POLI210: Political Science Research Methods

Lecture 9.1: Survey research II

Olivier Bergeron-Boutin

2021

Boring admin stuff

- Thanks for completing mid-term survey; my takeaways...
  - Readings are not helpful
    - Textbook readings or other readings? (POLLING)
  - In-person labs are too fast-paced
    - I've talked to the TAs about this
  - Assignment instructions are sometimes unclear
  - Disconnect between classes and labs
- Quiz 1: out now!
 - You have until Wednesday 11:59PM

What you think of your learning

```r
library(ggplot2)
library(tidyverse)
midcourse <- read.csv("midcourse.csv") %>% slice(-1:-2) %>%
  mutate(effective = factor(
    Q4, levels = c("Not effective at all", "Slightly effective",
                   "Moderately effective", "Very effective", "Extremely effective")))
ggplot(midcourse, aes(x = effective)) +
  geom_bar(fill = "steel blue", colour = "black") +
  scale_x_discrete(drop = FALSE) +
  theme_bw(base_size = 19)
```
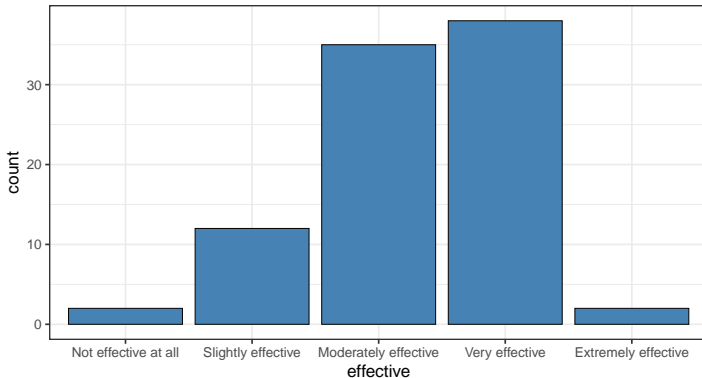
What you think of the workload

```
midcourse %>%
  mutate(workload = factor(
    Q9, levels = c("Light", "Moderate", "Heavy but manageable", "Heavy and unmanageable"))) %>%
ggplot(aes(x = workload)) +
  geom_bar(fill = "steel blue", colour = "black") +
  scale_x_discrete(drop = FALSE) +
  theme_bw(base_size = 19)
```

Issues in survey design
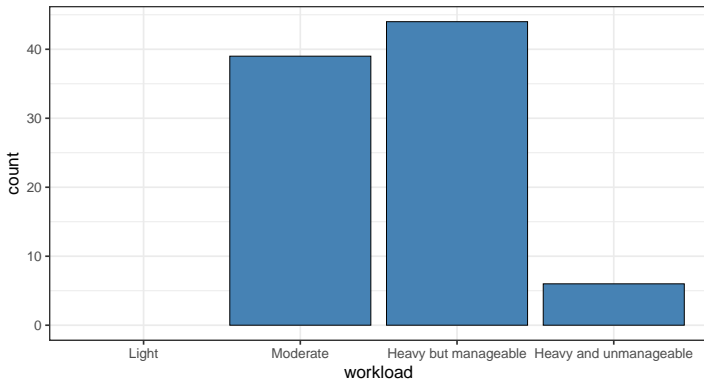
Let's assume that your sampling is unassailable:

- Your sampling frame is a perfect representation of the population
- You randomly draw a large-enough sample from this sampling frame
- Everyone who is contacted agrees to answer

Things can still go wrong! 3 areas of concern:

- Question wording
- Inattentiveness
- Social desirability bias

Question wording

How we phrase our questions matter

- Wordings that appear similar to us may elicit different responses
- An example we've seen: climate change vs global warming
    - Global warming: 44% of Republicans agree has been happening
    - Climate change: 60% of Republicans agree has been happening[1]
    - In both cases, we're trying to measure the same underlying concept
- We want **unambiguous and simple** question wording

---

[1]See Schuldt, Konrath, and Schwarz (Schuldt, Konrath, and Schwarz 2011)

"Should third-party advertising be allowed during the federal elections?"

"Should third-party advertising be allowed during the federal elections?"

- The language may be unfamiliar to respondents – what's "third-party advertising?"
- Solution: use simpler language or provide short introduction
- For instance: "Should political advertising by unions and corporations be allowed during the federal elections?"

"Did you vote in the last election?"

Question wording problems

"Should third-party advertising be allowed during the federal elections?"

- The language may be unfamiliar to respondents – what's "third-party advertising?"
- Solution: use simpler language or provide short introduction
- For instance: "Should political advertising by unions and corporations be allowed during the federal elections?"

"Did you vote in the last election?"

- The wording of the statement is vague
- Federal election? Provincial election? Municipal election?

Question wording problems

Michael Avenatti ✔
@MichaelAvenatti

Time for a poll. Vote as to whether I should a) stay on television and keep disclosing accurate information to the American public or b) get off television and stop disclosing accurate information.

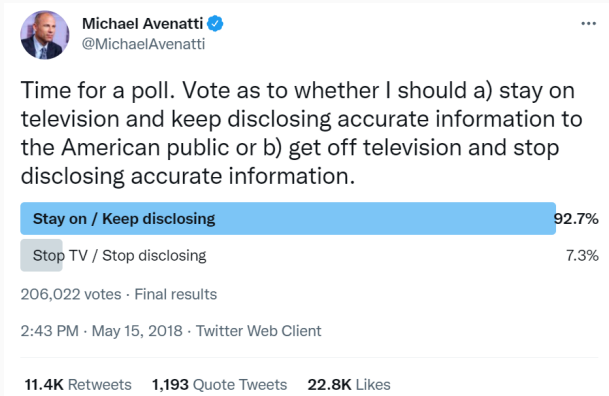|	
Stay on / Keep disclosing	**92.7%**
Stop TV / Stop disclosing	7.3%

206,022 votes · Final results

2:43 PM · May 15, 2018 · Twitter Web Client

11.4K Retweets **1,193** Quote Tweets **22.8K** Likes

Question wording problems

> **Michael Avenatti** ✔
> @MichaelAvenatti
>
> Time for a poll. Vote as to whether I should a) stay on television and keep disclosing accurate information to the American public or b) get off television and stop disclosing accurate information.
>
> | Stay on / Keep disclosing | 92.7% |
> | Stop TV / Stop disclosing | 7.3% |
>
> 206,022 votes · Final results
>
> 2:43 PM · May 15, 2018 · Twitter Web Client
>
> **11.4K** Retweets   **1,193** Quote Tweets   **22.8K** Likes

Double-barelled question

- The question has multiple components
- What if I want him to get off television, but continue disclosing information?

8

"Do you agree that the use of recreational marijuana should be legalized?"

- Leading question: encourages respondents to agree
- Simple fix: "Do you agree or disagree…"

Question wording problems

"Since the beginning of last year – that is since January of 2020 – have you attended a formal or informal meeting organized by yourself, by someone else you know personally, or by a religious, social, civic, governmental or political group to specifically discuss a local, provincial, national, or international issue – for example, neighbourhood crime, housing, schools, social security, electoral reform, global warming, or any other public issue that affects people?"

"Since the beginning of last year – that is since January of 2020 – have you attended a formal or informal meeting organized by yourself, by someone else you know personally, or by a religious, social, civic, governmental or political group to specifically discuss a local, provincial, national, or international issue – for example, neighbourhood crime, housing, schools, social security, electoral reform, global warming, or any other public issue that affects people?"

SO WORDY!!

- Your respondents are not going to read the full question
- Online survey: +/- 20 words
- Face-to-face or phone: a single breath
- Potential trade-off between providing context and increasing inattentiveness

- Some issues are complex and hard to summarize briefly
- One approach: leave the wording as simple as possible
    - "Would you support or oppose changing zoning laws to allow for the construction of multi-family homes?"
    - But what do people know about zoning policy?
    - People will probably answer nonetheless ⤳ non-attitudes
- Another approach: provide contextual information
- "When building new housing, any new construction must follow zoning laws which outline what kind of buildings can be built where. Some are now calling for changes to the zoning laws where they live to allow for the construction of multi-family homes. Supporters of this say that this will drive economic growth as more people will be able to move to high-opportunity regions with good jobs and will allow more Americans the opportunity to get affordable housing of their own, making it easier to start families. Opponents of this say that single-family zoning requirements are important to protect the distinct character of their neighborhoods. They also cite concerns that development could have negative impacts on the environment. Would you support or oppose changing zoning laws to allow for the construction of multi-family homes?"

11

A growing problem with online surveys

- In-person and phone: easy to tell if respondent is "satisficing"
  - "Satisficing": using shortcuts that speed up survey duration
- Online: hard to know exactly what respondents are doing
  - They may not read the question at all
  - Read it very quickly and miss important detail
    - Think again of climate change vs global warming...
  - Walk away from the survey and come back to it hours later
- I'll show you an example from my own research!

Inattentiveness in our conjoint experiment

On this and the next screens you will see pairs of candidates who are competing for a seat in Parliament. For each pair, please choose the candidate that you prefer.

| Candidate A	Candidate B
Male	Male
43	75
Member of Parliament	Mayor
New Democrat	Conservative
Says lockdowns should continue until there are fewer COVID-19 deaths	Says lockdowns should continue until there are fewer COVID-19 deaths
Says that a prime minister should work with Parliament even if it is obstructing his/her policies to combat a pandemic	Says that a prime minister should work with Parliament even if it is obstructing his/her policies to combat a pandemic
Says economic aid to address the COVID-19 crisis should mostly be given to businesses	Says economic aid to address the COVID-19 crisis should ensure a basic income of $1,000 per month for everyone

Which candidate do you prefer?

Candidate A

Candidate B

Figure 1: Example of a conjoint experiment from a 2019 survey (in collaboration with Dietlind Stolle and Elisabeth Gidengil)

Inattentiveness in our conjoint experiment

```
library(ggplot2)
load("C:/Users/olivi/Documents/Grad/McGill Fall 2021/POLI210/lectures/lecture_9.1/c
ggplot(conjoint, aes(x = time_conjoint1)) +
  geom_histogram(fill = "steel blue", col = "black") +
  geom_vline(xintercept = mean(conjoint$time_conjoint1), col = "red", linetype = "d
  theme_bw(base_size = 19)
```
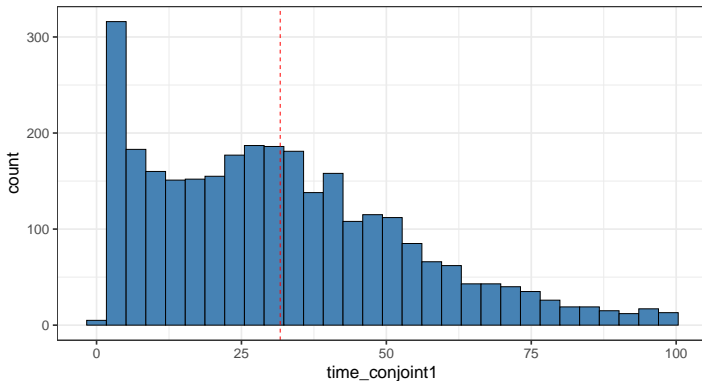
Inattentiveness in our conjoint experiment

```
boxplot(conjoint$time_conjoint1)
```
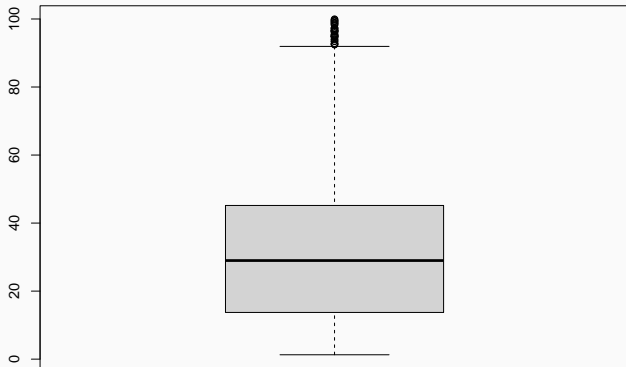
Inattentiveness: the fix

To deal with inattentiveness, survey design matters a lot

- Too long ⤳ respondents tune out
- Too repetitive ⤳ respondents tune out
- The questions are too complex ⤳ respondents tune out

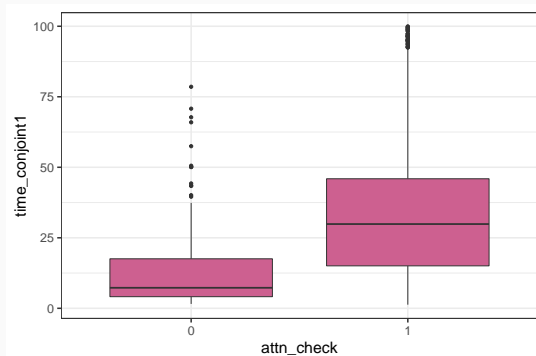Otherwise, **attention checks**

Attention checks in our survey

```
table(conjoint$attn_check)
```

```
##
##    0    1
##  131 2848
```

```
ggplot(data = conjoint, aes(x = attn_check, y = time_conjoint1)) +
  geom_boxplot(fill = "hotpink3") +
  theme_bw(base_size = 19)
```
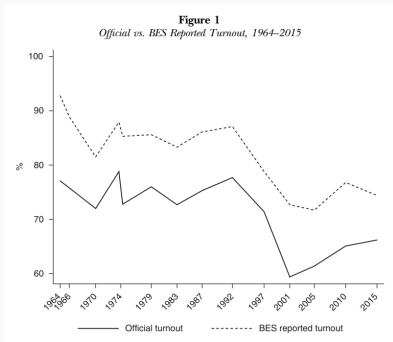
Social desirability bias

Respondents may not give truthful answers because they think of their answer as socially undesirable

- Often related to behavior/attitudes that are thought of as reprehensible
- e.g. support for violence; prejudice towards groups; illegal behavior
- Your respondents will still answer; but their answers will not be truthful
- Your estimate of the "undesirable" behavior/attitude will be **biased downward**

On the contrary, people may lean toward answers that they think are socially desirable

- e.g. turnout; self-reports of volunteering
- Here, your estimate will be **biased upward**

Figure 2: Turnout in British Election Study (Prosser and Mellon 2018)

The British Election Study systematically overestimates turnout

- One cause: social desirability bias
- Another cause: sampling bias!
  - The sorts of people who took the survey are more interested in politics

The state of political polling

Do you think election polls have been getting worse? (POLLING)

Are polls getting worse?

Do you think election polls have been getting worse? (POLLING)

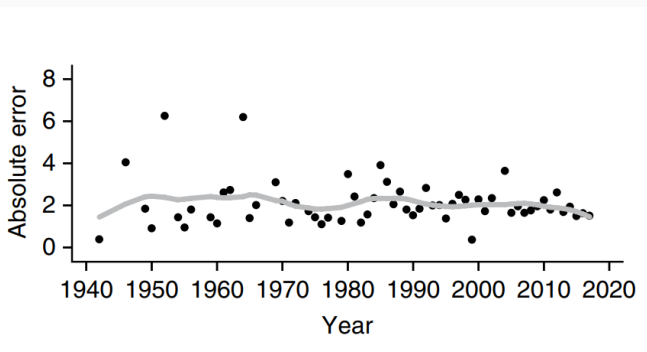

Figure 3: Absolute polling error across time in 45 countries, 1942-2017 (Jennings and Wlezien 2018)
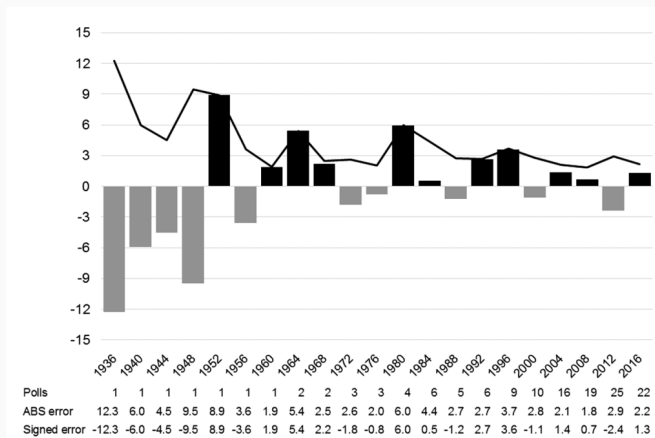
Polling error for party i: $|Poll_i - Election_i|$

Figure 4: Polling error in presidential elections, 1936-2016 (Kennedy et al. 2018)

If things haven't been getting worse, what explains polling misses?

- 2016 US election, Brexit…

2016 US election: national-level polls were okay

- But state-level polls were historically bad
- And because of the Electoral College, the state polls matter a lot!

	2000	2004	2008	2012	2016
Polls	220	231	332	280	422
ABS error	4.6	3.2	3.4	3.5	5.1
Signed error	-1.9	0.8	-0.1	-2.3	3.0
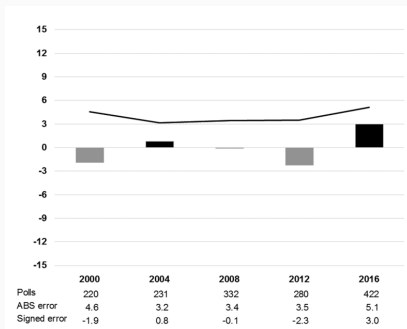
Potential explanations for 2016

2016 US election: some hypotheses (Kennedy et al. 2018)

- Nonresponse bias and deficient weighting
- Late deciders
- Misspecified likely voter models
- The "shy" Trump voters

Keep in mind: highly-paid experts fall prey to these issues!

Nonresponse bias: the people who decline to be interviewed are systematically different

- Solution: give greater weight to respondents from underrepresented groups

Unit	Female	Education	Height
1	1	College	4ft, 3in
2	1	Doctorate	3ft 7in
3	0	High school	4ft, 8in
4	1	Doctorate	5ft, 3in
5	1	Doctorate	4ft, 4in

Is my sample representative?

Nonresponse bias: the people who decline to be interviewed are systematically different

- Solution: give greater weight to respondents from underrepresented groups

Unit	Female	Education	Height
1	1	College	4ft, 3in
2	1	Doctorate	3ft 7in
3	0	High school	4ft, 8in
4	1	Doctorate	5ft, 3in
5	1	Doctorate	4ft, 4in

Is my sample representative?

- Well, depends what **population** we are interested in
- Maybe I'm interested in attitudes among highly-educated dwarves!

Unit	Female	Education	Height
1	1	College	4ft, 3in
2	1	Doctorate	3ft 7in
3	0	High school	4ft, 8in
4	1	Doctorate	5ft, 3in
5	1	Doctorate	4ft, 4in

But I'm probably not interested in dwarves

- The Canadian population:
    - About 50% female
    - Typical person has high school/college diploma
    - Mean height is probably about 5ft, 6in?
- In my sample, women, the highly-educated, and short people are overrepresented because more likely to agree to participate
- One fix: weights

Unit	Female	Education	Height	Weight
1	1	College	4ft, 3in	1
2	1	Doctorate	3ft 7in	1
3	0	High school	4ft, 8in	1
4	1	Doctorate	5ft, 3in	1
5	1	Doctorate	4ft, 4in	1

By default, everyone is weighted equally

- You can think of this as everyone having a weight of 1

To overcome nonresponse bias, I can:

- Give more weight to units whose characteristics are underrepresented (weight > 1)
- Give less weight to units whose characteristics are overrepresented (weight < 1)

Problem #1: Nonresponse bias and deficient weighting

Unit	Female	Education	Height	Weight
1	1	College	4ft, 3in	?
2	1	Doctorate	3ft 7in	?
3	0	High school	4ft, 8in	?
4	1	Doctorate	5ft, 3in	?
5	1	Doctorate	4ft, 4in	?

Who should be given more weight?

Problem #1: Nonresponse bias and deficient weighting

Unit	Female	Education	Height	Weight
1	1	College	4ft, 3in	?
2	1	Doctorate	3ft 7in	?
3	0	High school	4ft, 8in	?
4	1	Doctorate	5ft, 3in	?
5	1	Doctorate	4ft, 4in	?

Who should be given more weight?

- Unit 3 for sure! Males and low-education are underrepresented
- Maybe unit 1? College-educated are underrepresented

Problem #1: Nonresponse bias and deficient weighting

Unit	Female	Education	Height	Weight
1	1	College	4ft, 3in	?
2	1	Doctorate	3ft 7in	?
3	0	High school	4ft, 8in	?
4	1	Doctorate	5ft, 3in	?
5	1	Doctorate	4ft, 4in	?

Who should be given more weight?

- Unit 3 for sure! Males and low-education are underrepresented
- Maybe unit 1? College-educated are underrepresented

Which variables you weigh on is crucial

- You can't possibly weigh on everything
- You want to weigh on the variables that are most correlated with the outcome of interest

So what was the problem in 2016?

- Pollsters would weigh on gender, age, ethnicity, etc.
- But not on education! Why?
    - Education used to have a curvilinear relationship with vote choice
    - The (undersampled) poorly-educated: voted Democrat
    - The (oversampled) well-educated: voted Democrat
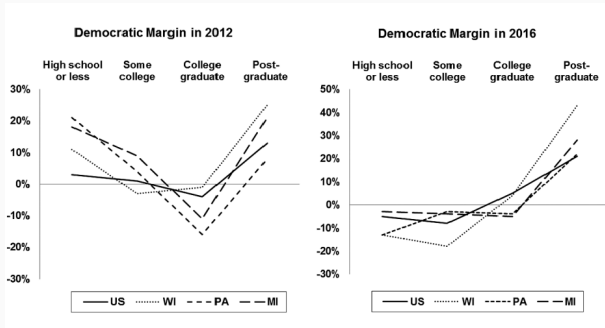
If 2020 polls were not adjusted for education, they would show bigger leads for Biden

Democratic margin in Muhlenberg College/Morning Call presidential election polls of Pennsylvania voters, with and without educational attainment weights (percentage points)
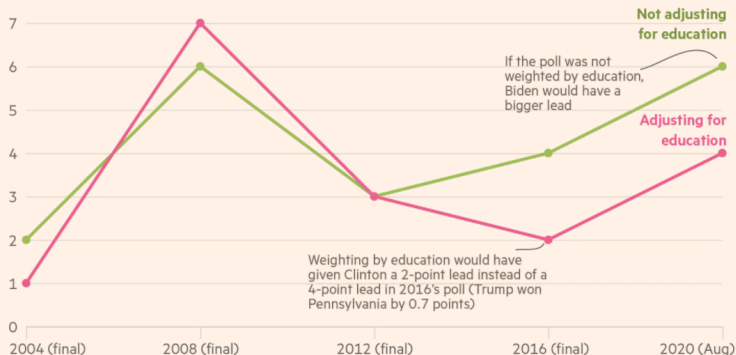
Figure 7: Democratic margin in Pennsylvania polls, with and without education weighting; see Financial Times

Problem #2: late deciders

The closer the poll to the election, the more accurate it tends to be

- People who don't know who to vote for may disproportionately break for one candidate
- Polls can miss these "late deciders"
- At the extreme, someone who decides in the voting booth

2016 featured far more undecided voters than recent past elections

CYCLE	COMBINED UNDECIDED + THIRD PARTY VOTE	
	WITH 100 DAYS TO GO	ON ELECTION DAY
1972	9.3%	8.1%
1976	10.0	8.7
1980	25.5	17.3
1984	7.4	4.7
1988	10.8	5.7
1992	19.1	22.1
1996	19.7	13.3
2000	17.8	8.9
2004	6.4	3.4
2008	13.1	3.6
2012	7.6	4.3
2016	**18.5**	**12.5**

Problem #3: misspecified likely voter models

Not everyone who is polled ends up voting

- I don't want to include people who I think won't vote
    - (if the goal is to predict the election's outcome)
- Using statistical methods that we'll learn about, you can look at a variety of variables…
    - Age, education, political interest…
- And produce a "guess" of whether any given person will vote or not
- This "model" of voter turnout is built using past data
    - e.g. Using data from the most recent Canadian election, I can predict that someone with a college degree is X percentage points more likely to vote than a person with a high school diploma
- My model fits the data from the last election
    - But what if the relationship between different IVs (education, income, age…) and turnout changes?
- **Misspecified**: your model does a poor job of predicting who will vote!

Gallup's surprisingly simple likely voter model: 7 questions

1. Thought given to election (quite a lot, some)
2. Know where people in neighborhood go to vote (yes)
3. Voted in election precinct before (yes)
4. How often vote (always, nearly always)
5. Plan to vote in 2012 election (yes)
6. Likelihood of voting on a 10-point scale (7-10)
7. Voted in last presidential election (yes)

Score 6 or 7: classified as a "likely voter"

See here for more info

Problem #4: "shy" Trump voters

One hypothesis: some people prefer not to declare support for Trump

- Remember "testable implications?"
- If the hypothesis above is correct, what would we expect to see in the data?
- Kennedy et al. (2018) point to 2 testable implications:
  - Differences in Trump support by mode
    - Telephone interview: ↑ pressure to self-censor
    - (How do we test this? An experiment!)
  - Trump should outperform polls more than other Republicans
    - Other Republicans: ↓ pressure to self-censor
- Findings: very little evidence of the "shy" Trump voter

Jennings, Will, and Christopher Wlezien. 2018. "Election Polling Errors Across Time and Space." Nature Human Behaviour 2 (4): 276–83. https://doi.org/http://dx.doi.org.proxy3.library.mcgill.ca/10.1038/s41562-018-0315-6.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, et al. 2018. "An Evaluation of the 2016 Election Polls in the United States." Public Opinion Quarterly 82 (1): 1–33. https://doi.org/10.1093/poq/nfx047.

Prosser, Christopher, and Jonathan Mellon. 2018. "The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses." Government and Opposition 53 (4): 757–90. https://doi.org/10.1017/gov.2018.7.

Schuldt, Jonathon P., Sara H. Konrath, and Norbert Schwarz. 2011. "'Global Warming' or 'Climate Change?': Whether the Planet Is Warming Depends on Question Wording." Public Opinion Quarterly 75 (1): 115–24. https://doi.org/10.1093/poq/nfq073.