

Comparing Data Quality and Sources of Error in Probability-based Online Panels and Online Opt-in Samples

Andrew Mercer

Principal Methodologist

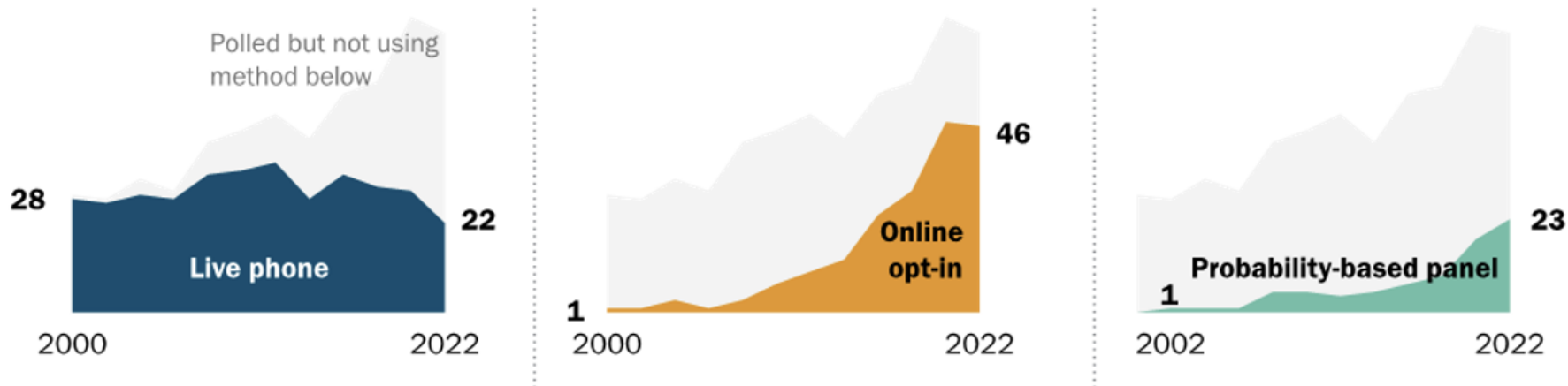
SOME BACKGROUND

Two broad types of online survey panels

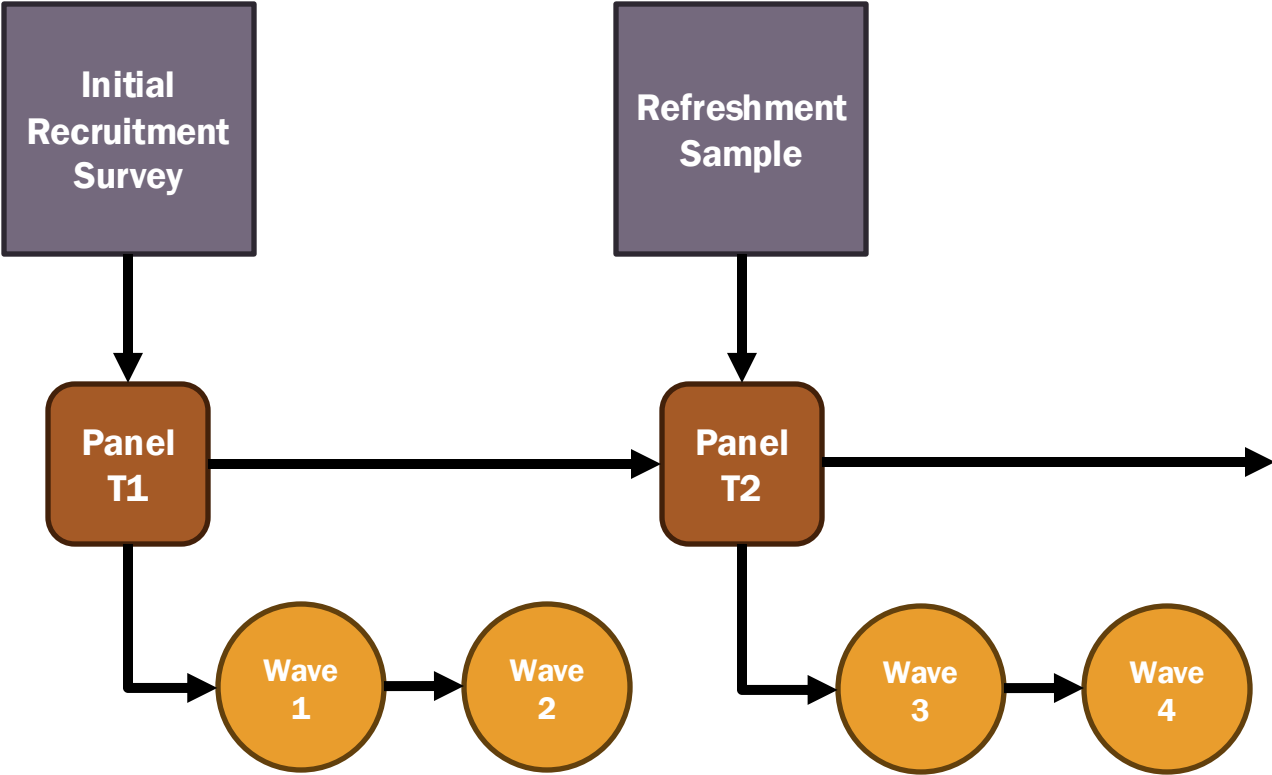
	Probability-based	Opt-in
Sources	Random sampling from U.S.P.S. residential address list	Convenience (e.g., social media ads, mobile apps, customer loyalty lists)
Examples	American Trends Panel, KnowledgePanel, AmeriSpeak, SSRS Opinion Panel	Cint, Qualtrics, Protégé, Prolific
Cost	Relatively high	Relatively low

Online opt-in now most common method for election polls

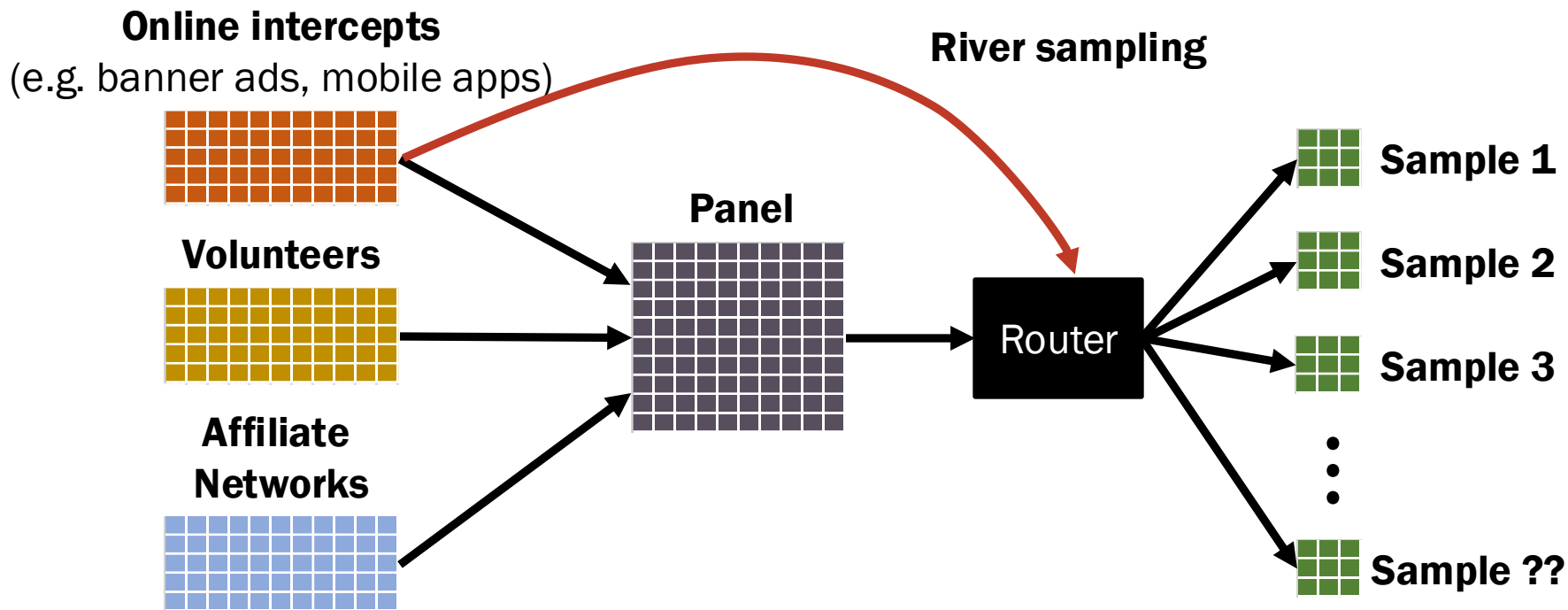
Number of U.S. pollsters using method alone or in combination with other methods



Probability-based Panels

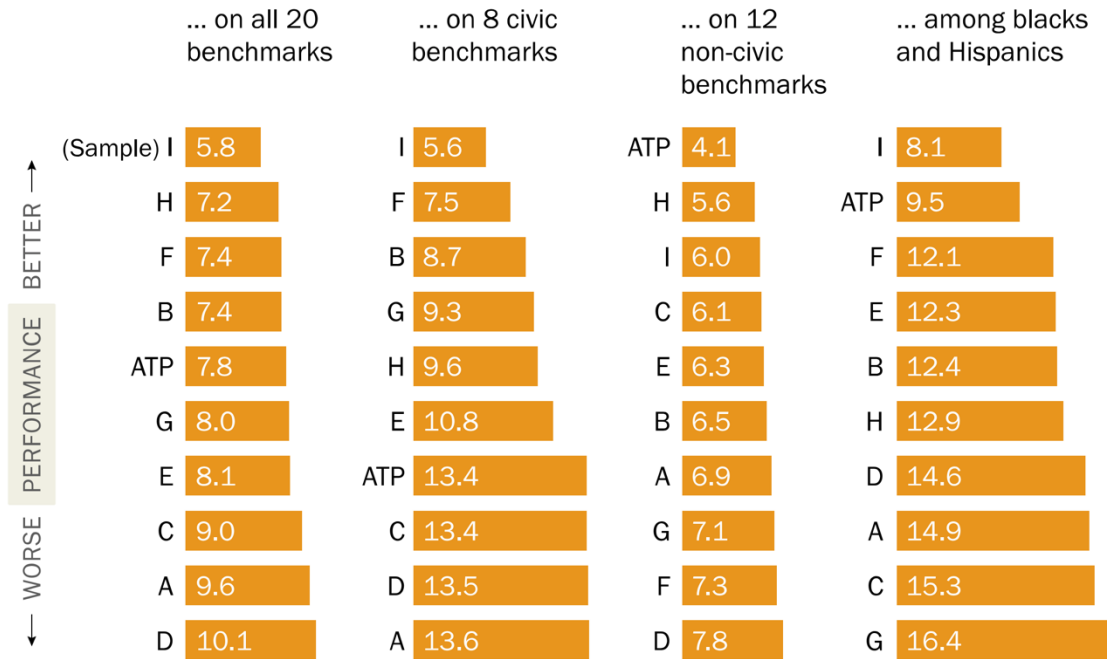


Online opt-in sampling



2016 Pew Research Benchmarking Study

Average absolute error...



Since then, we have seen...

- Many changes to the American Trends Panel's methodology.
- Growth in the number of commercially available probability-based panels.
- Changes in online opt-in methods, technology and business models.
- Large polling failures in the 2016 and 2020 presidential elections.

THE PRESENT STUDY

A new study for a new era

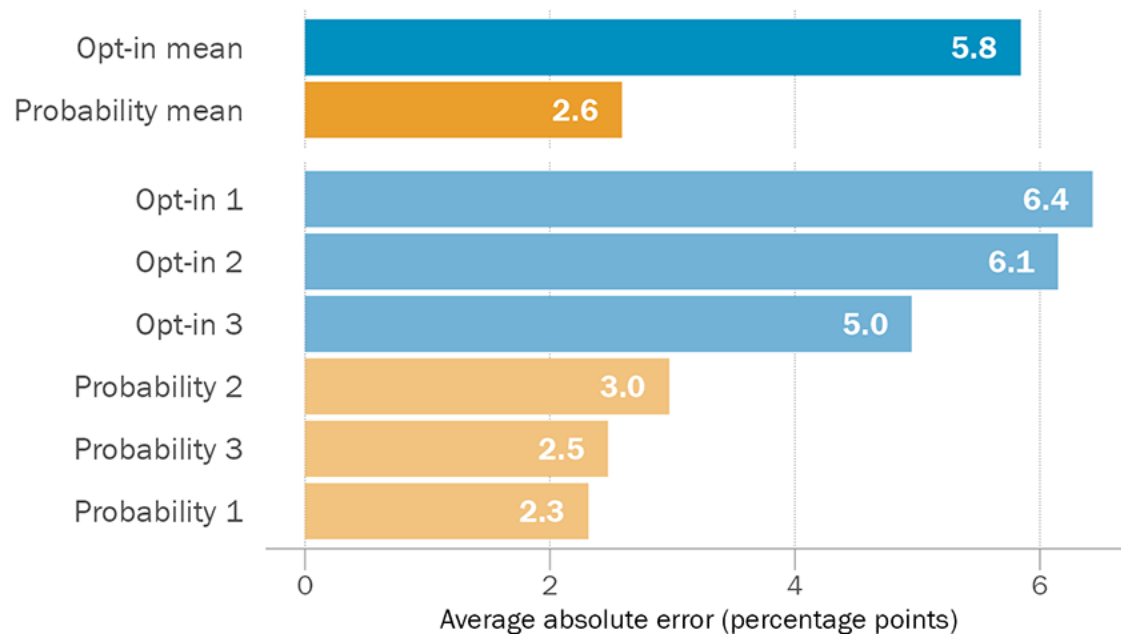
- Six online samples, all from different sample providers
- 15-minute questionnaire
- 28 variables with high-quality federal benchmarks
- Standard ATP weighting scheme applied to all samples

Sample	Field dates	n
Probability 1	June 14-28, 2021	5,027
Probability 2	June 14-27, 2021	5,147
Probability 3	June 29-July 21, 2021	4,965
Opt-in 1	June 15-25, 2021	4,912
Opt-in 2	June 11-27, 2021	4,931
Opt-in 3	June 11-26, 2021	4,955

WHAT DID WE FIND?

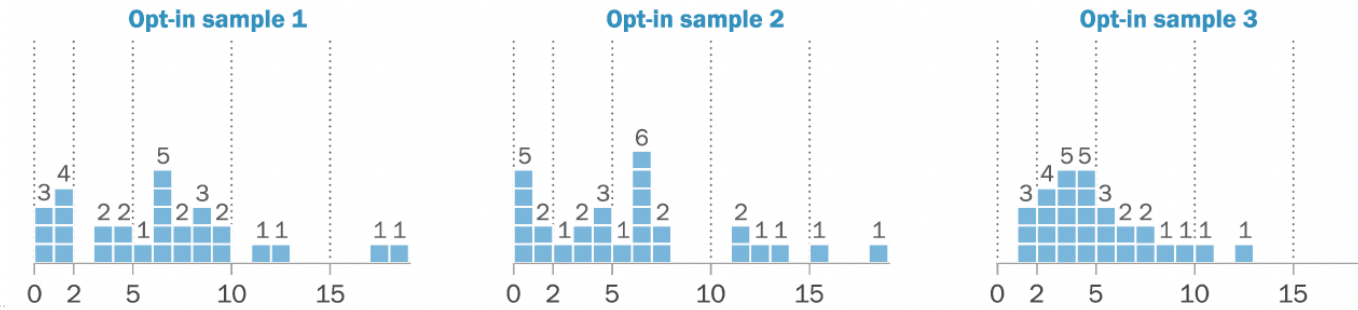
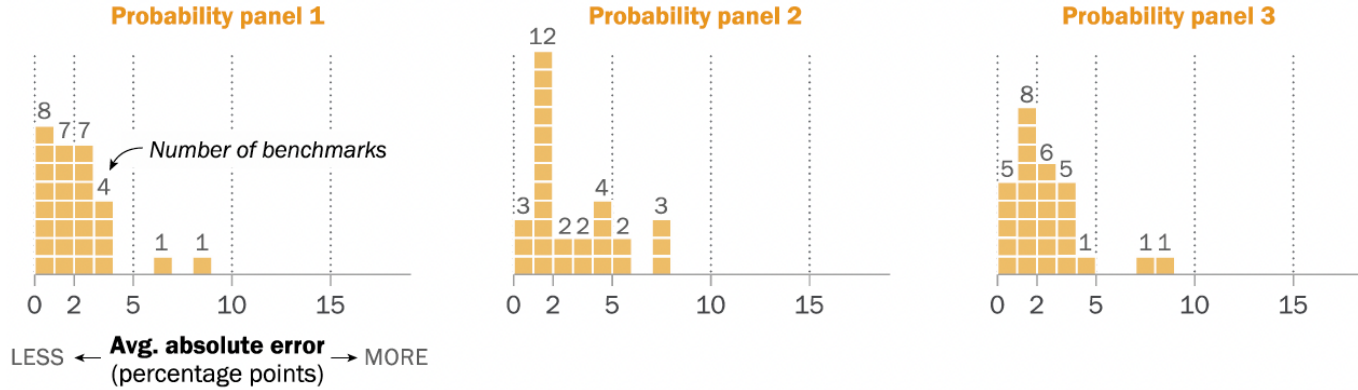
Average error on online opt-in samples was twice that of probability-based panels

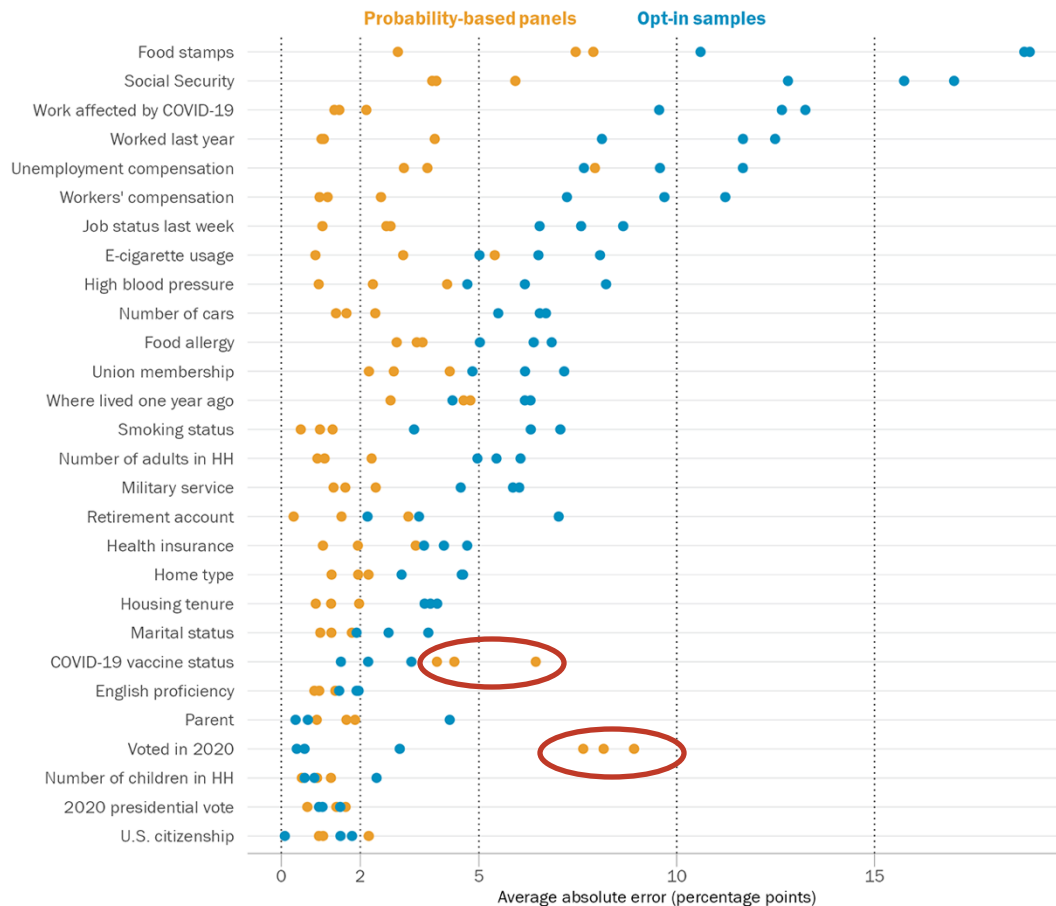
Average absolute error on 28 benchmark variables for estimates among U.S. adults



Opt-in samples had many more benchmarks with large errors than probability-based panels

Average absolute error for estimates among U.S. adults on 28 benchmark variables. About half of the benchmarks had more than 5 points of error on the opt-in samples compared with only a handful on the probability-based panels.

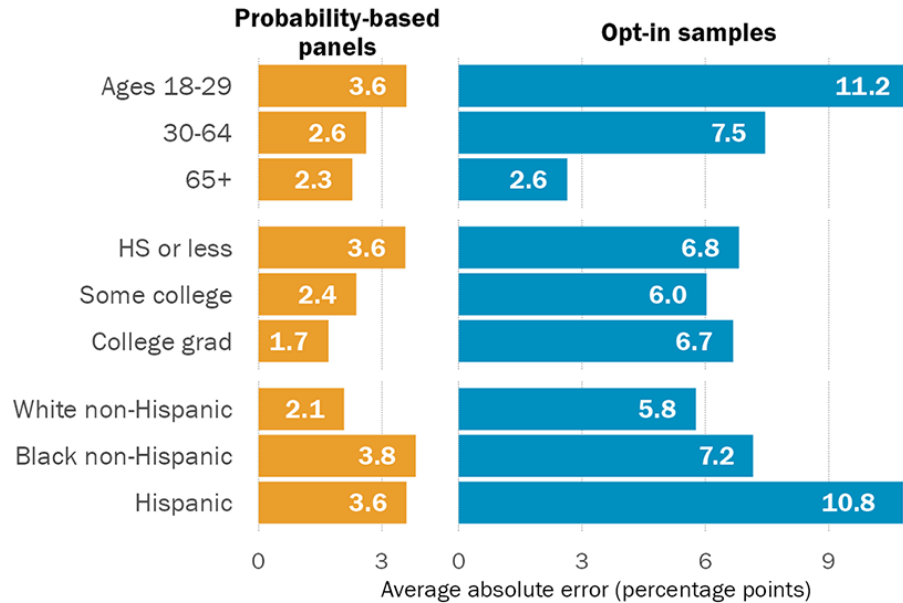




THE BIG STORY TURNED OUT TO BE SUBGROUPS

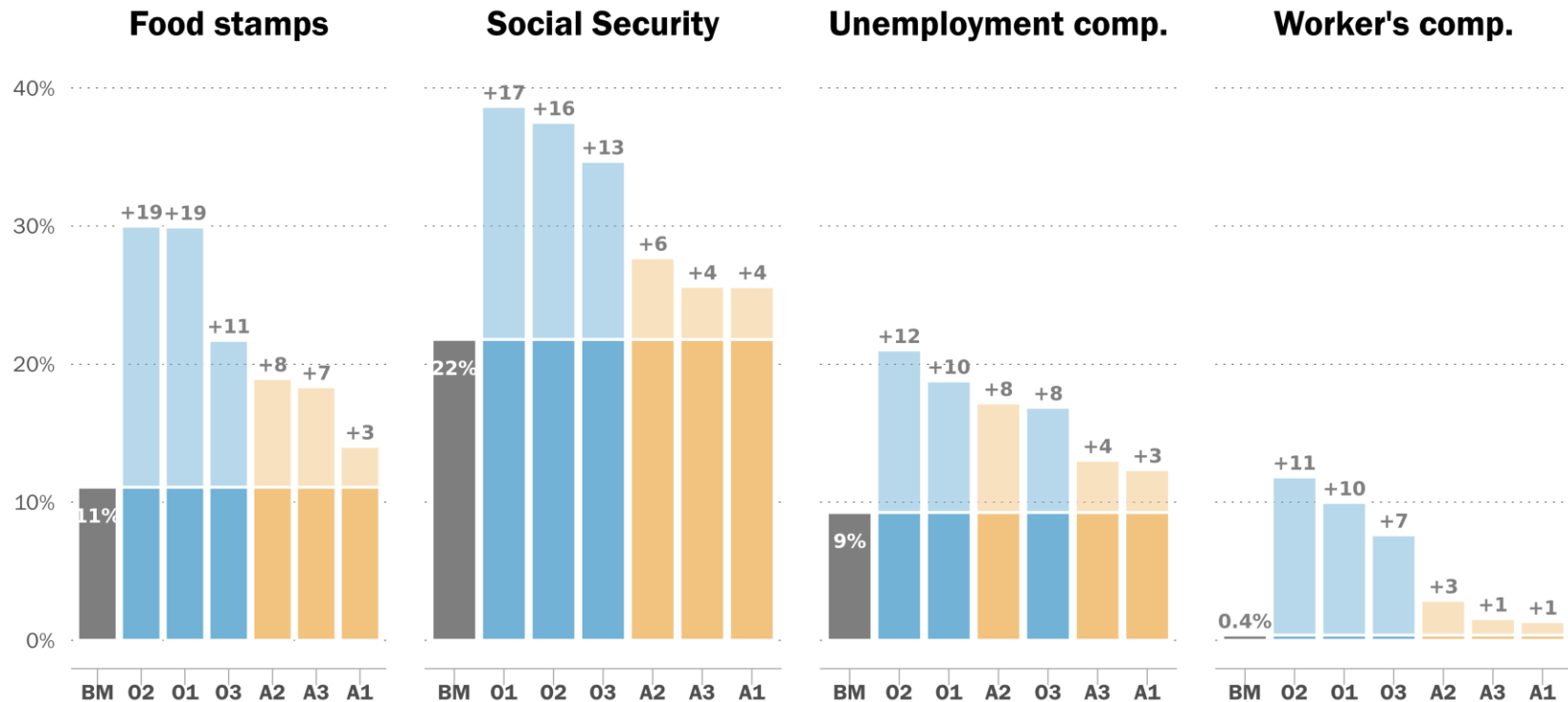
Large errors for 18- to 29-year-olds and Hispanic adults on opt-in samples

Average absolute error on 25 benchmark variables for estimates among demographic subgroups



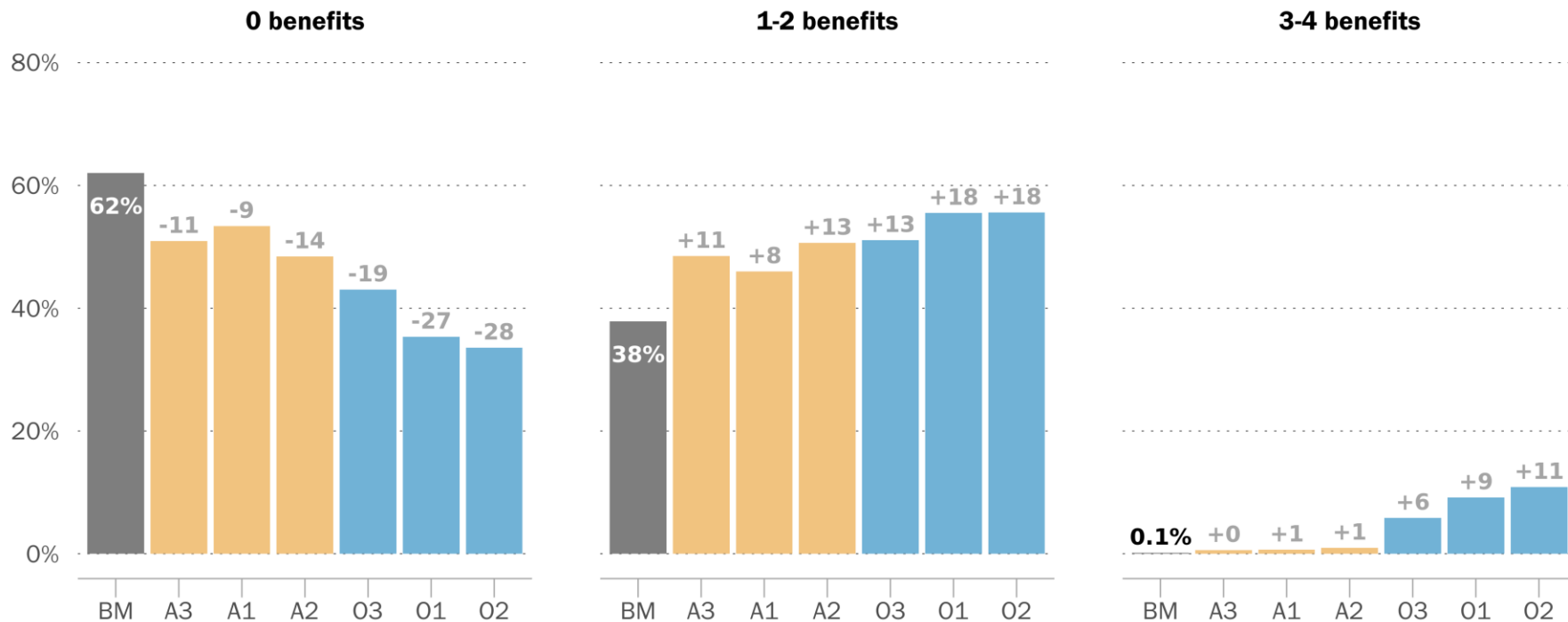
Receipt of government benefits

% of U.S. adults who received ___ in 2020



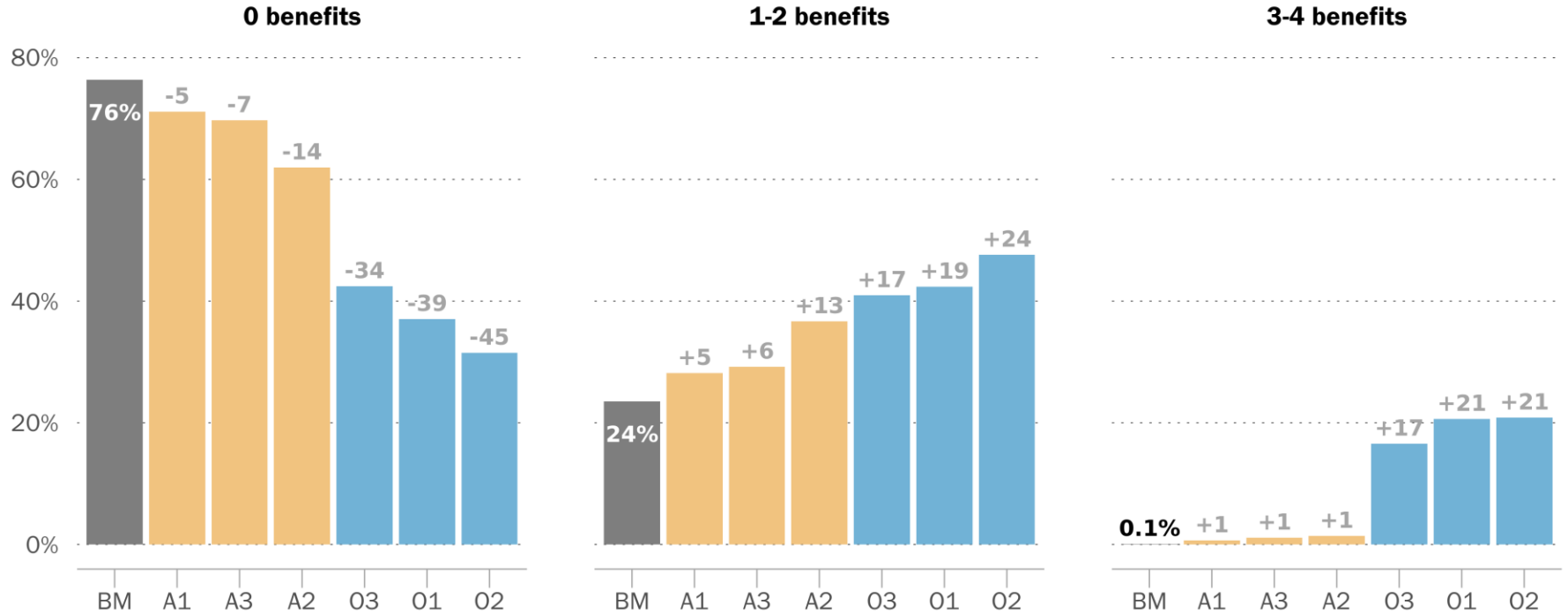
Opt-in overstated the share receiving 3-4 benefits

% of all U.S. adults who received ___ in 2020



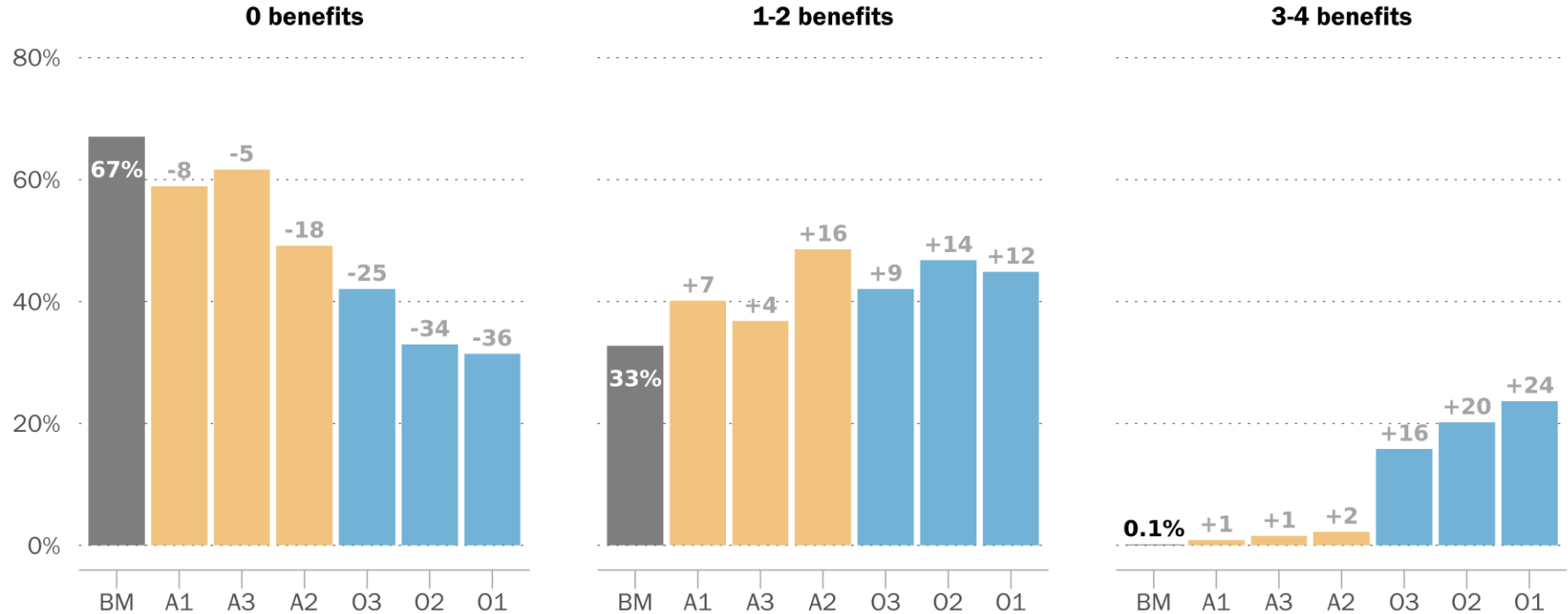
Number of benefits received among ages 18-29

% of adults ages 18-29 who received ___ in 2020



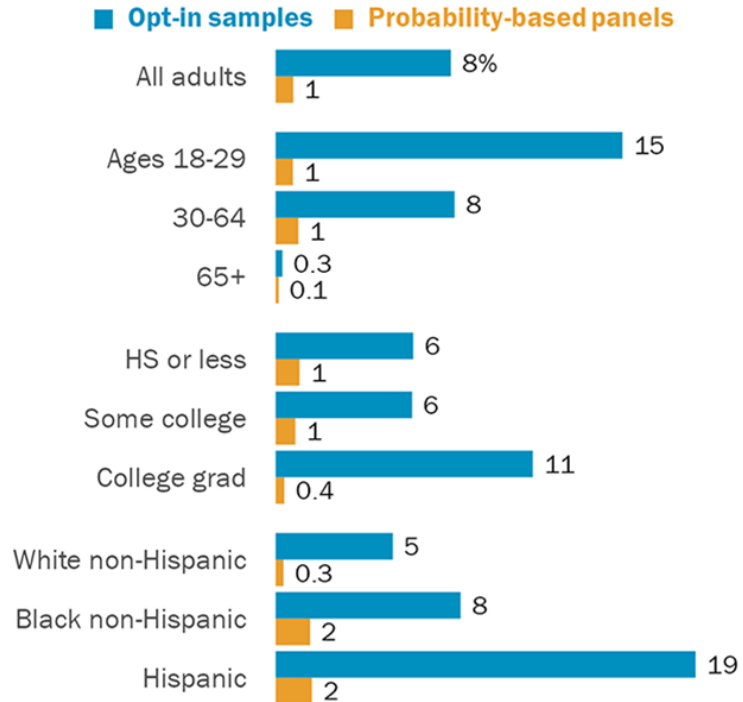
Number of benefits received among Hispanics adults

% of Hispanic adults who received ___ in 2020



Some opt-in respondents tend to answer 'Yes' regardless of the question

Average % of U.S. adults who answered "Yes" to at least 10 of 16 Yes/No questions



The problem of “bogus respondents” in opt-in samples

- Respondents who make little to no effort to answer survey questions truthfully.
- Disproportionately concentrated in estimates for key demographic subgroups.
- Results in a pattern of error that is dramatically larger within those groups.
- This is measurement error that’s not easily fixed through weighting.
- Does not manifest in probability-based samples.

We did a follow-up study in Feb. 2022

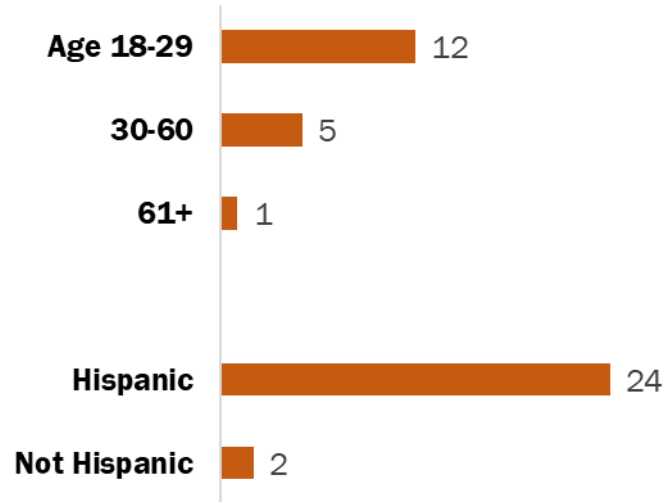
Are you licensed to operate a class SSGN submarine?

- Yes
- No



Same pattern of results

% of respondents saying they are licensed to operate a class SSGN sub





Andrew Mercer

@awmercercer.bsky.social

Just to be clear, we don't think that people answering this way are actually under 30 or Hispanic.

It's that people giving inaccurate data on questions about submarines are also giving inaccurate data about their demographics, falsely claiming to belong to these groups.

IS THIS ALL JUST ACADEMIC?

United States | Young opinions

One in five young Americans thinks the Holocaust is a myth

Our new poll makes alarming reading

Dec 7th 2023 | WASHINGTON, DC

This newer brand of antisemitism has led a generation of young people to dispute the Holocaust. One in five American adults under 30 say they agree the Holocaust was a myth, a new poll conducted by The Hill and NewsNation shows. Many believe

THE HILL

News | Policy | Business | Health | Opinion | Events | Jobs

1 in 5 young Americans believe Holocaust was a myth

BY NICK ROBERTSON - 12/08/23 2:15 PM

NEWSNATION

Poll: 20% of surveyed young Americans believe Holocaust may be myth

- New poll suggests some young people think Holocaust was a myth
- Many states don't require Holocaust education

Daily Mail .com

Home | Showbiz | Femal | Royals | Sports | Health | Wellness | Science | Politics | Money | U.K. | Breaking News | Australia | Video | University Guide | China | Debate | Meghan Markle | Prince Harry | King Char

ADVERTISEMENT



Electric Car

Browse the top senior searches.

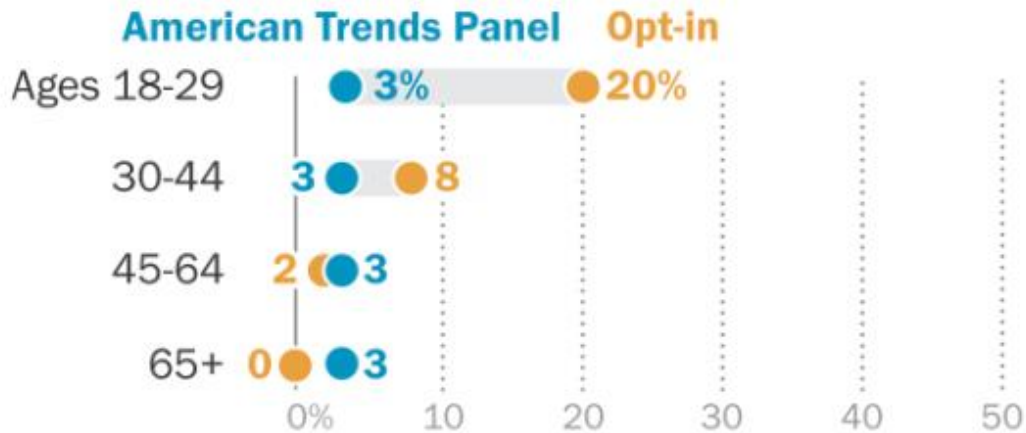
Senior Electric Car D

One in five young Americans believes the Holocaust is a myth - while another 30% say they are unsure if the genocide ever took place, poll finds



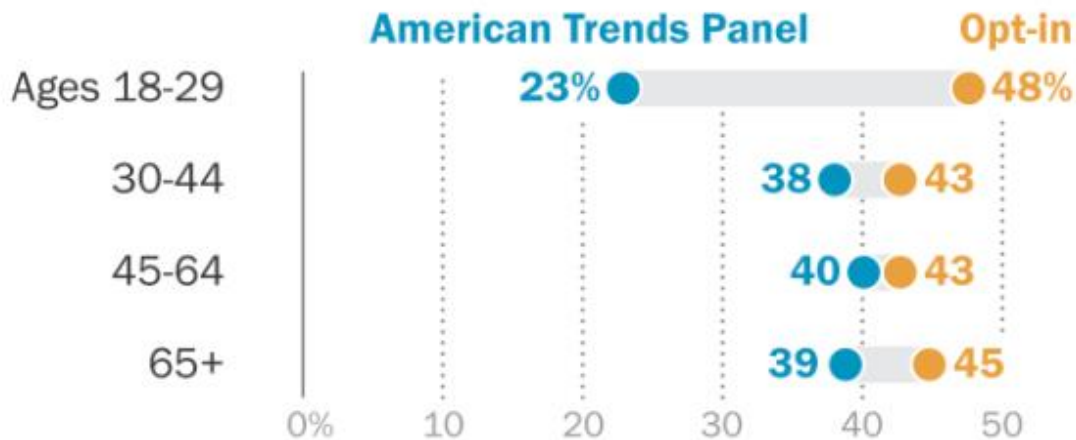
We tried to replicate this result on the ATP

% of U.S. adults who say they agree with the statement “The Holocaust is a myth”



Findings on abortion attitudes also failed to replicate

% of U.S. adults who say abortion should be illegal or only legal in special circumstances



What does probability vs. nonprobability mean today?

- Both have problems but they're different problems
- Representation vs. Measurement error
- Implications for blended samples
- Methods for detecting and dealing with bogus respondents
- All of our inferences are model-based now...
- ...but there are other benefits to probability-based methods

Thank you!

Andrew Mercer
Senior Research Methodologist

@awmercercer.bsky.social

