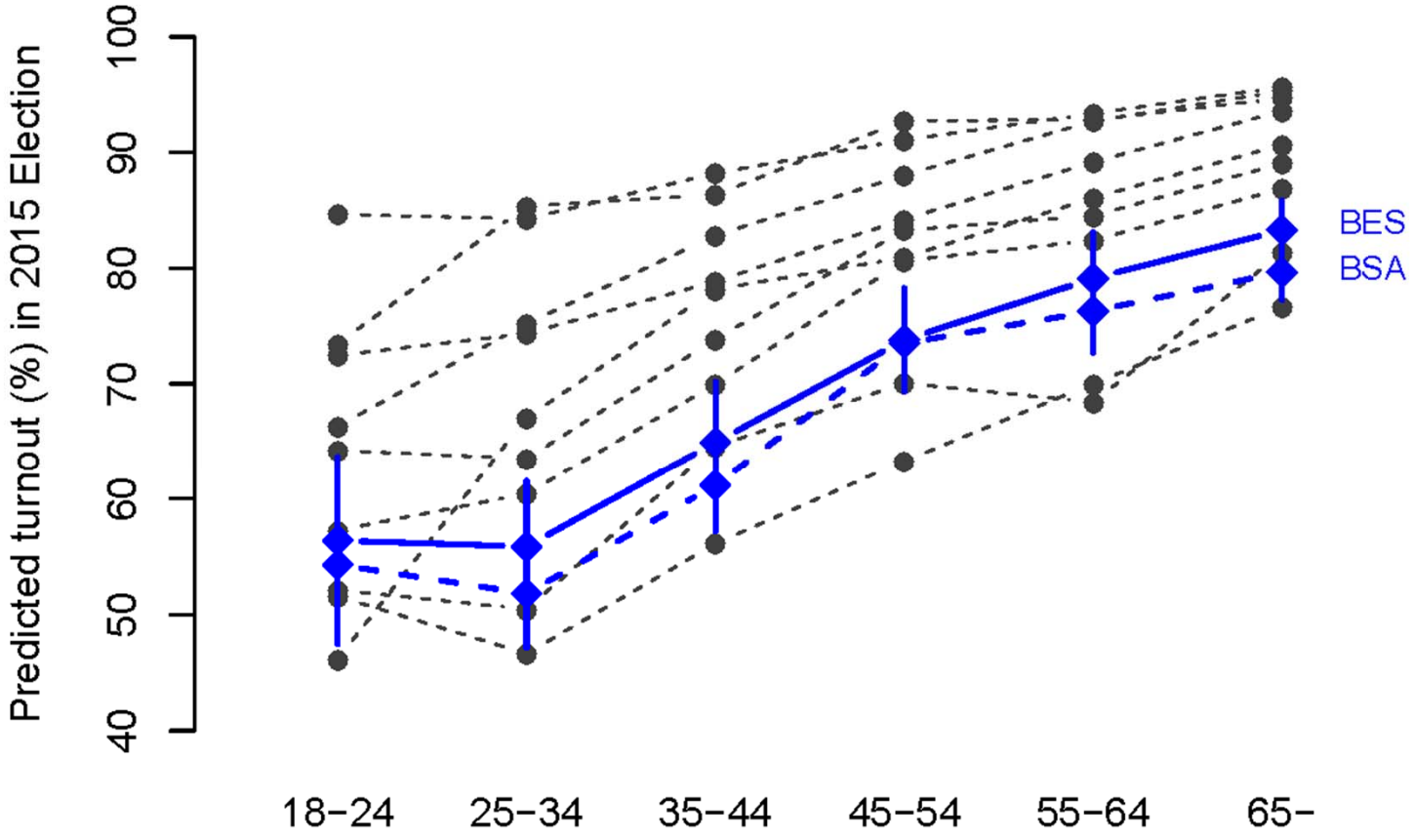


Overrepresentation of the politically engaged?

- Rivers & Wells (2015) and Mellon & Prosser (2015) find there were too many *politically engaged* people in poll samples (from one company) compared to BES

- We consider here predicted turnout (average of turnout probabilities) in the polls – with similar conclusion

Age vs. predicted Turnout in 2015

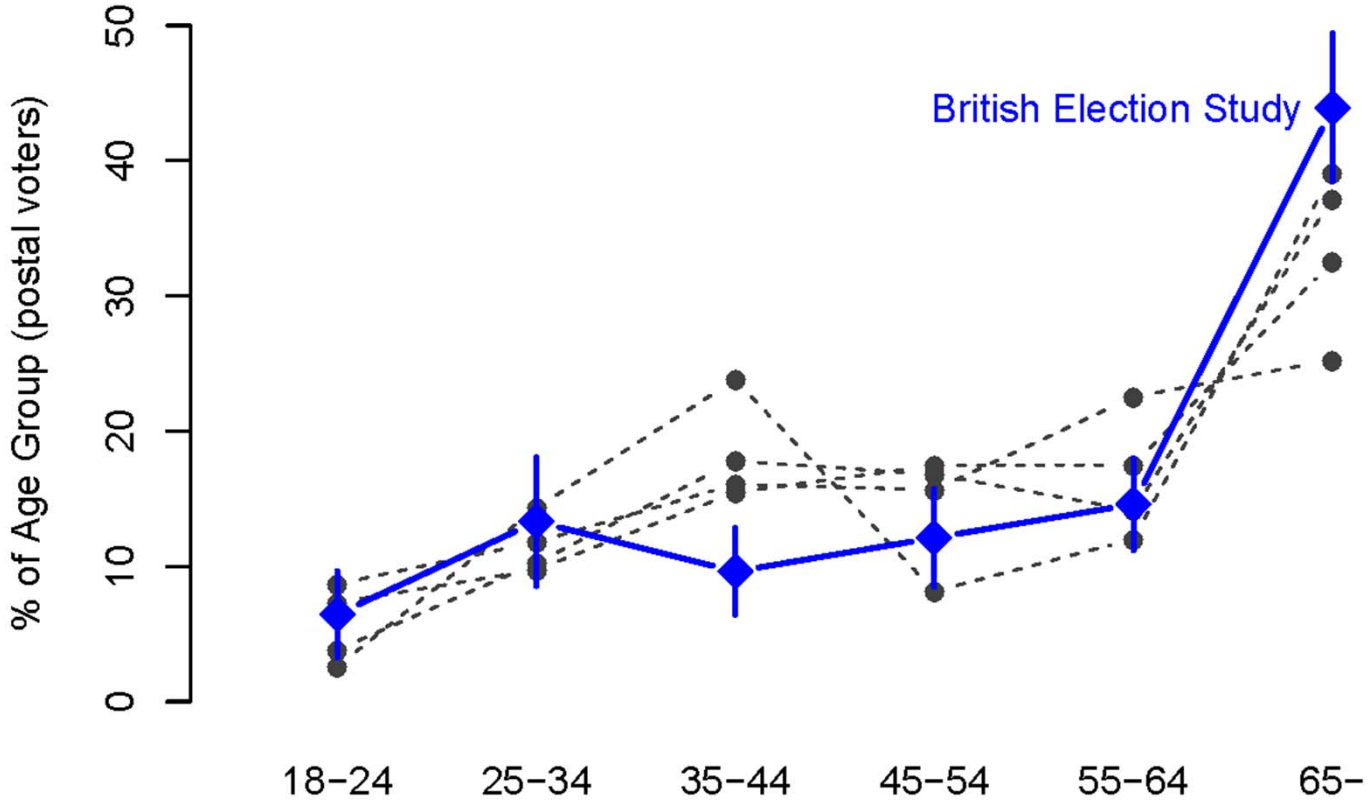


From unrepresentative samples to error in polls

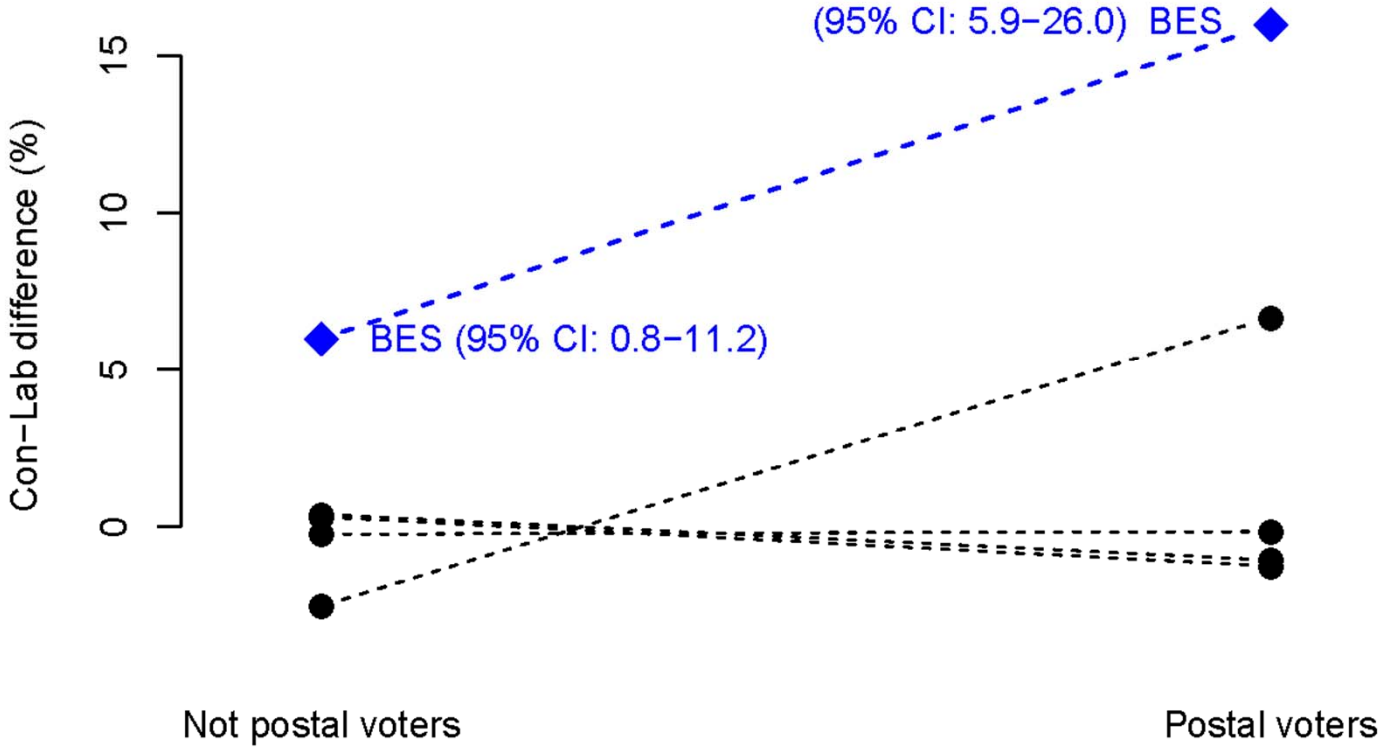
- Having too many or too few of some types of people in sample translates into error in predicted vote if those different types of people also vote differently
- For example, level of political engagement is associated with vote preferences
- Full explanation of the link between samples and error is likely to be complicated

- A small illustrative example: samples, age and vote among *postal voters*

Ages of postal voters (four polls)



Con-Lab among postal and non-postal voters



How could the sampling be improved?

- Within current (quota) methodology, two broad approaches:

    1. Obtain a more representative sample of people within levels of current weighting variables
    2. Improve weighting, by using more/better weighting variables

- Report and recommendations will say more about these

NCRM National Centre for Research Methods

Herding

What is herding?

- *"Herding specifically refers to the possibility that pollsters use existing poll results to help adjust the presentation of their own poll results. Herding strategies can range 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 existing polls."*
  - American Association for Public Opinion Research http://www.aapor.org/AAPORKentico/Education-Resources/Election-Polling-Resources/Herding.aspx

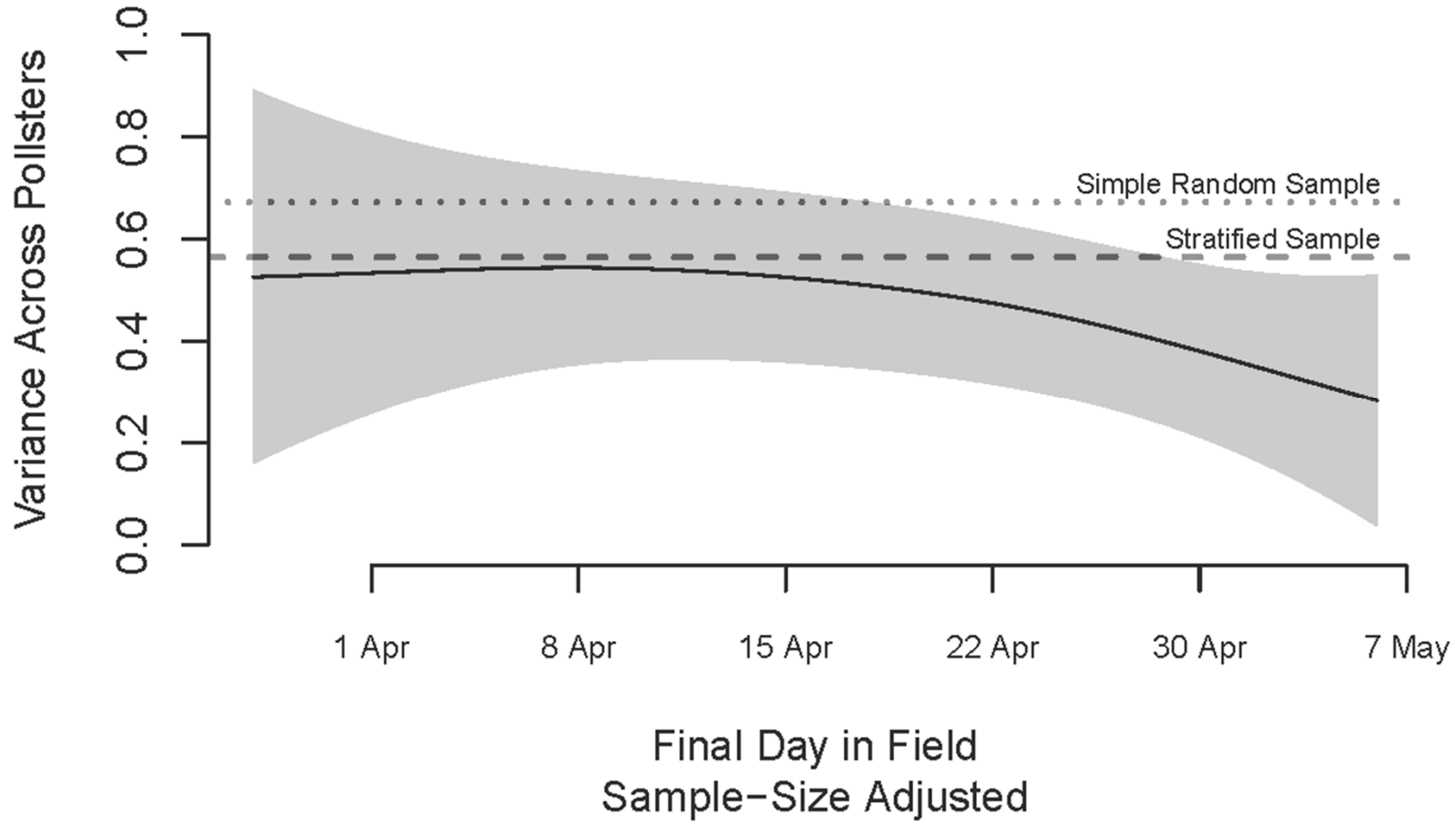
Why herd?

- Herding can arise because the pollsters are trying to be accurate:
  - Consciously or unconsciously compare the initial results of a new poll to previous polls in order to assess whether analysis procedures are working well.
- Today, we focus on the Con-Lab difference...
  - Almost all pollsters published final polls with Con-Lab differences between +1 and -1.
  - This is the headline result, the most electorally important result, and is largely what pollsters are judged on.

Variability of final polls

- Did the reported Con-Lab differences across pollsters vary less than they *ought* to have varied in the final polls?
- We have compared the level of variation in the final polls to three benchmarks:
  - Polls at earlier times in the campaign (adjusted for sample size).
  - Theoretical variability of the pollsters designs (treated as simple random samples as well as stratified samples).

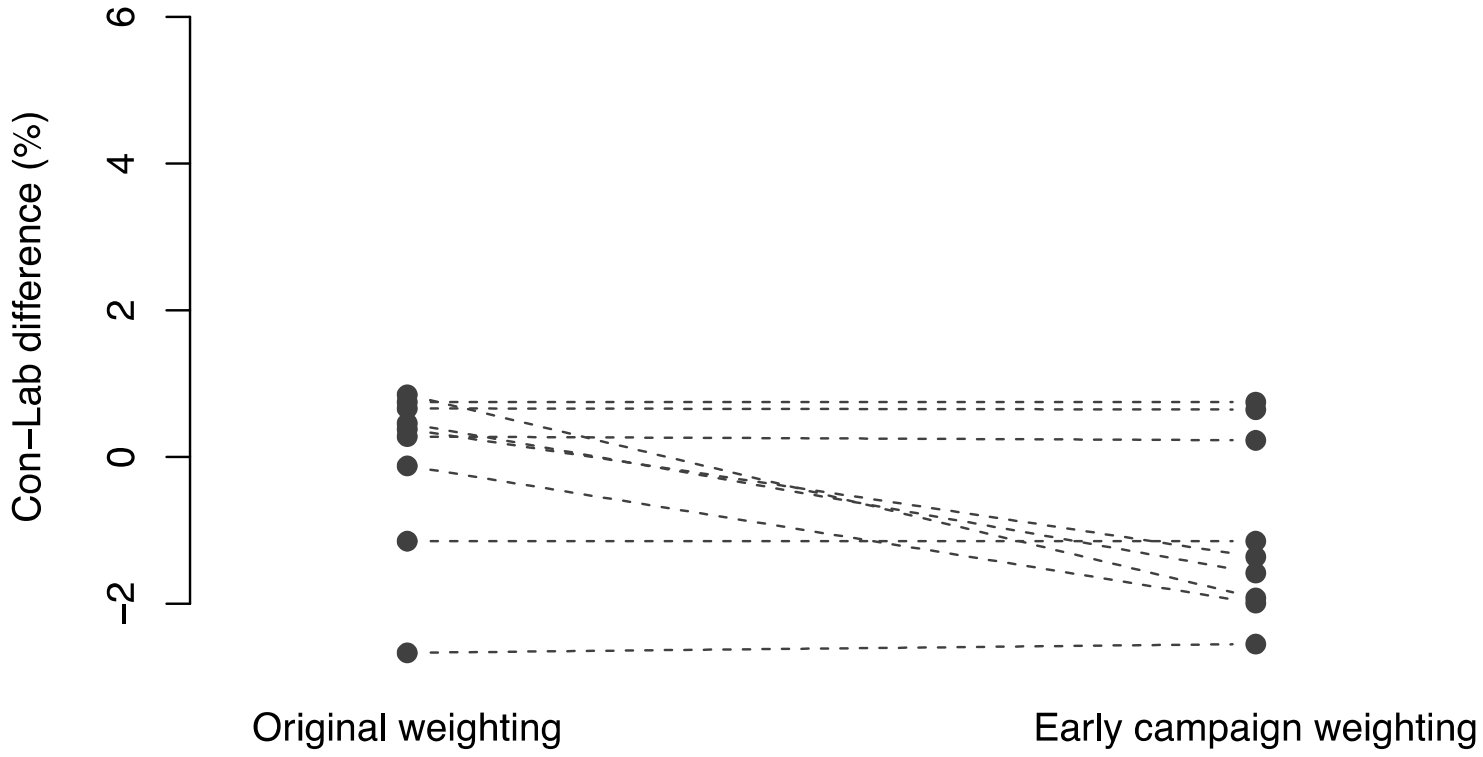
Variability of final polls



Final polls weighted as earlier polls

- We have reanalysed the pollsters final polls, using the weighting approach they followed in their earlier polls.
  - Did the changes pollsters made as the election approached reduce the observed variation across pollsters?

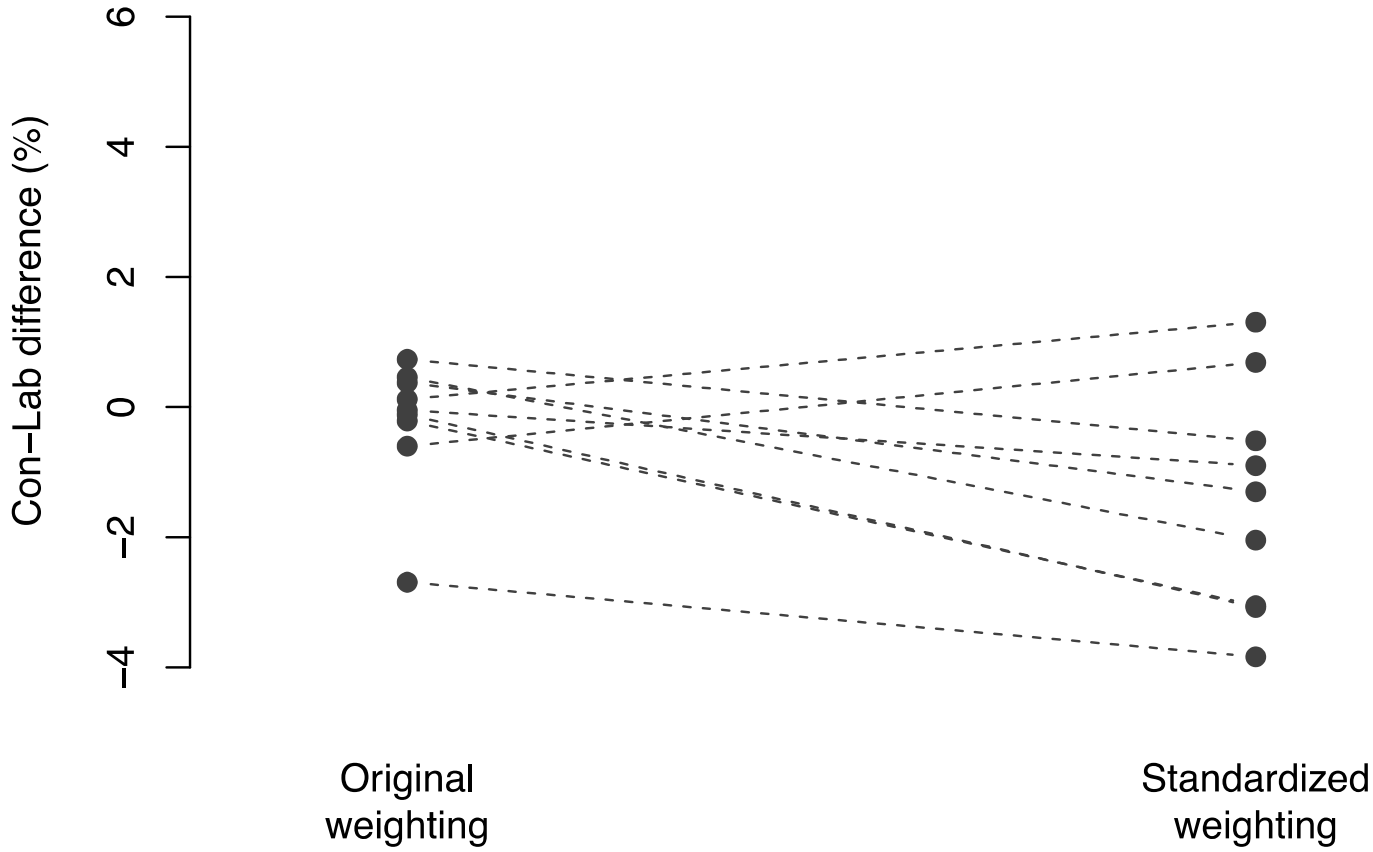
Final polls weighted like early campaign polls



Final polls weighted identically for all pollsters

- Did adjustment methods of some/all polls change?
  - Must have evidence that methods changed before final polls.
- Reanalysed the final polls using a common weighting approach
  - Given the raw data the pollsters were collecting and common weighting approaches, would they have reported more variation than they did using the their own methods?

Final polls weighted identically for all pollsters



Summary

- There was less variation in the Con-Lab lead across pollsters at the end of the campaign than early in the campaign.
- If the pollsters had analysed the raw data from the final polls in the ways they analysed their earlier polls, they would have reported slightly more variable (but also generally worse) results on the Con-Lab lead.
- If the pollsters had all analysed the raw data from their final polls in exactly the same way, they would have reported more variable results on the Con-Lab lead.

Comments

- None of this implies malpractice.
  - Nor does it imply that *all* pollsters were “herding”, however it is suggestive that some were
  - Herding could come about through pollsters doing their best to get the right vote distribution
  - There is no real evidence that this is responsible for the fact that the polls were off the mark on average

What is to be done?

- Recommendations in report, March 2016
  - Changes to current methodologies
  - No recommendation that all polls should move to random probability designs
  - BPC regulations on transparency
  - Reporting & interpretation of polls
  - Future data collection
- There will be no ‘silver bullet’, the risk of polling misses in the future can be reduced, not removed