Missing From The Party?

Using Administrative Data to Measure and Correct Partisan Nonresponse in ABS Samples

ssrs research. refined.

Jordon Peugh
Cameron B. McPhee
Mickey Jackson
Arifah Hasanbasri

The Washington Post

Scott Clement
Emily Guskin

SCHAR School of Policy and Government
GEORGE MASON UNIVERSITY

Mark J. Rozell

AAPOR
Annual Conference
2022
Chicago, IL

May 12, 2022

Background

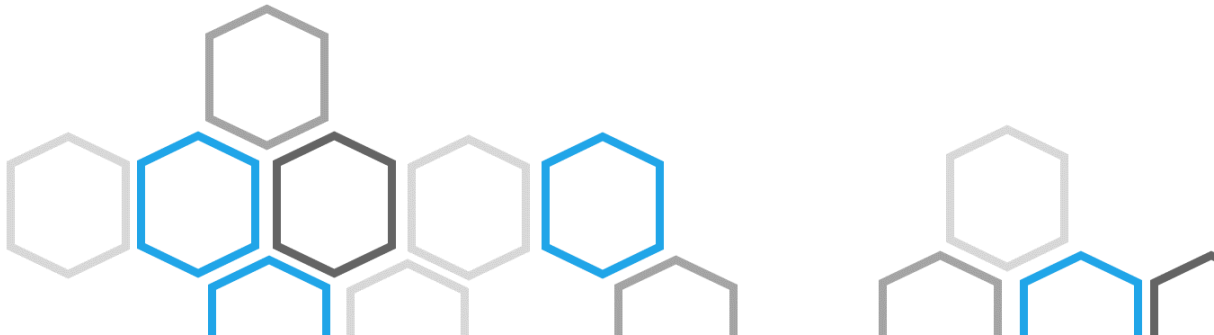
- The 2020 U.S. pre-election polls had the highest error in national polls in 40 years.
- AAPOR task force suggested the polling error may have been caused in part by non-ignorable partisan nonresponse, even after controlling for weighting demographics.
- If there is systemic nonresponse bias in election polls, other polls may also suffer from biases.
- Consequently, measuring and correcting for non-ignorable partisan nonresponse is a high-priority research area for public opinion research.
- However, even assessing partisan nonresponse is challenging.

Challenges in Assessing Partisan Nonresponse

- Benchmarking using Party-ID (PID) population distributions has limitations.
  - PID is not collected on high quality federal surveys typically used for demographic benchmarks.
  - Self-reported PID is sensitive to response mode and item wording.
  - PID is an attitude, not a fixed characteristic, changes over time.
- Benchmarking using certified results of election vs. self-reported recalled vote has limitations.
  - Sample must be limited to respondents who voted in the prior election.
  - Voting tends to be overreported.
  - Voters may misreport their candidate choice.
- Because of these issues, comparisons of the responding sample to external benchmarks are unsatisfactory for assessing partisan nonresponse.

Can appended administrative data help improve identifying and correcting partisan nonresponse?

- Address-based samples (ABS) can be linked to many auxiliary administrative data sources, including marketing databases and voter files at the address level; publicly available estimates for small areas such as census tracts and blocks; and electoral precincts.
- Some auxiliary data is correlated with PID and could be used to analyze nonresponse bias.
  - Comparing composition of respondents to the initially selected sample.
  - Potentially correcting for nonresponse by adjusting respondents back to their sample distribution.
- Appended data on PID or related characteristics could help sidestep the limitations of traditional PID benchmarking.



Research Questions

Can appended indicators of PID and related characteristics be utilized to assess partisan nonresponse in mixed-mode ABS designs?

What do such "sample-based" PID indicators tell us about partisan nonresponse bias in a political survey of the U.S. adult population?

Does weighting on sample-based PID indicators yield a more politically representative sample?

Methods

- The Washington Post and the Schar School of Policy and Government at George Mason University (WP/GMU) Poll
- Explored attitudes toward technology platforms and regulation
- Data collection by SSRS
- ABS sample of 10,000 U.S. residential addresses
- Fielded November 2021 – 3-week field period
- Appended data:
  - Address-matched voter registration data from L2, including individual level PID and political ideology based on consumer file data
  - Precinct level 2016 and 2020 election results, using geographic match based on latitude and longitude
  - FIPS code to county-level 2020 election results from Atlas of U.S. Presidential Elections

Appended Data Used to Create Modeled PartyID

- Modeled PID ("MPID") was calculated using a gradient boosted model trained on SSRS Opinion Panel recruitment sample (nationally representative address-based sample).
- Dependent variable in MPID model was self-reported PID (including leaners).
- Predictors in MPID model were:
  - Contextual variables –precinct-level presidential election results
  - Administrative variables – voter file (including inferred PID) and consumer file (including ideology) variables; address-level but missing for some addresses
- Sampled addresses matched to external files were assigned a supplementary predicted MPID variable.
- MPID was a model-based aggregation of these contextual and administrative variables, available regardless of response status to the survey.

Data Collection Protocol and Response Rate

- **Mailing 1: Invitation**
  - Paper invitation with URL and login information for the online questionnaire (bilingual)
  - Hard copy of the questionnaire
  - Spanish hard copy of the questionnaire included for likely Spanish speaking households (20% of the sample)
  - Prepaid return envelope
  - $1 or $2 cash incentive
- **Mailing 2:** Bilingual reminder postcard
- **Promised Incentive:** $10 for completing the online questionnaire or $5 for mailing back the paper survey
- **Response Rate:** AAPOR RR3 of 11.6%
  - 1,122 completed interviews (772 online; 350 by mail)

Weighting

Multiple weighting approaches were explored. For this presentation, we will primarily discuss the final weight used.

Weighting on the following:

- Pre-raking nonresponse adjustment using MPID and county-level election results
- Demographic raking targets: gender, age, race/ethnicity, education, education X race/ethnicity, Census region, population density
- Final weight included MPID as a raking dimension (with full-sample distribution as target)

Results

Research Question 1: Utility of Sample-based PID Indicators

Household-level indicators from administrative and contextual data were strongly predictive of self-identified PID, a clear indication that they are useful in measuring and correcting for partisan nonresponse bias.

DISTRIBUTION OF SELF-REPORTED PARTY IDENTIFICATION, BY MODELED PID

Sample-based Party Identification Indicator	Self-reported Party Identification		
	Democrat/ Lean Democrat	Republican/ Lean Republican	Independent/ Other, no lean
Modeled Party Identification (MPID)			
Democrat	66.5%	14.1%	19.4%
Republican	18.4%	64.4%	17.2%
Accuracy		*65.4%*	
Coverage		*92.6%*	

Research Question 2: Evidence of Partisan Nonresponse Bias

Households predicted to be Republican were overrepresented in the unweighted sample, but underrepresented (relative to Democrats) in the demographically weighted sample.

SAMPLE DISTRIBUTION VS. UNWEIGHTED AND WEIGHTED RESPONDENT DISTRIBUTIONS, BY SAMPLE-BASED PARTY IDENTIFICATION INDICATORS

Sample-based Party Identification Indicator	Percent of Sampled Households	Percent of Respondents		Difference from Sampled Households	
		Unweighted	Weighted by Demographics	Unweighted	Weighted by Demographics
Modeled Party Identification					
Democrat	40.9%	41.8%	43.9%	0.9%	3.0%
Republican	45.2%	50.8%	46.2%	5.6%	1.0%
Missing	13.9%	7.4%	9.9%	-6.5%	-4.1%
Dem – Rep margin	*-4.3%*	*-9.0%*	*-2.3%*		

Research Question 3: Utility of Weighting by Sample-based PID

Weighting by MPID produced a more accurate estimate of recalled 2020 vote choice when compared with a demographics-only weighting protocol.

EFFECTS OF SAMPLE-BASED PARTY WEIGHTING ON RECALLED VOTE

Measure	Population Parameter	Unweighted Percent of Respondents	Weighted Percent of Respondents	
			Weight 1 (Demo Only)	Weight 4 (MPID)
Recalled 2020 Vote (Among Self-reported Voters)				
Biden	51.3%	49.3%	51.5%	50.4%
Trump	46.9%	44.0%	41.4%	42.5%
Other/Refused	1.8%	6.7%	7.0%	7.1%
Biden - Trump margin	*4.4%*	*5.3%*	*10.1%*	*7.9%*

Research Question 3: Utility of Weighting by Sample-based PID, *Continued*

Looking at validated voters, weighting by MPID produced estimates closer to certified election results.

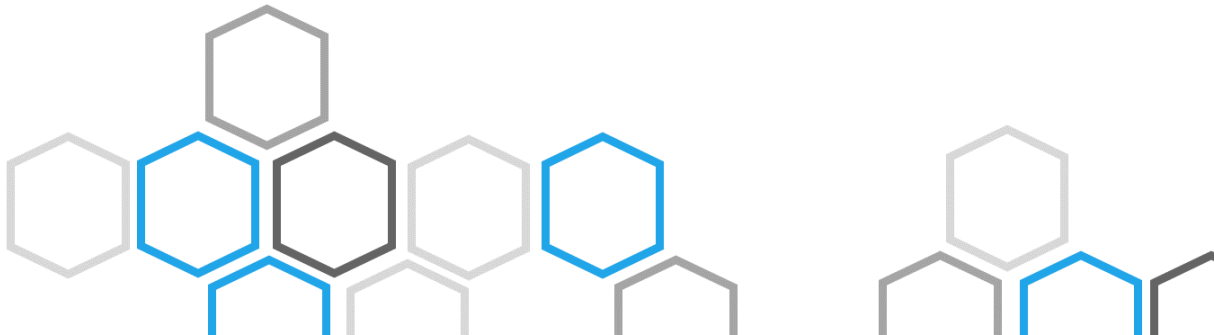
ESTIMATED TURNOUT AND RECALLED-VOTE DISTRIBUTION, BY VOTER DEFINITION (WEIGHT 4 - MPID)

Voter Definition	Estimated Turnout (Voting-age Population)	Recalled 2020 Vote		
		Biden	Trump	Biden - Trump Margin
Self-identified Voters	75.3%	50.4%	42.5%	*7.9%*
Validated Voters - Address Level[2]	61.0%	49.2%	43.3%	*5.8%*
Validated Voters - Strict Person-level[3]	52.8%	48.7%	43.9%	*4.8%*
Validated Voters - Modified Person-level[4]	58.5%	49.1%	43.1%	*6.0%*
Certified Election Results	*61.5%*	*51.3%*	*46.9%*	*4.4%*

Discussion and Limitations

Discussion

- ABS sample linked to administrative and contextual indicators of likely political attitudes allowed for the development of a model-based party identification (MPID) indicator for the majority of the sample and was 65% accurate in predicting self-reported party leaning, with fewer than 20% of respondents predicted to belong to the wrong party.
- Appending PID predictions allows for:
  - Monitoring of response rates by predicted partisanship during data collection
  - Measurement of non-ignorable partisan nonresponse
  - Final sample of respondents to be weighted to match predicted partisan makeup of population, without relying on another survey for weighting targets
- Method can be applied to cross-sectional and panel surveys where addresses are available.

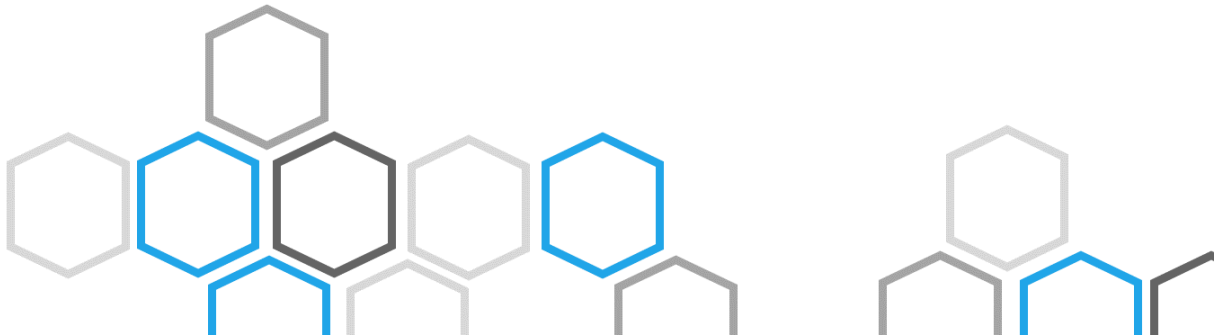


Limitations

This is one national survey; additional research is needed to refine utility of sample-based PID indicators in reducing partisan nonresponse.

Practical Challenges

- Lot of effort needed to analyze existing address-matched survey data to estimate PID prediction model and to append administrative and geographic predictors to ABS sample;
- Long field period (3 weeks) needed, longer than a typical pre-election survey



JORDON PEUGH

ssrs
research. refined.

Chief Business Officer

484.840.4337
jpeugh@ssrs.com
@jxpeugh

AAPOR
Annual Conference
2022
Chicago, IL

May 12, 2022