

Reinforcing Predictions: Replicating Analytic Models Across Samples and Modes

Madeline S. Gosz, Randall K. Thomas, & Frances M. Barlas

Ipsos Public Affairs

The logo for the American Association for Public Opinion Research (AAPOR), featuring the acronym "AAPOR" in a serif font with a checkmark over the "O", and the full name "AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH" in a smaller font below it.
77th Annual Conference
May 11-13, 2022
Sheraton Grand Chicago • Chicago, IL
www.aapor.org #aapor   

Come Together
Advancing Inclusion and Equity
Through Data Collection,
Measurement, and Community



Study Background & Purpose



There are generally two main types of online samples used in polling – probability-based and opt-in non-probability samples.

Though online samples are often used for general population national studies, there has been increasing interest in their use for smaller groups or geographies.

In addition, there are increasing attempts to develop alternative methodologies that can target specific areas, including Redirected Inbound Call Sampling (RICS), which uses IVR technology to present and receive answers that can identify cases relevant to a study.

Study Background & Purpose



The RICS method and sampling technique has been described by Levine, Krotki, & Lavrakas (2019, Public Opinion Quarterly) and has been shown to be a promising development that could help improve access to samples that may be more difficult to obtain than standard techniques.

With the rise of RICS, it is important to understand how it compares to, or can be used to supplement, traditional probability-based samples as well as non-probability opt-in sample to help reach smaller groups or geographic regions.

Study Background & Purpose



To compare these various sample sources, we conducted a political poll in Georgia in late summer 2020 using three separate sample sources:

- probability-based KnowledgePanel (KP),
- an opt-in non-probability online sample, and
- a RICS sample methodology.



Study Background & Purpose



In a study we reported on last year (Li, Thomas, Lien, & Barlas 2021), we computed the average deviation from 7 benchmark values for Georgia for each sample. We found that the KnowledgePanel sample had the lowest bias, followed by opt-in, with the RICS methodology having the highest bias.

Sample	Average Absolute Difference from Benchmarks
KnowledgePanel	4.50%
Opt-in	7.00%
RICS	9.10%

Study Background & Purpose



The elections in Georgia in 2020 (and into 2021) were high-stakes elections in which there were few percentage point differences between candidates. Joseph R. Biden (Dem) won the Presidential election and Jon Ossoff (Dem) won the U.S. Senate election.

These questionnaires captured demographics, voting patterns/decisions, and political attitudinal items from respondents in Georgia in the time before these elections took place.

In the study we are reporting here, we were interested in examining how predictive multiple regression models could obtain similar results with the different samples/modes.

Method

Study Design – Fielding

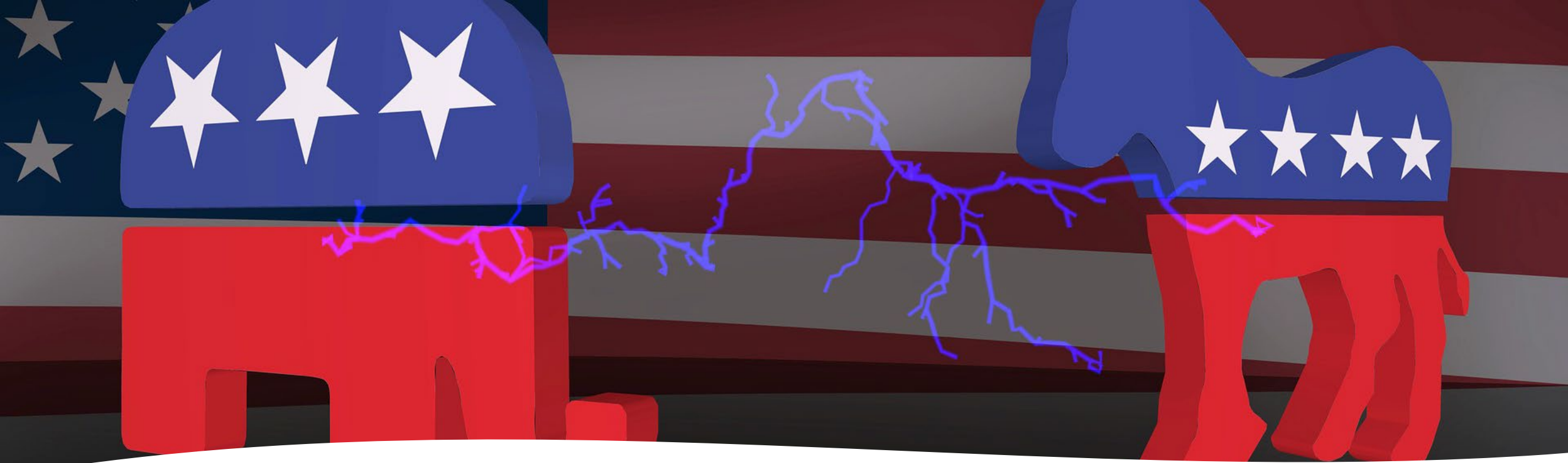


Respondents:

- **Probability-based KnowledgePanel (KP) sample: 525 respondents**
- **Opt-in non-probability online sample: 1,283 respondents**
- **Redirected Inbound Calling Sample (RICS): 1,180 respondents**

Field Period: August 13, 2020 – September 26, 2020

Modes of Administration: Web-based questionnaire for KP and opt-in samples and IVR technology to present questions and receive answers for RICS sample



Study Design – Random Assignments

To avoid question order bias, respondents were randomly assigned one of two paths as to receive questions. Questions from Path A were presented in one order (e.g. If Republican, press 1. If Democrat, press 2), while questions from Path B were read in the reverse order (e.g. If Democrat, press 1. If Republican, press 2).

Study Design – Weighting



Before our analyses, we first weighted each sample file using demographics derived from ACS and CPS benchmarks for the following variables:

- **Age-gender**
- **Region of country**
- **Education**
- **Income**
- **Race-ethnicity**

Study Design – Main Dependent Variables



There were four dependent variables of interest which we ran regression models on:

- **Vote likelihood** (Definitely will not vote, Probably will not vote, Uncertain to vote, Probably will vote, and Definitely will vote)
- **Political party ID** (Strong Republican, Moderate Republican, Lean Republican, No lean, Lean Democrat, Moderate Democrat, and Strong Democrat)
- **Presidential candidate choice** (Donald Trump, Another candidate/undecided, and Joe Biden) – filtered to registered voters
- **Senate candidate choice** (David Perdue, the Republican, Another candidate/undecided, and Jon Ossoff, the Democrat) – filtered to registered voters



Results

Results

Because the opt-in, KP, and RICS models were each weighted to common targets, we report the weighted regression models with the political attitudinal items as predictors.

Results – Predicting Likelihood to Vote



Likelihood to Vote (7=Definitely Will)	KP	Opt-in	RICS
Adjusted R ² of Model	64.7%	62.4%	43.1%
	Beta	Beta	Beta
Currently registered to vote	0.502	0.422	0.388
Voted in almost every election	0.256	0.278	0.197
Voted in 2016 election	0.213	0.201	0.187
Important to control gun ownership	0.138	-0.055	0.015
Hours of online activity for personal use	-0.039	0.043	0.034
Federal government should reduce wealth gap	0.031	-0.016	-0.011
TV viewing in average day	0.022	0.048	0.053
Country headed in right direction	-0.011	-0.002	-0.049

Results – Predicting Party Identification



Party Identification - 7 category (7=Strong Democrat; 1=Strong Republican)	KP	Opt-in	RICS
Adjusted R² of Model	63.7%	41.0%	46.0%
	Beta	Beta	Beta
Support the Black Lives Matter movement	0.34	0.225	0.389
Abortion should be illegal	-0.193	-0.096	-0.047
Federal government should ensure health care for all	0.149	0.12	0.045
Important to control gun ownership	0.136	0.199	0.214
Immigrants who enter illegally should be allowed to stay	0.127	0.139	0.129
Federal government should increase defense spending	-0.071	-0.145	-0.046
Federal government should reduce wealth gap	0.059	0.088	0.101

Results – Predicting Presidential Vote Choice



Presidential Vote Choice (1=Trump; 2=Neither; 3=Biden)	KP	Opt-in	RICS
Adjusted R² of Model	88.1%	81.3%	80.6%
	Beta	Beta	Beta
President Trump Approval - 7 category	-0.648	-0.465	-0.585
Party Identification - 7 category (1=Strong Republican; 7=Strong Democrat)	0.234	0.399	0.224
Important to control gun ownership	0.076	0.047	0.045
Federal government should increase defense spending	-0.061	-0.003	0.022
Immigrants who enter illegally should be allowed to stay	0.041	0.024	0.076
Federal government should ensure health care for all	0.034	-0.011	0.017
Support the Black Lives Matter movement	-0.014	0.072	0.091
Abortion should be illegal	0.013	0.001	-0.025
Federal government should reduce wealth gap	-0.012	0.028	-0.048

Results – Predicting Senate Vote Choice



Senate Vote Choice (1=David Perdue, Republican; 2=Neither; 3=Jon Ossoff, Democrat)	KP	Opt-in	RICS
Adjusted R ² of Model	80.8%	73.1%	65.3%
	Beta	Beta	Beta
Party Identification - 7 category (1=Strong Republican; 7=Strong Democrat)	0.566	0.545	0.382
President Trump Approval - 7 category	-0.187	-0.261	-0.332
Federal government should ensure health care for all	0.178	-0.009	0.016
Federal government should reduce wealth gap	0.048	0.03	0.026
Federal government should increase defense spending	-0.048	-0.012	0.027
Abortion should be illegal	0.021	-0.04	-0.028
Immigrants who enter illegally should be allowed to stay	0.018	0.016	0.041
Important to control gun ownership	-0.009	0.041	0.055
Support the Black Lives Matter movement	-0.006	0.05	0.084

Discussion

Conclusions and Discussion



We found that, while the models identified generally common strong predictors that had similar strength and direction of effect, some weaker predictors were more variable in their influence across the dependent variables and in the different samples.

Similarities across samples:

- **Vote likelihood** – All three samples have the same top three predictors of likelihood to vote
- **Political party ID** – The models have the same top predictor in common; the other predictors are more variable in their predictive utility
- **Presidential candidate choice** – All three samples have the same top two predictors of presidential candidate choice
- **Senate candidate choice** – All three samples have the same top two predictors of senate candidate choice

Conclusions and Discussion



Generally, we were able to replicate regression models across samples to some extent, especially across KP and opt-in samples. RICS, however, may suffer from having more noise in the measurements, or variability in how well the predictors predict the dependent variables.

Conclusions and Discussion



These findings are consistent with what we saw in the study last year (Li, Thomas, Lien, & Barlas 2021), in which the average bias was higher for RICS than it was for KP or opt-in samples.

This, as well as finding RICS to have lower predictive utility in these regression models, can be considered an indicator of sample quality. While RICS can be a way to capture hard-to-reach respondents in surveys for targeted geographies, it may bring more noise to the data compared to probability-based and opt-in non-probability online samples.

Thank you!

Madeline Gosz

madeline.gosz@ipsos.com