

# **Polling in the 2020 Election**

**Clyde Tucker**

# Outline

- **Issues in Pre-election Polls**
- **Electoral Choice in Survey Research**
- **The Current Political Environment**
- **Review of Current Polls**
- **Changing American Political Culture**

# Issues in Pre-election Polls

- Recognize that pre-election polls are just a snapshot in time and may not reflect election results. Only the final polls would be useful for election predictions (more on this later).
- Polls are most useful for understanding which voters support which candidate and why.
- Questionnaire Design and Administration
  - Horserace question should be first, before presidential approval and issue questions.
  - House effects
    - Reputation
    - Political bias—incorporated at any stage
    - Methodological sophistication
    - Time and Funding
    - Probability or non-probability sampling
    - Mode (interviewer administered by phone, robo calls, internet either based on random sample or opt in)
- Cell Phone Only Households
  - About half of households are cell only
  - Increases to over 60% when “cell mostly”
  - Excluding cell-only households restricts access to young voters

# Advances in Communication Technology

- **Cell Phone Only Households**
  - About 60% of households are cell only
  - Another 15% “cell mostly”
  - Only 6% landline only
  - Over 80% of adults have smartphones
  - 15-20% smartphone dependent
  - Blacks and Hispanics use smartphones more than Whites
- **Internet Usage**
  - 90% of adults use the internet
  - Over 75% of households have broadband
  - Even 60% of those 65+ have broadband
  - About 80% of Whites have broadband and 60-65% of Blacks and Hispanics also have broadband

# How Technology has Changed Survey Administration

- **In-person interviewing all but impossible except for government surveys.**
- **Can probably get over 95% coverage using either smartphones or the internet.**
- **So creating an adequate sampling frame can be done unless need to cover the landline only households (probably less likely to vote).**
- **Which method of administration will be used to contact potential respondents will depend on cost and personal preference.**
- **Probably best to use both smartphones and the internet, but how to achieve a random sample?**
- **Both technologies present challenges.**

# Determining Likely Voters

- **Likely voter screens are usually based on answers to a series of questions, including political interest, past voting history, length of time in the community, and likelihood of voting.**
- **The likely voter screen is used either to exclude respondents with a low score or to assign a probability of voting to each respondent, often aided by information from voter validation studies.**

# Problems with Identifying Likely Voters

- **Inaccurate predictions of likelihood of voting**
  - Using self reports is prone to error
  - Some questions may be worse than others
  - Assigning a probability of voting may be better than having cutoffs
  
- **Missing some likely voters due to nonresponse**

# Allocating Undecideds

- **There are essentially three ways to allocate the undecideds:**
  - **Evenly divide them between the candidates (most often the two leading ones)**
  - **Assign undecideds proportionately based upon the proportions for each candidate among the decided voters**
  - **Assign an undecided voter based on the characteristics of the voter (models will differ depending on the analyst)**
- **The correct allocation is probably dependent on the specific election context.**



# Sampling Error

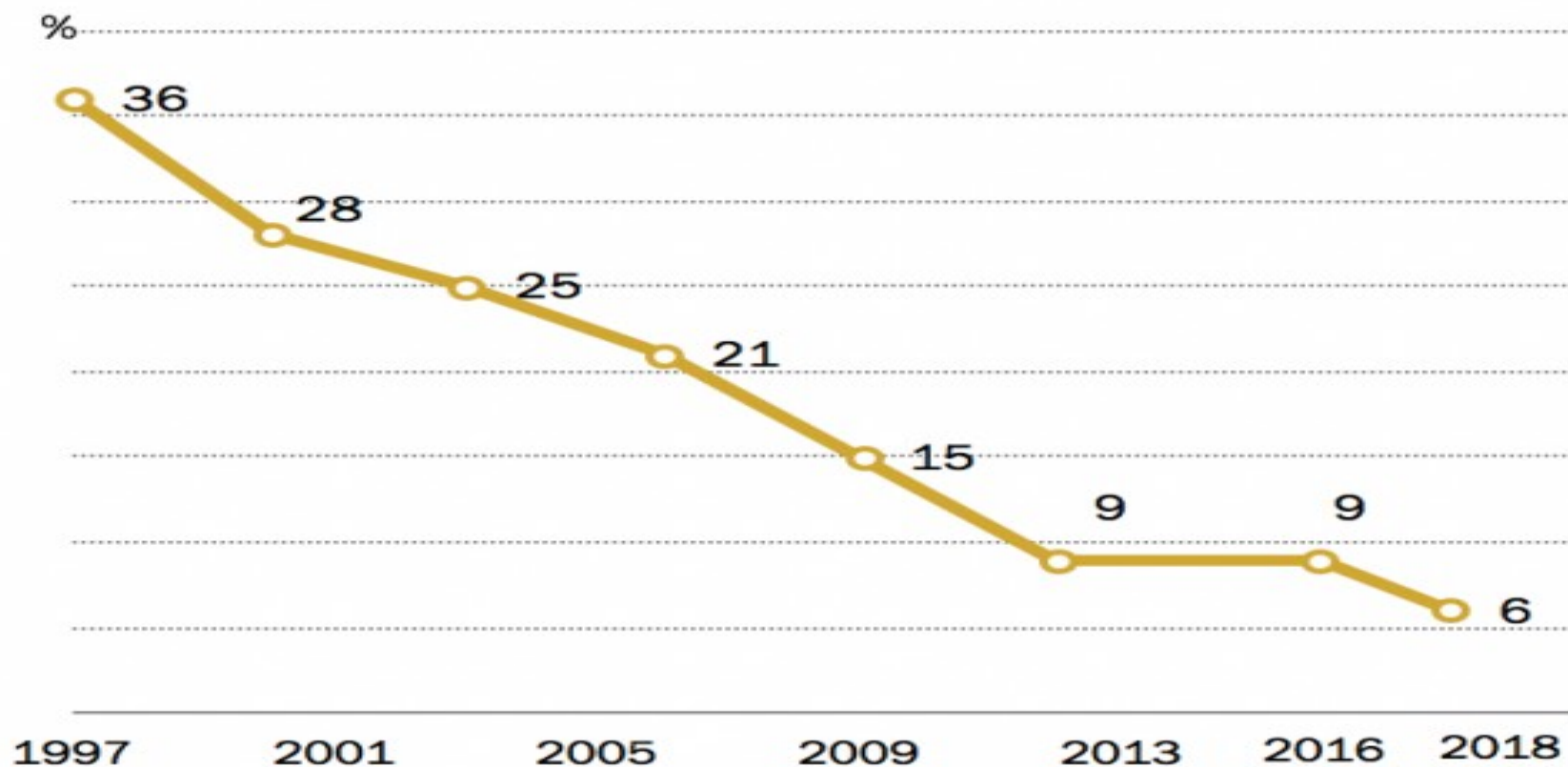
- **The error in election predictions includes both sampling error and bias.**
- **Many reports on election polls include a statement on margin of error, but this is usually based only on a calculation of sampling error.**
- **Even if a phone poll starts with a random sample of numbers; given high levels of nonresponse, it is unlikely that the effective sample will be random, making it impossible to accurately measure the sampling error.**
- **Although an internet poll could be based on creating a random panel of respondents, an expensive proposition, it is likely that responses from this panel over time will involve panel members choosing to opt in. This will lead to a convenience sample from which a measure of sampling error cannot be calculated.**
- **Thus, it is likely any measure of sampling error attached to a poll will be inaccurate. But the public wants a point estimate anyway. However, there may be some virtue to at least begin with a random sample.**

# **The Nonresponse Problem**

- Bias is largely the result of nonresponse.**
- All polls have it to some degree, and it is difficult to calculate.**
- Over the last four decades, response rates in almost all non-government surveys (like election polls) have been in rapid decline.**
- Contributing to this decline have been the following:**
  - A growing number of single-person households (28%)**
  - Changing technologies make contacting and interviewing respondents more difficult.**
  - The electorate is inundated with surveys or selling under the guise of a survey.**
  - Today's electorate has a much less positive view of the government and the media.**

# After brief plateau, telephone survey response rates have fallen again

*Response rate by year (%)*



Note: Response rate is AAPOR RR3. Only landlines sampled 1997-2006. Rates are typical for surveys conducted in each year.

Source: Pew Research Center telephone surveys conducted 1997-2018.

**PEW RESEARCH CENTER**

# Coping with Nonresponse 1

- **Given some very low response rates, bias can be quite large.**
- **It would be great if the nonresponse was random, but it usually is not.**
- **The bias in election polls most often results in an overstatement of Democratic percentages. This occurs in all election surveys—telephone, internet, tracking and exit polls.**
- **The first method for reducing nonresponse bias is to repeatedly contact the nonrespondents, but this costs money and time and may not be very effective.**
- **Other methods involve statistical adjustments, some more complicated and time consuming than others.**

# Coping with Nonresponse 2

- **The most common form of statistical adjustment is modifying survey weights.**
- **Post-stratification involves multiplying the initial respondent weights in different population subgroups by a ratio of known subgroup proportions in the population divided by the subgroup proportions in the effective sample.**
- **Calibration reweights the sample to match the known population means of certain population characteristics.**
- **Another reweighting technique uses the inverse of the propensity for respondents to participate in the survey.**

# Coping with Nonresponse 3

- In the first two cases, the survey methodologists must know which respondent characteristics are most related to nonresponse bias and have the correct population information.
- In the case of propensity weighting, the survey methodologist needs to know which survey variables are useful for determining the propensities and be able to verify that these propensities hold for the entire population.
- Unfortunately, nonresponse is often related to voter preference or some issue position. In these cases, population information is unlikely to be known since nonresponse is related to the variable of substantive interest, making nonresponse not missing at random.
- Also, the election survey needs to have some minimum number of respondents in the required subgroups.
- If the above conditions are not met, reweighting the sample might do more harm than good.

# Predictive Models 1

- **As Kenett, Pfeffermann, and Steinberg (2018) make clear, most users of polls are interested in a prediction, and, as mentioned earlier, only the last poll will be useful for that.**
- **When using the final polls for prediction, a weighted average of the last polls is more accurate than results from a single poll. RealClearPolitics assigns a weight to each poll based on its historical accuracy.**
- **Kenet and his colleagues do note the recent development of prediction models.**
- **Nate Silver has the most famous one for predicting presidential elections. It is used to predict which presidential candidate will win every state.**
- **Many simulations are run while altering assumptions. The winner is the candidate who wins the majority of simulated races (with 270 or more electoral votes) and the result is expressed as the proportion of time the candidate won.**
- **Silver begins early in the election year with a model that includes the demographics of each state as well as past voting history, estimates of turnout, and historical biases in the state's polls.**

# Predictive Models 2

- The polling results also are added to the model as time goes on. In fact, the polls receive greater weight as the election draws near and the weights on the other variables are reduced. As with RealClearPolitics, the weights are based on how accurate the poll is expected to be.
- Silver has to use state polls in his model, which is tricky. There is little polling in some states, and those doing the polling may not be very sophisticated when it comes to survey methodology.
- In a time when everyone is recommending the use of multiple data sources, note that this is exactly what Silver does when building his model.
- Also, as Kenett et al. point out, successful presidential campaigns are doing the same thing. Obama appears to have bypassed the DNC and develop his own team of data scientists who combined multiple data sources and sophisticated analytical methods to guide decision making during the campaign. The Trump campaign did the same.
- But what about the results being 90-10 or 70-30? What do these mean to the average person? And what about fifty-fifty?



# What Are Pollsters To Do ? 1

- Kenet et al. suggest improvements in election surveys based on an information quality framework. Below are my own recommendations, which have some relationship theirs . Of course, in both cases, some will require a good deal of time, money, and expertise. Unfortunately, only campaigns may have enough of these necessary resources.
- I suggest the following:
  - Improve prediction by using multiple data sources and not just polling results. For instance, demographic information (e.g., age, race, sex and education), geographic information (e.g., place size and changes in population characteristics), and past voting history could be used to develop predictions by state or maybe by regions. Like Silver's model, the results would be built up to estimate candidate percentages of the vote at the national level.
  - Improve the measures of likelihood of voting using voting records and historical trends in turnout at appropriate levels of geography. These measures could be enhanced by using measures of political interest at the same geographic levels.

# What Are Pollsters To Do ? 2

- **Develop a ground game that compliments polling results by tracking trends in social media (e.g., sentiment measures) while gathering what personal information may be available for these social media users. This information could not only be used to verify or improve polling estimates, but, more specifically, it could inform the allocation of undecideds.**
- **Using the polling data, determine which voter characteristics (as well as interactions between these characteristics) that are most related to voter preference. The results would provide information for improving weighting to reduce nonresponse bias.**

# Current Political Environment

- **Pandemic leading to much greater mail-in voting and concerns about valid results and when they will be known.**
- **Possible interference by outsiders.**
- **High interest with above average turnout expected.**
- **Two strong opposing camps.**
- **Black Lives Matter and Black turnout expected to be larger than 2016.**
- **White suburbs a battleground.**
- **Women more likely to vote for Biden than Trump.**
- **Democrats targeting young voters.**
- **How large a lead in the national vote do Democrats need to win in the electoral college (maybe 5%)?**

# National Presidential Polls

	Biden	Trump	Spread
<b>RCP Average 9/21/20</b>	<b>49.5</b>	<b>43.0</b>	<b>6.5D</b>
<b>YouGov</b>	<b>51</b>	<b>42</b>	<b>9D</b>
<b>The Hill</b>	<b>45</b>	<b>39</b>	<b>6D</b>
<b>Reuters</b>	<b>50</b>	<b>41</b>	<b>9D</b>
<b>Rasmussen</b>	<b>46</b>	<b>47</b>	<b>1R</b>
<b>USC</b>	<b>51</b>	<b>42</b>	<b>9D</b>
<b>NBC/WSJ</b>	<b>51</b>	<b>43</b>	<b>8D</b>
<b>JTN/RMG</b>	<b>48</b>	<b>43</b>	<b>9D</b>
<b>IBD/TIPP</b>	<b>50</b>	<b>44</b>	<b>6D</b>
<b>NPR/PBS</b>	<b>52</b>	<b>43</b>	<b>8D</b>
<b>FOX News</b>	<b>51</b>	<b>46</b>	<b>5D</b>

# Battleground States (RCP Average)

	<b>Biden</b>	<b>Trump</b>	<b>Spread</b>
<b>WI</b>	<b>50.1</b>	<b>43.4</b>	<b>6.7D</b>
<b>FL</b>	<b>48.6</b>	<b>47.0</b>	<b>1.6D</b>
<b>NC</b>	<b>47.3</b>	<b>46.4</b>	<b>1.6D</b>
<b>MI</b>	<b>47.8</b>	<b>43.0</b>	<b>4.8D</b>
<b>AZ</b>	<b>49.2</b>	<b>44.2</b>	<b>5.0D</b>
<b>OH</b>	<b>46.7</b>	<b>44.3</b>	<b>2.4D</b>
<b>VA</b>	<b>53.0</b>	<b>39.0</b>	<b>14.0D</b>
<b>NH</b>	<b>48.2</b>	<b>42.5</b>	<b>5.5D</b>
<b>PA</b>	<b>48.7</b>	<b>44.7</b>	<b>4.0D</b>
<b>IA</b>	<b>45.0</b>	<b>46.7</b>	<b>1.7R</b>
<b>CO</b>	<b>55.0</b>	<b>45.0</b>	<b>10.0D</b>
<b>NV</b>	<b>46.5</b>	<b>40.5</b>	<b>6.0D</b>
<b>ME</b>	<b>52.2</b>	<b>38.4</b>	<b>13.8D</b>
<b>MN</b>	<b>51.6</b>	<b>41.4</b>	<b>10.2D</b>
<b>TX</b>	<b>43.8</b>	<b>47.3</b>	<b>3.5R</b>
<b>GA</b>	<b>45.0</b>	<b>46.3</b>	<b>1.3R</b>
<b>NM</b>	<b>53.5</b>	<b>39.0</b>	<b>14.5D</b>

## Senate Races (RCP Average)

	Dem	Rep	Spread
AZ	Kelly 50.0	McSally (I) 43.3	6.7D
NC	Cunningham 46.1	Tillis (I) 42.5	3.6D
MI	Peters (I) 46.8	James 43.2	3.6D
ME	Gideon 47.2	Collins (I) 41.0	6.0D
IA	Greenfield 44.7	Ernst (I) 44.3	0.4D
MT	Bullock 44.7	Daines (I) 46.3	1.6R
CO	Hickenlooper --	Gardner (I) --	--
MN	Smith (I) 48.0	Lewis 39.3	8.7D
AL	Jones (I) 42.0	Tuberville 50.0	8.0R
GA	Ossoff 42.5	Perdue (I) 46.8	4.3R

# Presidential Election Landslides

**1964**

**Johnson**

**Goldwater**

**61%**

**39%**

**1972**

**Nixon**

**McGovern**

**61%**

**38%**

**1984**

**Reagan**

**Mondale**

**59%**

**41%**

# Changes in the American Political Culture

- **The Powell memo**
- **The election of Ronald Reagan**
- **Growing distrust of American institutions and less civic engagement**
- **Reduced tax revenue and collapsing public infrastructure**
- **The triumph of individual will over the collective will**
- **Rapid technological change and fear of the future**