Elections and Polling Methods

How are polls carried out?

Grinnell College

Politics

Failed Polls in 2016 Call Into Question Profession's Precepts
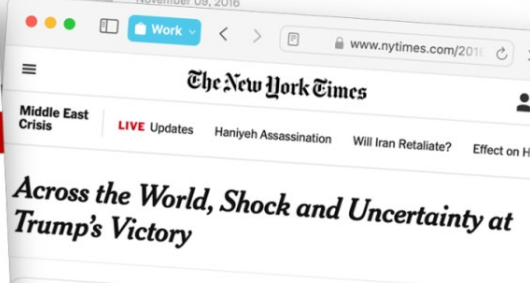
- Electorate is tougher to gauge in world of mobile telephon[...]
- Seeing mortal peril for 'a business that's told us so much'

Analytics And Data Science

Why Pollsters Were Completely and Utterly Wrong

by Dan Cassino

November 09, 2016

NEWS

Reality Check: Should we give up on election polling?

⏱ 11 November 2016

Middle East Crisis · LIVE Updates · Haniyeh Assassination · Will Iran Retaliate? · Effect on H[...]

Across the World, Shock and Uncertainty at Trump's Victory

Pew Research Center illustration

Polling Outline

Polling is **complicated**

These are a summary of some important things about polling mixed in with ideas we have seen covered in class already.

Most of the new stuff I am taking from articles published by:

▶ Pew Research Center

▶ FiveThirtyEight.com

Background

How US elections work (extremely simplified):

Every 4 years there is a presidential election
- ▶ president and VP are elected
- ▶ some senate seats are elected (# varies)
- ▶ all house seats are elected (438 total)
- ▶ many other positions at the state level are elected

Every 2 years (between presidential elections)
- ▶ some senate seats are elected (# varies)
- ▶ all house seats are elected (438 total)
- ▶ many other positions at the state level are elected
- ▶ frequently called 'mid-term' elections

Sometimes there are also amendments to state laws that are voted on too.

Background

In 2016 and 2020, polling did not do a very good job (in many people's opinions) of predicting the presidential election results.

In both years, polls underestimated the strength of Republican candidates.

Polls generally performed better in the mid-term elections in 2018 and 2022.

Big Things in Polling

In my opinion, there are really 2 big areas of concern in polling. But there is much that goes into each.
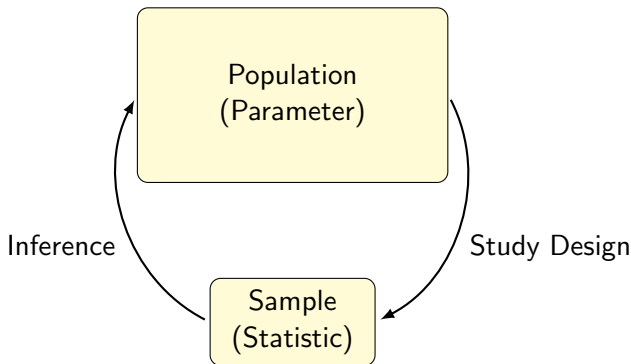
1. Sample Selection (hard)
   - How do we pick people
   - Representative?

2. Estimation (easy)
   - confidence intervals
   - frequent use of the term Margin of Error (ME)

Review – The Statistical Framework

▶ almost always the parameter and statistic of interest in polls is a proportion (p) or difference in proportions ($p_1 - p_2$)

▶ we can use confidence interval stuff we covered to estimate these

Review – Sample Selection

How do we select people?

We want our sample to be **representative** of our population.

- ▶ this means that our sample is nearly the same as our population, only smaller
    - ▶ i.e.: same proportions M/F, same age/ethnic demographics

- ▶ a representative sample allows us to generalize our results from the sample to the pop.

Review – How do we select people?

Random Sample
We can choose people at random from our population to reduce the chances of getting a biased sample.

- ▶ usually the best way to get a representative sample
- ▶ allows us to **generalize** from our sample to the pop.
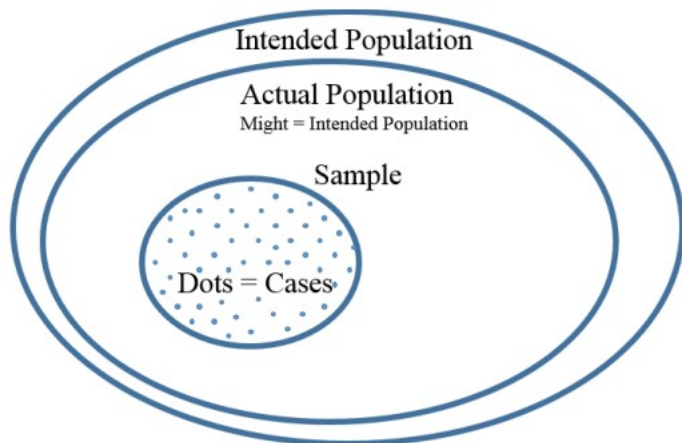
Sample Size (n = ?)

- ▶ number of people we survey is important (more people = more info)
- ▶ sample size is very important, but proportion of pop. surveyed is not
- ▶ affects ME's and CI's we've seen before
- ▶ you can get good results using a sample of 1000 people for a pop. of 300million

Review – Sampling Frame

Sampling Frame – List of people we have access to sample from.

▶ What happens when certain groups of people aren't fully representative in the sampling frame?

Review – Intended vs. Actual population

Intended Population

Actual Population
Might = Intended Population

Sample

Dots = Cases

source: Dr. Ziegler's Stat 104 notes (ISU)

Poll Accuracy

Polls were historically accurate in 2021-22

Weighted-average error of polls in the final 21 days* before presidential primary and presidential, Senate, House and gubernatorial general elections since 1998

CYCLE	PRESIDENT		SENATE	HOUSE	GOV.	COMBINED
	PRIMARY	GENERAL				
1998	–	–	7.5	7.1	8.1	7.7
1999-2000	7.9	4.4	6.0	4.3	4.9	5.5
2001-02	–	–	5.5	5.6	5.2	5.4
2003-04	7.0	3.3	5.3	5.8	5.5	4.8
2005-06	–	–	5.2	6.5	5.1	5.7
2007-08	7.7	3.5	4.7	5.9	4.4	5.5
2009-10	–	–	4.9	7.0	4.7	5.8
2011-12	8.9	3.7	4.7	5.5	4.9	5.3
2013-14	–	–	5.3	6.8	4.5	5.3
2015-16	10.2	4.9	5.0	5.8	5.4	6.8
2017-18	–	–	4.2	4.9	5.2	4.9
2019-20	10.2	5.0	5.8	6.5	6.4	6.3
2021-22	**–**	**–**	**4.8**	**4.0**	**5.1**	**4.8**
All years	9.2	4.3	5.4	6.1	5.4	6.0

Source: FiveThirtyEight.com

Changes to the Polling Scene

"...[the] number of active polling organizations has grown significantly, indicating that there are fewer barriers to entry into the polling field. The number of organizations that conduct national election polls more than doubled between 2000 and 2022."

- ▶ issues? (more on this later too)

"This growth has been driven largely by pollsters using inexpensive opt-in sampling methods. But previous Pew Research Center analyses have demonstrated how surveys that use [non-random] sampling may have errors twice as large, on average, as those that use [random] sampling."
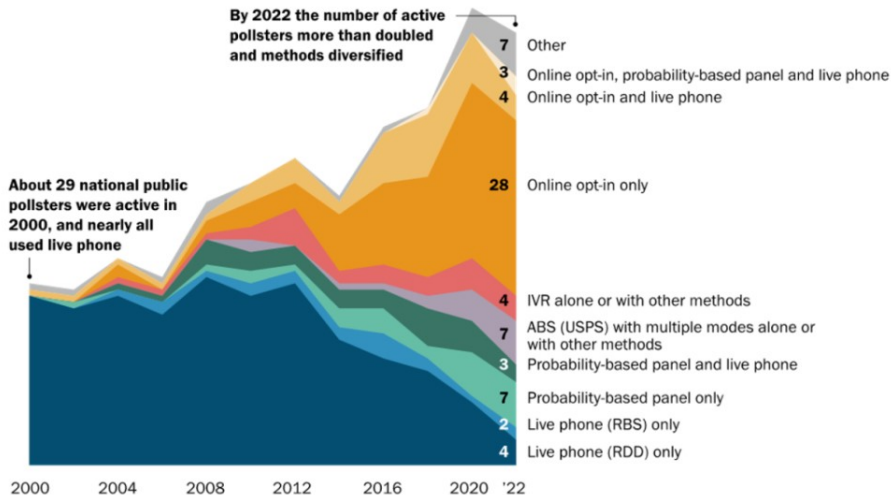
Polling Methods

How have surveys been conducted mostly until the last few years?

- ▶ land-line phone surveys
- ▶ cell-phone surveys
- ▶ internet
- ▶ emails
- ▶ 'snail' mail

Polling Methods

Polling has entered a period of unprecedented diversity in methods

Number of national public pollsters in the U.S. using method(s)

By 2022 the number of active pollsters more than doubled and methods diversified

- 7 — Other
- 3 — Online opt-in, probability-based panel and live phone
- 4 — Online opt-in and live phone
- 28 — Online opt-in only

About 29 national public pollsters were active in 2000, and nearly all used live phone

- 4 — IVR alone or with other methods
- 7 — ABS (USPS) with multiple modes alone or with other methods
- 3 — Probability-based panel and live phone
- 7 — Probability-based panel only
- 2 — Live phone (RBS) only
- 4 — Live phone (RDD) only

2000   2004   2008   2012   2016   2020   '22

Changes to the Polling Scene

"… many of the more prominent polling organizations that use [random] sampling – including Pew Research Center – have shifted from conducting polls primarily by telephone to using online methods, or some combination of online, mail and telephone. The result is that polling methodologies are far more diverse now than in the past."

Margin of Error

The Margin of Error was used in CI's to quantify uncertainty in our estimation, but...

Sampling error is not the only kind of polling error

Error from ...	Error name	Reflected in margin of error
Excluding parts of the population	Noncoverage	No
Low response rates from certain groups	Nonresponse	No
People misunderstanding the question or misreporting their opinions	Measurement	No
Interviewing a sample rather than entire population	Sampling	Yes

PEW RESEARCH CENTER

Selecting People for Surveys

MAJOR issues we've seen
- under-representation in samples
- non-response bias

How do we account for these and get accurate results?

Survey Weighting

Weighting is a term for adjusting our survey sample so that it is more representative of the population.

- ▶ we give more 'weight' to groups that are under-represented
- ▶ idea is to even out the fact that there are less people in this group than we should have in the sample

"Historically, public opinion researchers have adjusted their data using a core set of demographic variables to correct imbalances between the survey sample and the population."

Survey Weighting

Common demographic adjustments

- ▶ Age
- ▶ Race/Ethnicity
- ▶ Sex/Gender
- ▶ Education

"But there is a growing realization among survey researchers that weighting a poll on just a few variables like age, race and gender is insufficient for getting accurate results. Some groups of people – such as older adults and college graduates – are more likely to take surveys, which can lead to errors that are too sizable for a simple three- or four-variable adjustment to work well."

Survey Weighting

American Trends Panel weighting dimensions

Variable	Benchmark source
Age (detailed)	2022 American Community Survey (ACS)
Age x Gender	
Education x Gender	
Education x Age	
Race/Ethnicity x Education	
Black (alone or in combination) x Hispanic	
Born inside vs. outside the U.S. among Hispanics and Asian Americans	
Years lived in the U.S.	
Census region x Metropolitan status	
Volunteerism	2021 CPS Volunteering & Civic Life Supplement
Party affiliation x Voter registration	2020 CPS Voting and Registration Supplement
Party affiliation x Race/Ethnicity	2024 National Public Opinion Reference Survey (NPORS)
Frequency of internet use	
Religious affiliation	

Note: Estimates from the ACS are based on noninstitutionalized adults. Voter registration is calculated using procedures from Hur, Achen (2013) and rescaled to include the total U.S. adult population.

PEW RESEARCH CENTER

Who Shows up to Vote?

"Predicting who will vote is critical – and difficult. Preelection polls face one crucial challenge that routine opinion polls do not: determining who of the people surveyed will actually cast a ballot."

None of the polling matters if those who show up to vote are different than the group we surveyed.

▶ pollsters make their own educated guesses about turnout

▶ based on historical data

▶ based on measures of voter enthusiasm

▶ EXTREMELY hard to get this right

Poll Timing

"Public opinion on most issues is remarkably stable, so you don't necessarily need a recent poll about an issue to get a sense of what people think about it. But dramatic events can and do change public opinion, especially when people are first learning about a new topic. For example, polls this summer saw notable changes in voter attitudes following Joe Biden's withdrawal from the presidential race."

- ▶ Polls taken right after an event may not pick up shifts in opinion
- ▶ Dramatic events can affect voter enthusiasm $\rightarrow$ turnout
- ▶ opinion and enthusiasm have a complicated relationship $\rightarrow$ makes accurate polling even harder
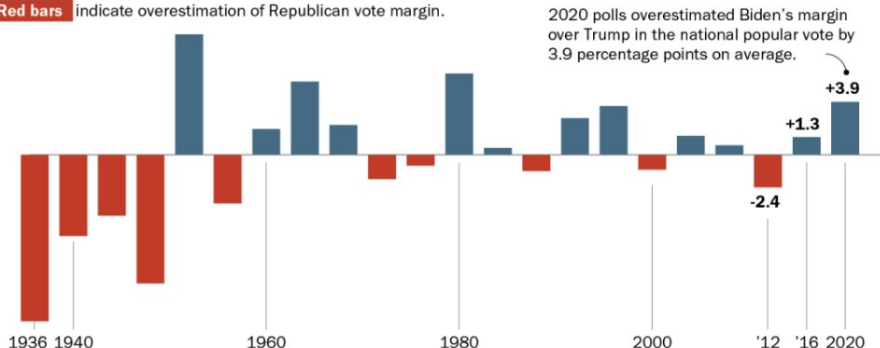
Pew Research Center illustration

Polling Accuracy

Polling errors in U.S. presidential elections

Bars represent average signed error

Blue bars indicate overestimation of Democratic vote margin.
Red bars indicate overestimation of Republican vote margin.

2020 polls overestimated Biden's margin over Trump in the national popular vote by 3.9 percentage points on average.

+3.9

+1.3

-2.4

1936 1940   1960   1980   2000   '12 '16 2020

Note: The average signed error is the difference between the actual margin separating the candidates in the general election and the average margin in the polls. A negative error means that the Republican candidate's margin was overstated and a positive error means the Democratic candidate's margin was overstated.
Source: American Association for Public Opinion Research (AAPOR) Task Force on 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls

Polling Accuracy

American Association for Public Opinion Research (AAPOR):
"...the 2020 polls featured polling error of an unusual magnitude: It was the highest in 40 years for the national popular vote and the highest in at least 20 years for state-level estimates of the vote in presidential, senatorial, and gubernatorial contests."

Frame Title

PRC: "The 2022 midterms saw generally accurate polling, despite a wave of partisan polls predicting a broad Republican victory. In fact... Moreover, a handful of contrarian polls that predicted a 2022 "red wave" largely washed out when the votes were tallied. In sum, if we focus on polling in the most recent national election, there's plenty of reason to be encouraged."

FiveThirtyEight: "...polls were more accurate in 2022 than in any cycle since at least 1998, with almost no bias toward either party."

Sampling Error (again)

Back to error... people think polls are more accurate than they really are

Sampling error is not the only kind of polling error

Error from ...	Error name	Reflected in margin of error
Excluding parts of the population	Noncoverage	No
Low response rates from certain groups	Nonresponse	No
People misunderstanding the question or misreporting their opinions	Measurement	No
Interviewing a sample rather than entire population	Sampling	Yes

PEW RESEARCH CENTER

Polling Transparency

Transparency in how a poll was conducted is associated with better accuracy.

"Participation in these transparency efforts does not guarantee that a poll is rigorous, but it is undoubtedly a positive signal. Transparency in polling means disclosing essential information, including the poll's sponsor, the data collection firm, where and how participants were selected, modes of interview, field dates, sample size, question wording, and weighting procedures."

Polling Effects on an Election

Results from polls before an election can influence who turns up to vote.

"Following the 2016 election, many people wondered whether the pervasive forecasts that seemed to all but guarantee a Hillary Clinton victory – two modelers put her chances at 99% – led some would-be voters to conclude that the race was effectively over and that their vote would not make a difference."

Further Complications

The presidential election is note determined by national popular vote!

- ▶ each state has a number of 'electoral votes' = # senators + # representatives
- ▶ for most states whoever wins the state popular vote gets that state's EV's
- ▶ 538 EVs up for grabs, need 270 to win president race
  - ▸ barring other legal shenanigans that can throw a wrench into this

Sources

Pew Research Center: https://www.pewresearch.org/short-reads/2024/08/28/key-things-to-know-about-us-election-polling-in-2024/

FiveThirtyEight.com
https://fivethirtyeight.com/features/2022-election-polling-accuracy/

AAPOR: https://aapor.org/wp-content/uploads/2022/11/AAPOR-Task-Force-on-2020-Pre-Election-Polling_Report-FNL.pdf