Improving Polling Accuracy:
GSG Review of 2020 Data & Plans for Future

GSG
GLOBAL STRATEGY GROUP

Scope of Analysis

- **Data collected by GSG**
- **10 states, 9 with Senate Races**
  - AK, AZ, CO, IA, IL, ME, MI, MT, NC, and WI (no Senate race)
- **39 surveys**
  - Conducted post-Labor Day
  - Mix of phone, text, and voter-file matched panel
- **27,800 total interviews**
- **Primary focus was Presidential and Senate Races**

Summary of Findings

- **Final polling had 2.6 points of pro-Biden bias with "base-weighting" approach alone**

  Base weighting – weighting to a projected electorate based on the voter file and modeling. GSG uses additional approaches to control partisanship, including weighting on recalled 2016 vote and weighting metrics like Party ID across multiple surveys ("composite weighting"), so our final materials showed less than 2.6 points of bias.

- **Turnout error accounted for ~35% of bias and 4% of error in presidential results**

  Turnout error is bias and error that results from projected electorates differing from the actual electorate.

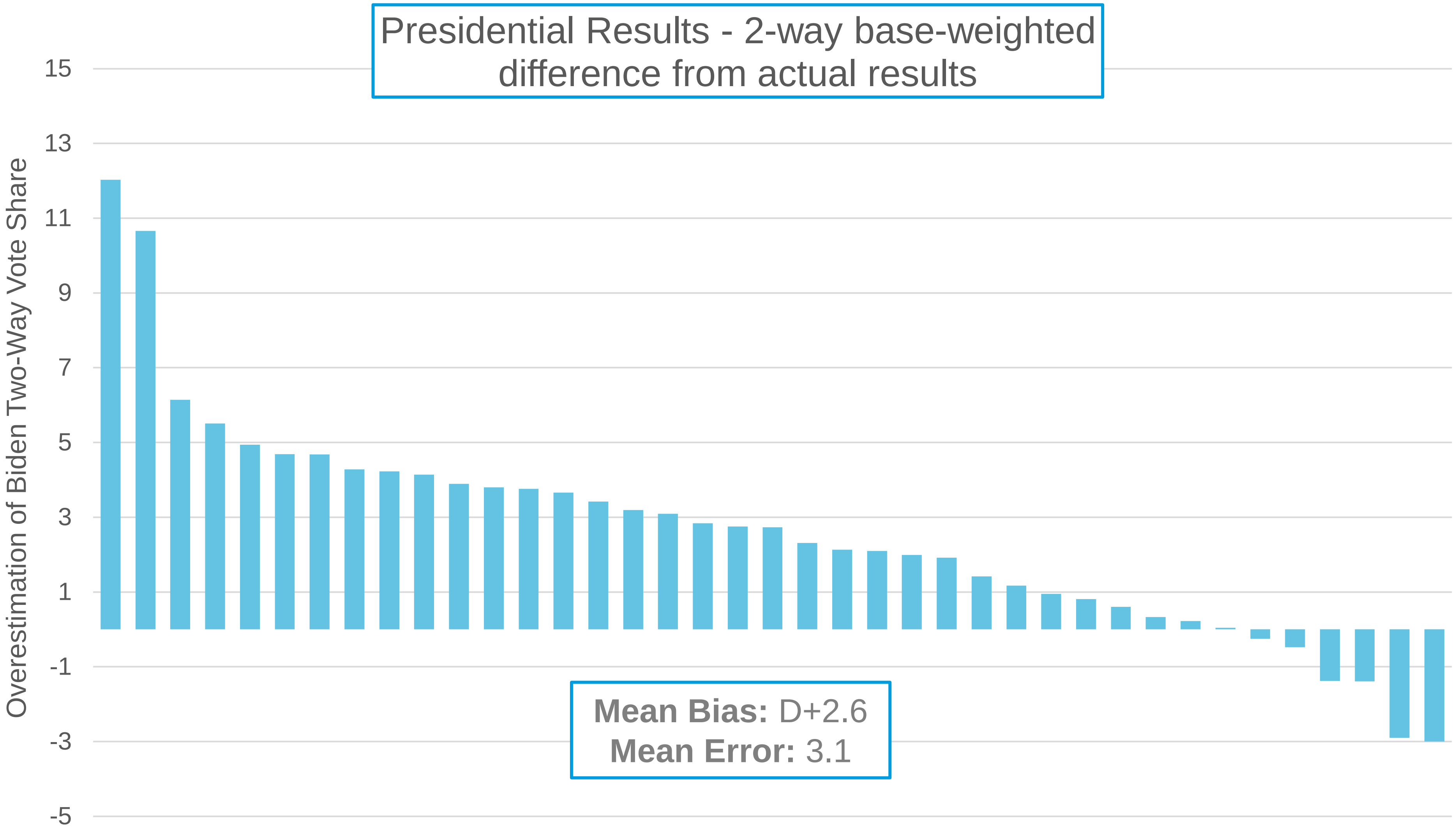
- **Partisan differential response bias left uncorrected by base weights accounted for at least 9% of bias and 38% of error**

  Partisan differential response happens when Republicans and Democrats respond to surveys at different rates. Partisan and other weights (registration, scores) correct for a large amount of this, but these are imperfect measures of partisanship.

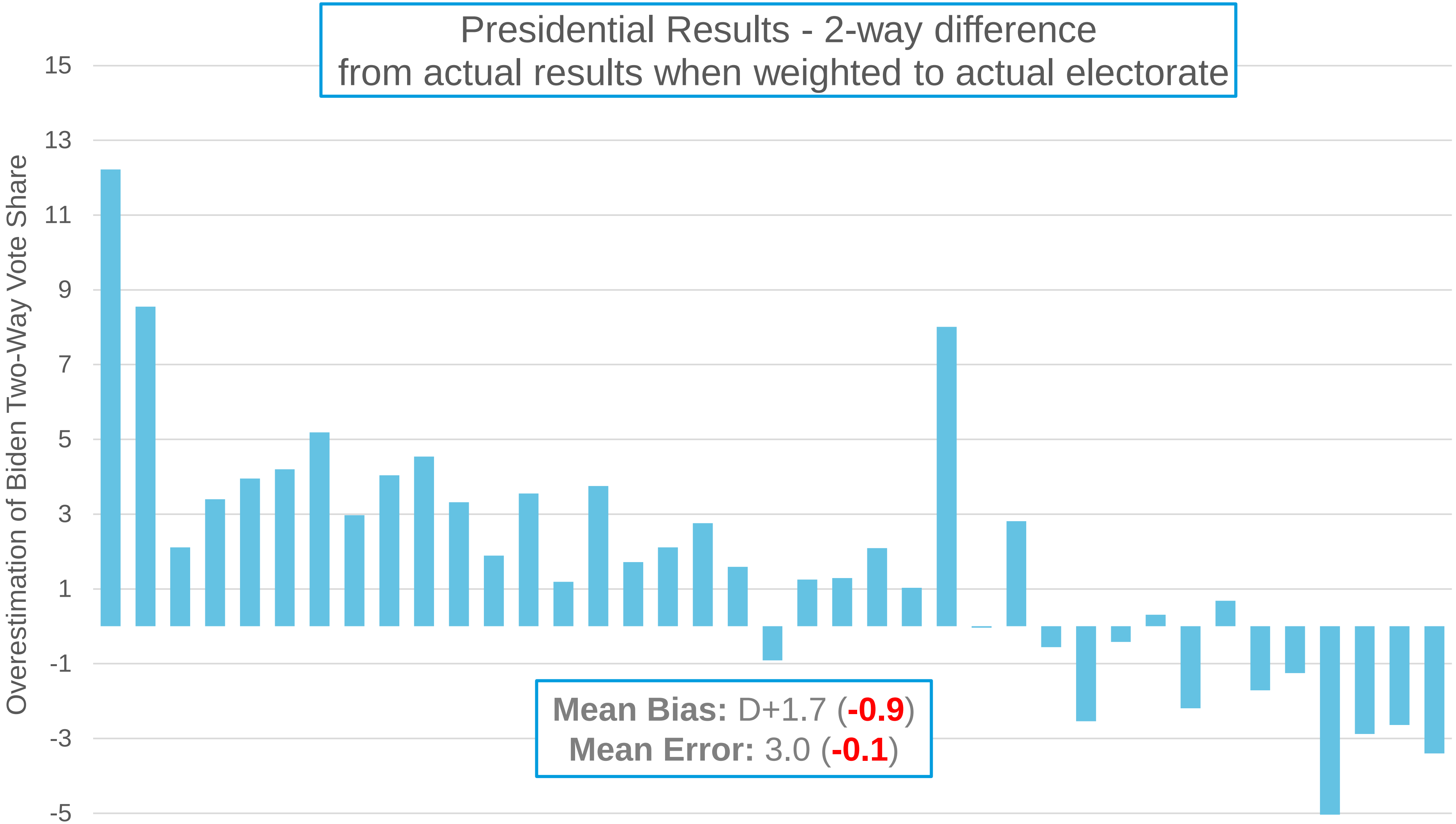
- **Non-partisan differential response accounts for some of the remaining bias and error**

  Non-partisan differential response happens when underlying traits or attitudes other than partisanship (for example: pro-Trump attitudes *among* Republicans) affect response rates to surveys and are not accounted for by existing weights.

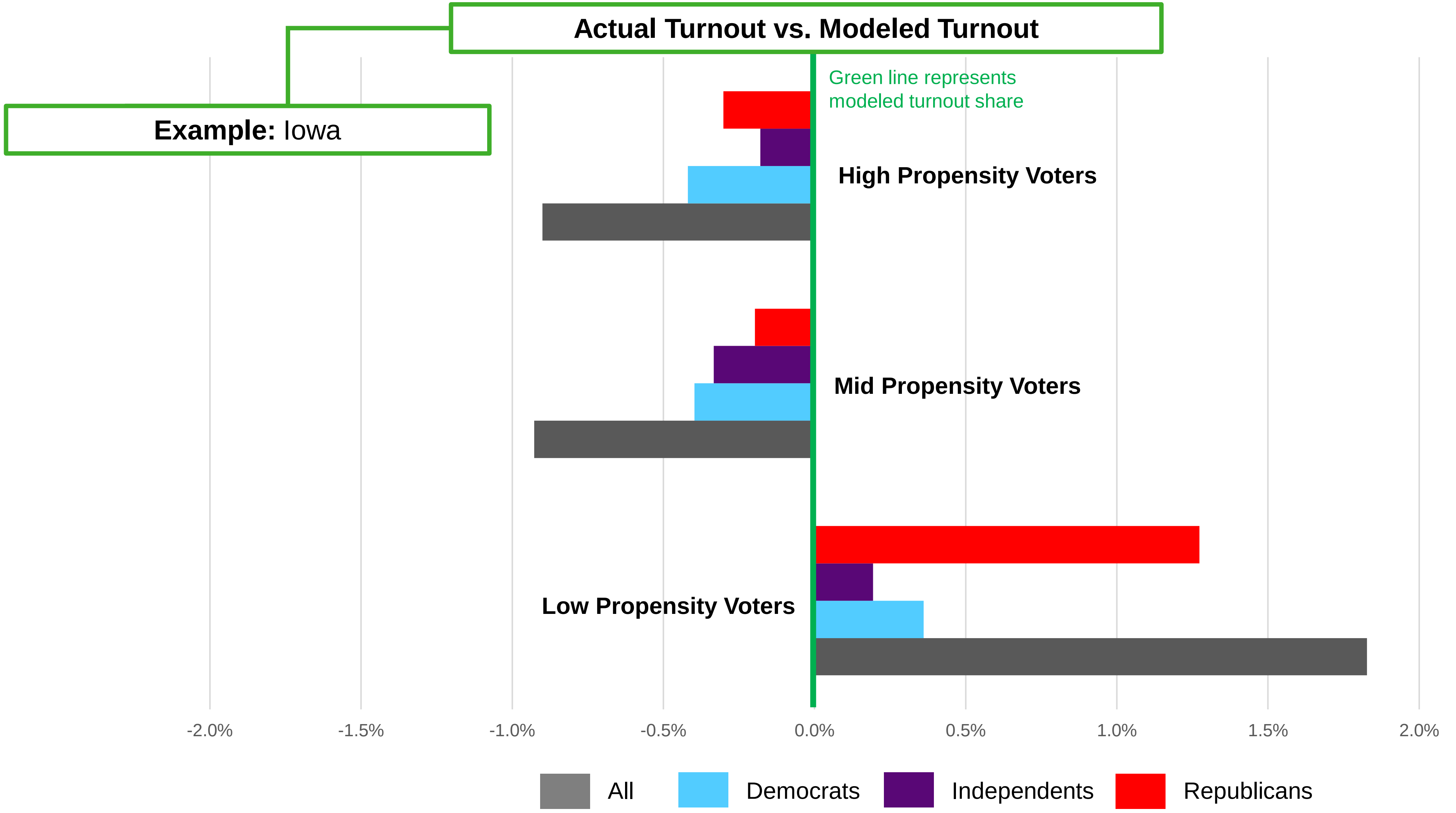
Visualizing Bias: On average, our polls across these 10 states had 2.6 points of pro-Biden bias using base weights



Turnout Error: When weighted to the actual electorate, the mean bias is reduced to 1.7 points on average

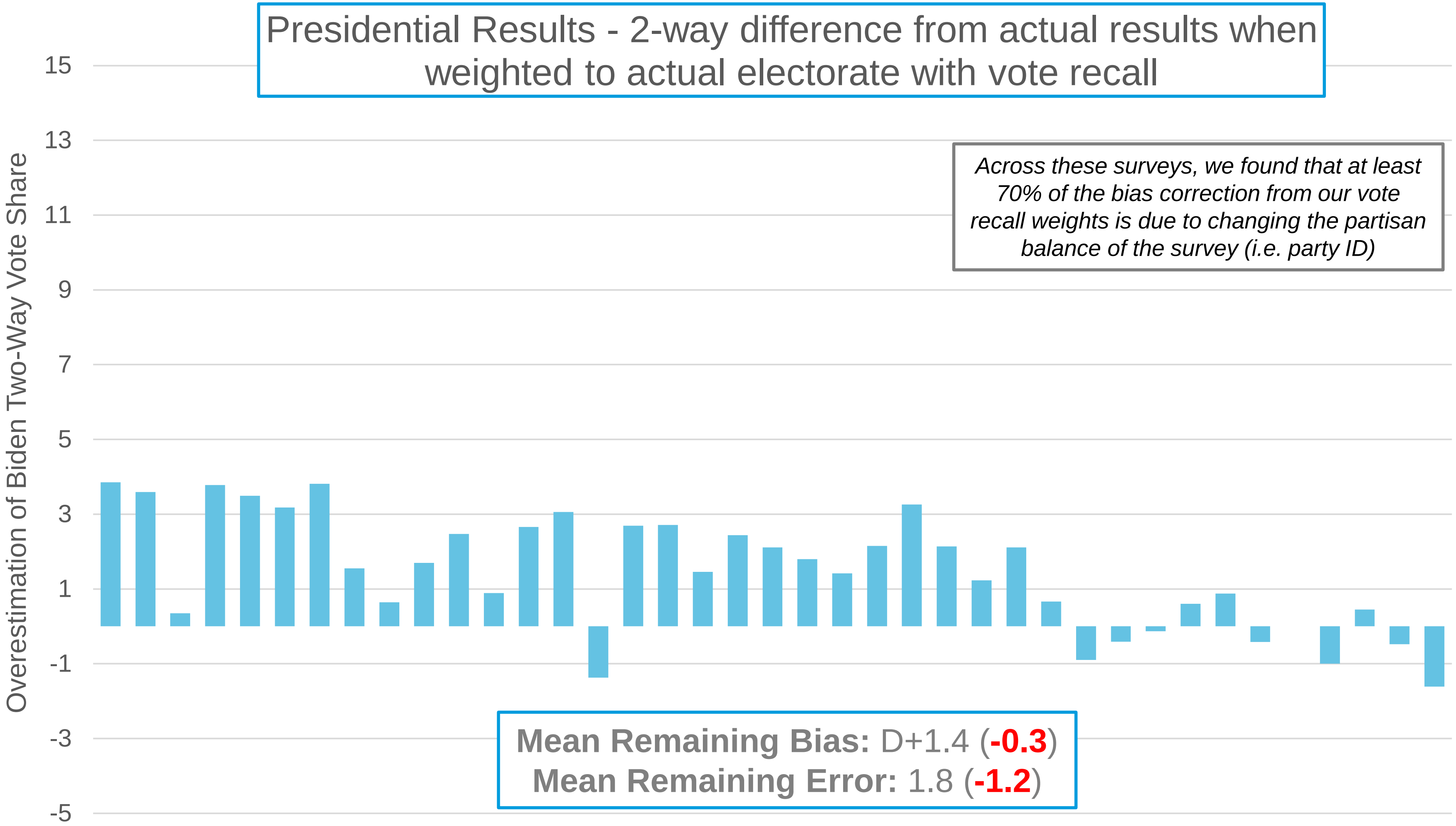


Real Life Example of Turnout Error: Low propensity voters exceeded their expected share; most of that gain came from Republicans

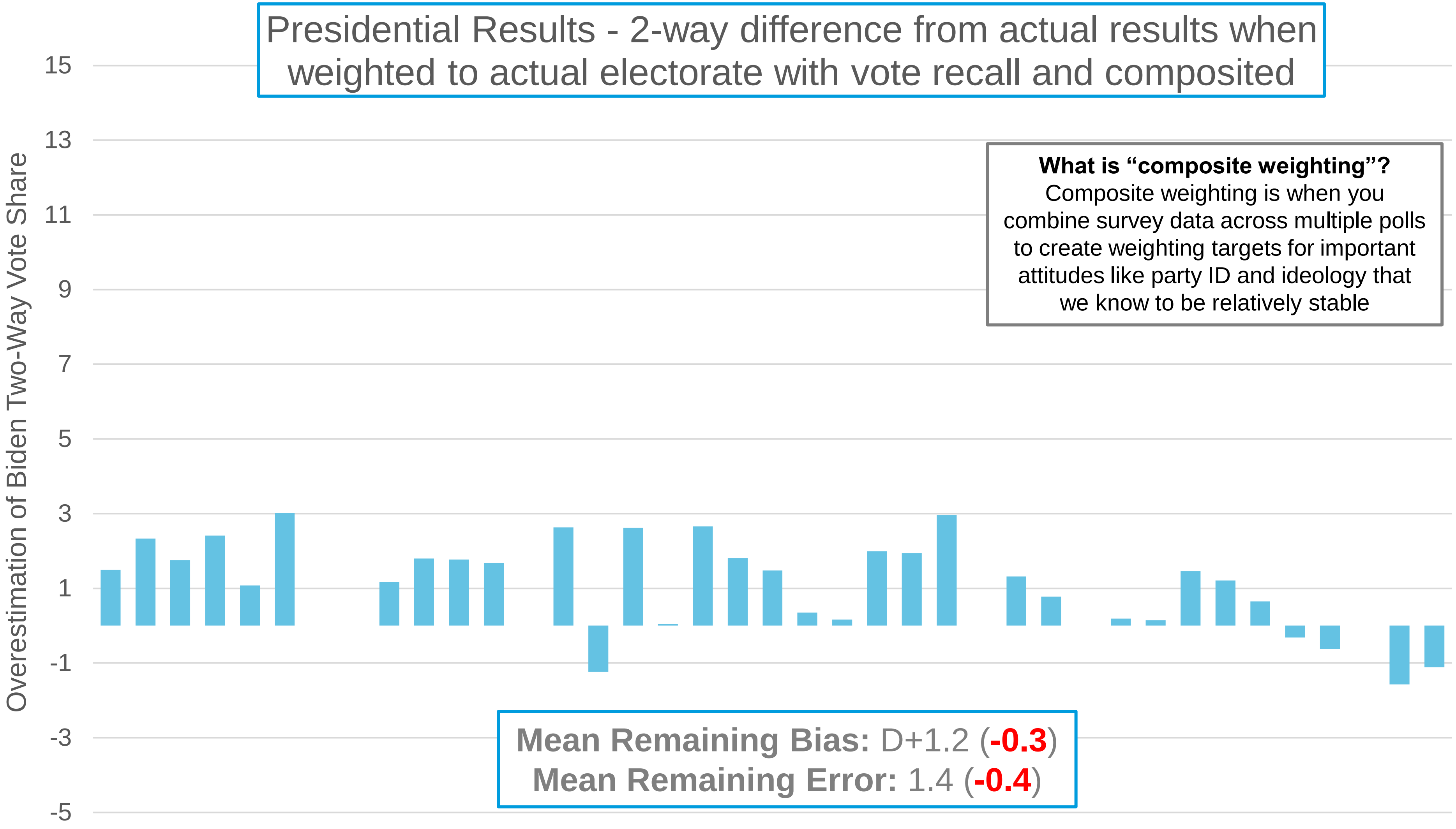


Similar propensity-based turnout error occurred in other states, including Arizona, Montana, and North Carolina

Further Reducing Error w/ Vote Recall: Weighting to 2016 Vote recall corrects some additional bias and reduces 1.2 additional points of error



Further Reducing Error w/ "Composite Weighting": Compositing surveys based on past survey data reduces both error and bias further*



**Surveys that were the first in the post-Labor Day period are omitted from this slide because they could not be composited. Just among these surveys, bias was D+1.5, and error was 1.8.*

Composite Weighting in Practice

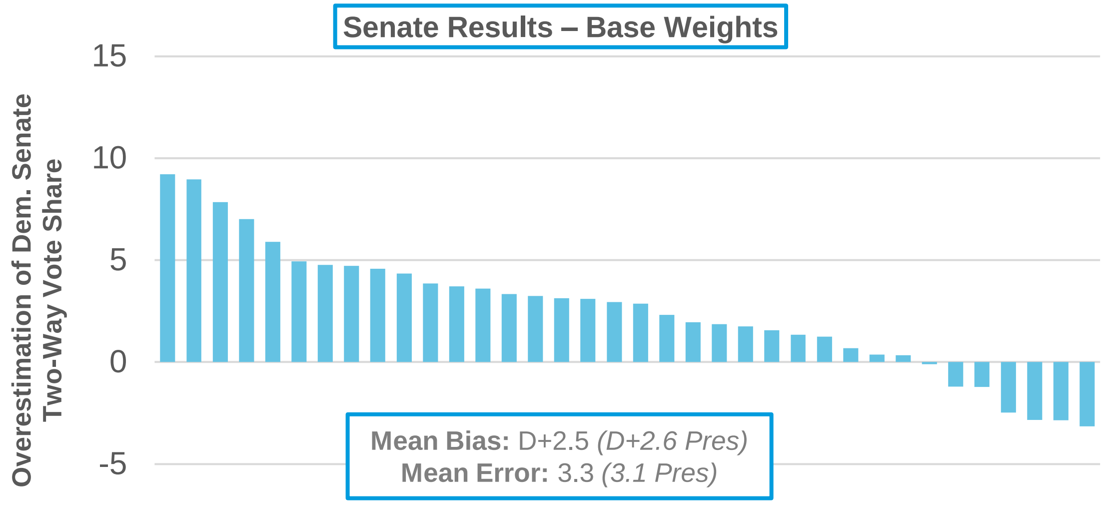
- **We can combine survey data across multiple polls to leverage information on <u>stable</u> attitudes (PID, ideology) and create weighting targets**

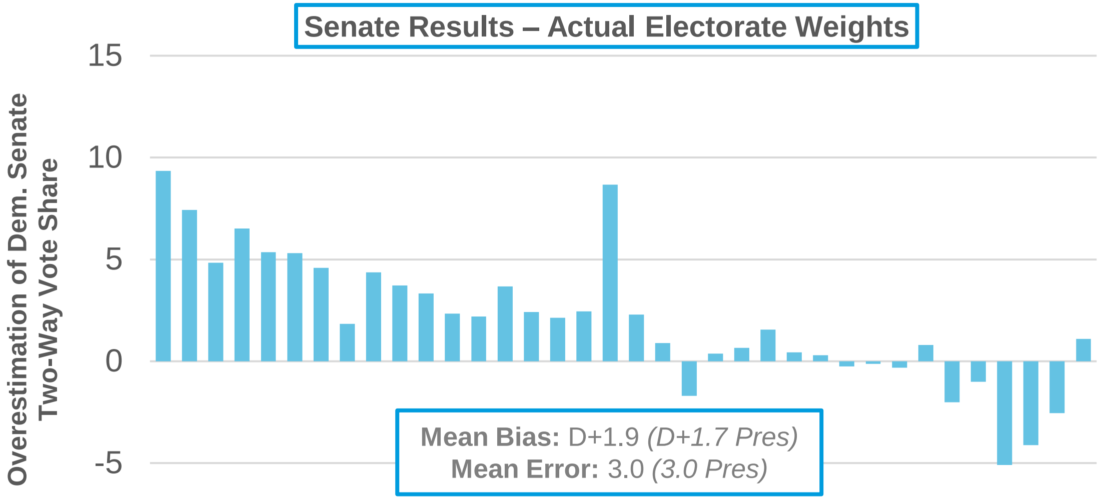
  - Avoids "wasted" data on key partisan metrics, creating a larger effective sample size

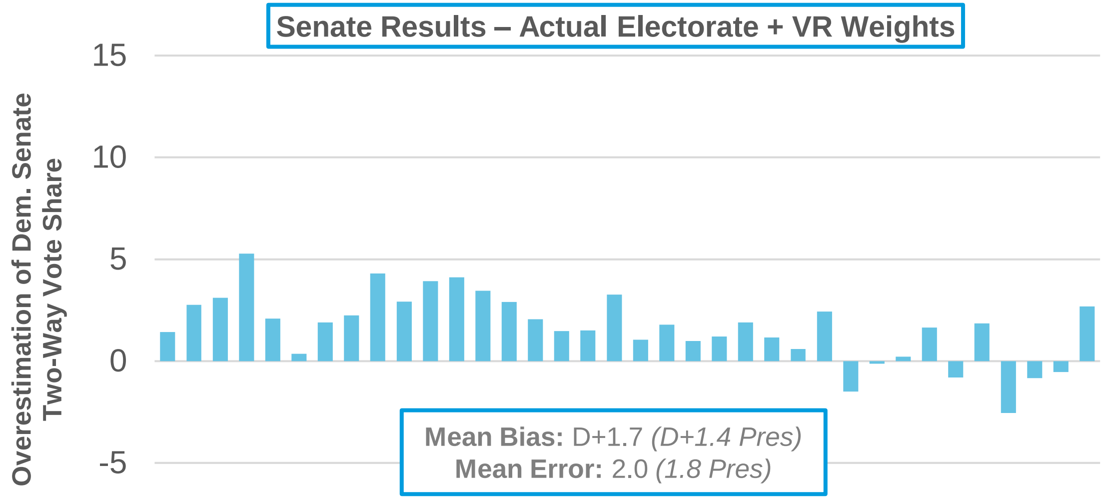
  - Reduces error – statistical noise – in assumptions about relationships *between* partisanship and important, but smaller subgroups (race x age x region, etc.)

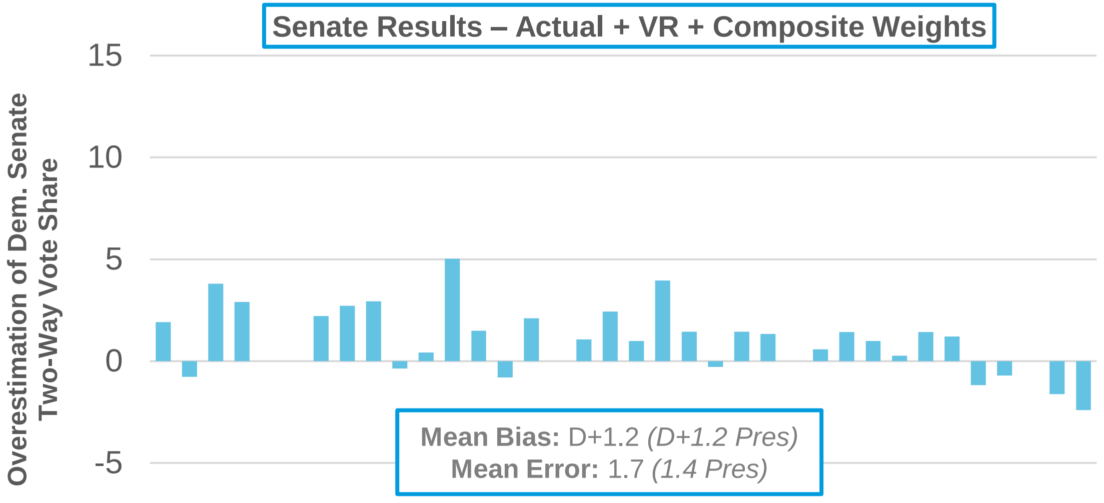
			Last 6 surveys - Base weighted						
		"Truth"	*Aug.*	*Sept.*	*Sept.*	*Oct.*	*Oct.*	*Nov.*	Target
All likely voters	PID	R+5	R+3	R+11	R+1	R+3	R+9	R+1	R+4.6
	Interviews	n/a	600	600	600	600	600	600	3600
Hispanic subgroup	PID	D+31	D+39	D+40	D+30	D+32	D+30	D+26	D+32.8
	Interviews	n/a	130	102	146	115	143	102	738

Senate Results: In general, the same trends exist for Senate Races

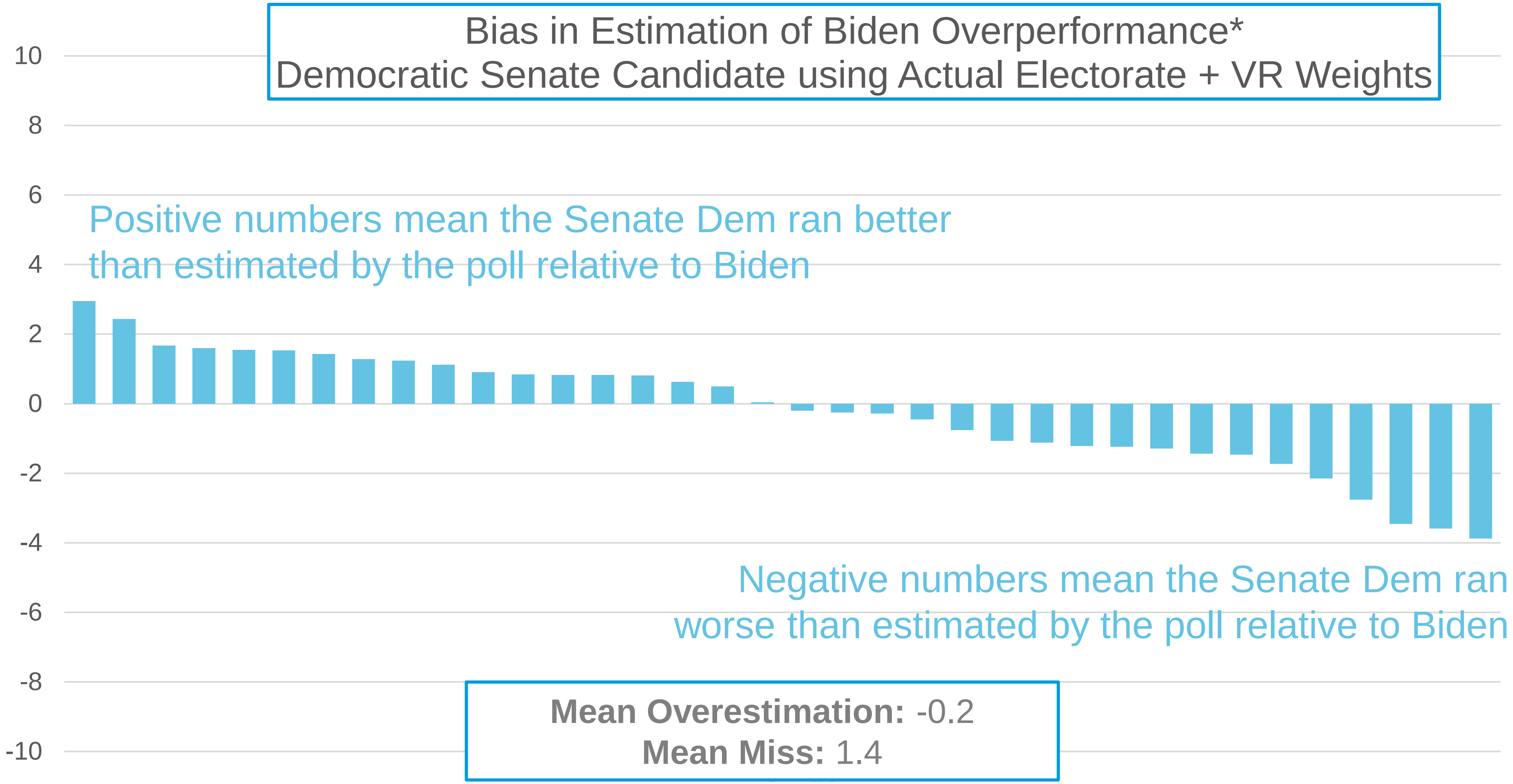








Margin Between Presidential and Senate: With our best weights applied, there is no consistent trend of over or under-estimating Senate candidates relative to Presidential



** (Estimated Presidential Result - Estimated Senate Result) - (Actual Presidential Result - Actual Senate Result)*

Why is partisan response bias under-corrected when we weight only using voter file party metrics?

- **Partisan voter file variables used by pollsters to weight (party registration, party scores, etc.) are not perfect measures of individual-level partisanship.**
  - For example, in Iowa, across 1,308 post-Labor Day interviews with registered Democrats, 76% identified as Democrats and 16% as Republicans. Among those who said they voted in 2016, only 74% were Clinton voters and 15% Trump voters.

- **The effectiveness of weights in adjusting for partisan non-response is diminished the more these partisan voter file variables deviate from underlying partisan preferences.**

- **We simulated additional error in our data to identify how well different types of weighting corrects error. In this analysis, we see partisan weights reduce far more error in states with party registration or strong partisan modeling than in states with weak modeling.**

	AVERAGE ERROR CORRECTION		
	Party Reg. States	Good PSCORE States	Bad PSCORE States
Level 1: Only demographics (age, race, etc.)	10%	13%	8%
Level 2: Demographics + Party on File	36%	35%	13%
Level 3: Both of the above + 2016 Vote Recall	70%	67%	66%

**Good PSCORE states record which party's primary voters have voted in, while bad PSCORE states only mark whether a voter voted in any primary; Averages across both metrics*

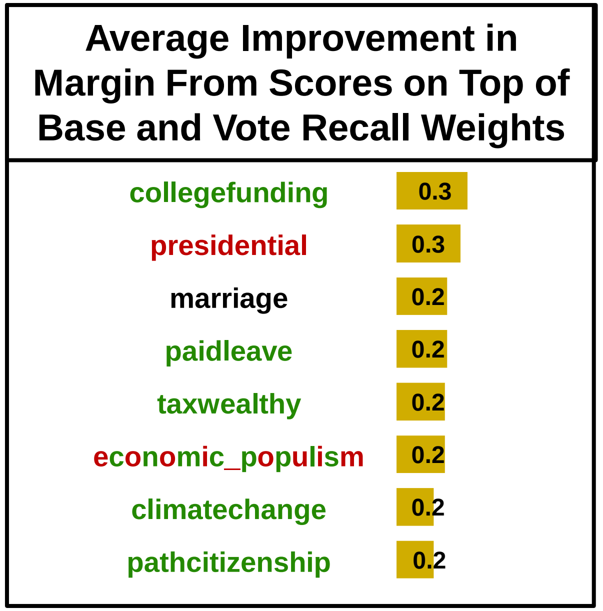
How do we deal with response bias beyond partisanship?

- **After weighting on 2016 vote recall, 1.4 points of bias remain. After composite weighting, 1.2 points of bias remain.**

- **Initial analysis suggests potential sources of nonpartisan response bias include:**
  - Pro- and anti-Trump attitudes not explicitly tied to partisanship
  - Pro-government spending attitudes

In order to investigate sources of bias making up what is left unaccounted for, we wrote an algorithm to re-weight our data using 53 different TargetSmart scores. We found that many scores effectively reduced error and bias, even on top of vote recall.

Many of the best performing scores were related to attitudes toward Trump or the role of government and spending in various aspects of life.



- **Not a magic bullet**
  - None improved data 100% of the time
  - Impact small given remaining error/bias
  - BUT: offer insight into the kinds of attitudes that we might be over- or under-representing in the electorate to inform future research and exploration

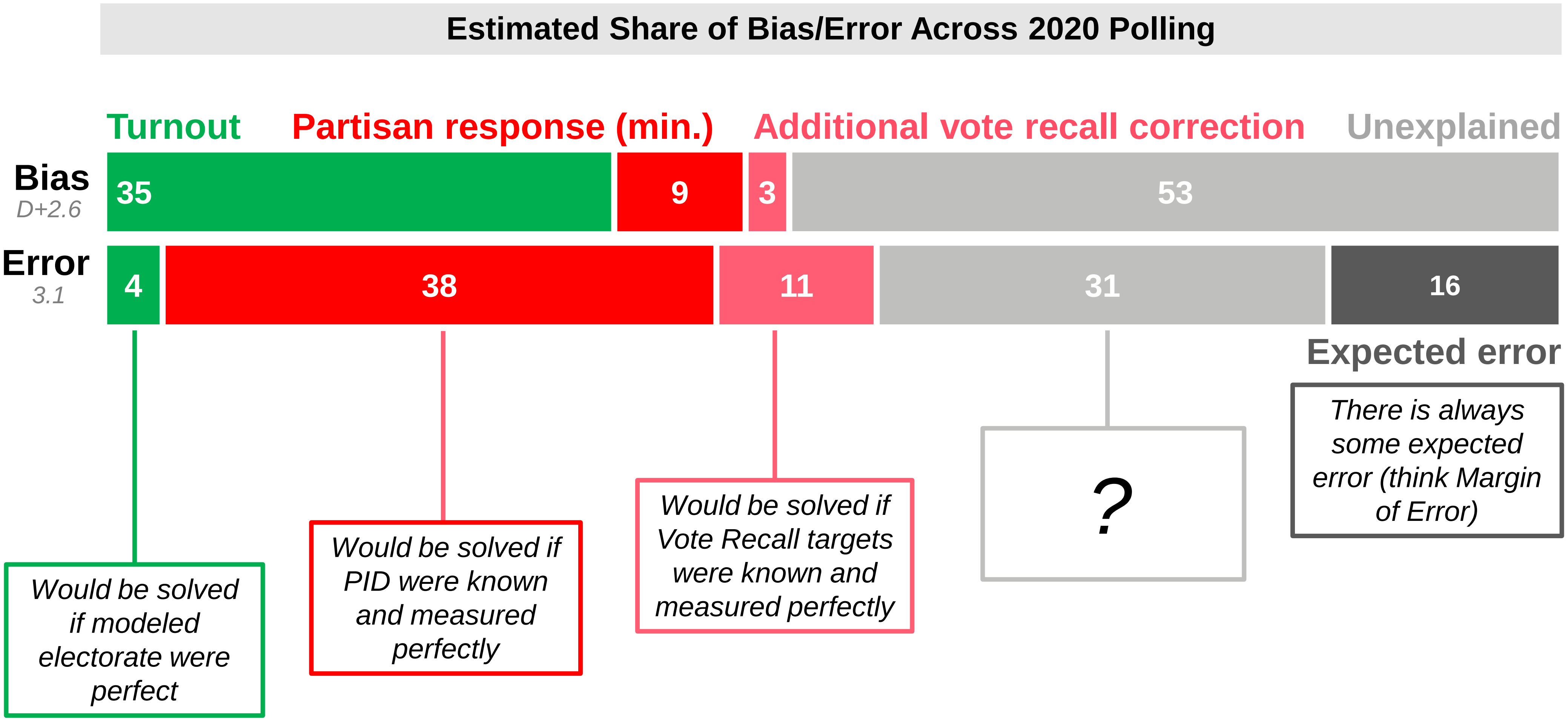
Real Life Example of Reducing Error

	Presidential Vote			Senate Vote		
	Biden	**Trump**	*Margin*	**Greenfield**	**Ernst**	*Margin*
GSG Polling	48.5	51.5	***-3.0***	48.7	51.3	***-2.6***
Weighted to actual turnout	47.8	52.2	***-4.4***	48.0	52.0	***-4.0***
Weighted to actual turnout + 2016 vote recall	47.5	52.5	***-5.0***	47.7	52.3	***-4.6***
Weighted to actual turnout + 2016 VR + presidential*	47.3	52.7	***-5.4***	47.5	52.5	***-5.0***
Actual Election Day Result	45.8	54.2	***-8.4***	46.6	53.4	***-6.8***

Presidential score measures the likelihood that an individual views Trump as presidential using questions about his appropriateness, morality, and work ethic

Source: GSG polling on behalf of Theresa Greenfield; TargetSmart; Iowa Secretary of State

Breaking down error and bias



What’s next

- **GSG has already implemented findings from this accuracy study into its day-to-day workflow. However, while we believe better partisan weighting and modeling will go along way, it may not solve the problem on its own.**

- **We need to identify <u>new</u> attitudinal measures and benchmarks to weight on in order to correct further, and understand what is driving non-response.**

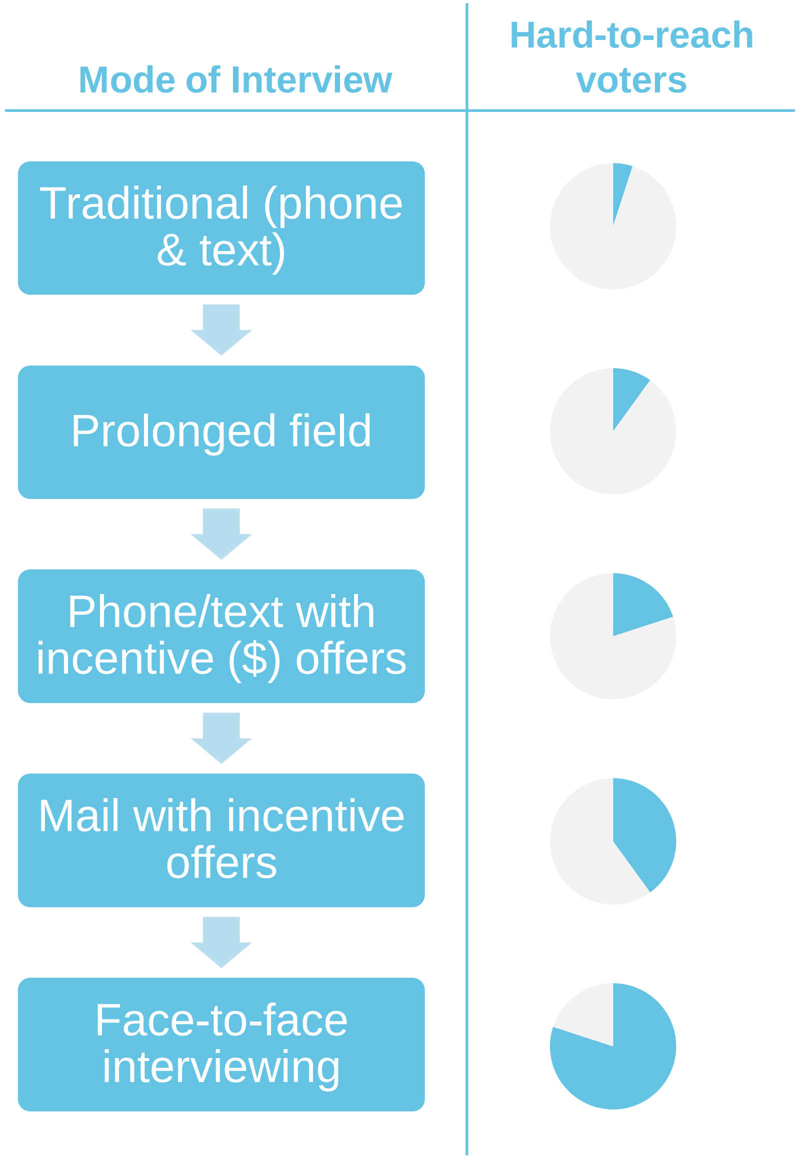
- **This is a key goal of the Wisconsin Project, a collaborative effort with five Democratic Polling firms (GSG, GBAO, ALG Research, Garin Hart Yang, Normington Petts) and Open Labs.**

- **GSG plans to continually revise its best practices over the next several months as new findings emerge from this work.**

The Wisconsin Polling Project: Approach

Use (costly, unconventional) methods to dramatically increase response rates and ideally identify those "missed" in 2020 polls:

- **First, conduct a survey of registered voters in Wisconsin using traditional methods**
- **Then, conduct a parallel survey using increasingly aggressive sampling procedures to improve response rates**
- **Finally, compare results from higher response rate (and higher cost) methods to traditional, lower response rate methods**



The Wisconsin Polling Project: Goals

- **Create representative baseline for population attitudes and characteristics that may be used to weight future surveys and/or give us an understanding of the types of people that are "missed." Survey would include questions that touch on various theories for differential response rates, including:**
  - Social trust
  - Racial resentment
  - Views on covid precautions
  - Views on corporal punishment
  - Views on the role of government

- **Learn from experience and potentially replicate some version of this effort in additional states/locales heading into 2022.**

Final Thoughts: Our Approach to Survey Weighting

- **What we are doing right now with our standard “base weighting” approach:**
  - Continue to weight on typical demographic and partisan information – weighting data in the standard way significantly reduces error from unweighted samples.
  - For now, weight on 2020 vote recall but BE CAREFUL:
    - Attitudes may shift over time – we will be monitoring this
    - Midterm & special electorates may have different partisan compositions – so don’t just weight to 2020 result.
    - This is not an end point or magic bullet.

- **What we are examining for the future with “base weighting”**
  - Weighting on engagement variables, as other studies have shown this to be a key factor
  - Weighting on new attitudinal benchmarks identified in upcoming research

- **Continuing to lean on large sample sizes with “composite weighting”**
  - Larger data sets can continue to provide targets for important attitudinal variables like party identification and ideology to reduce the variance inherent in small sample sizes.
  - If base weighting is not fixed, this alone will not solve the problem, but as base weights get better, this tactic can reduce error in smaller sample single polls.

Final Thoughts: Other Considerations Beyond Weighting

- **Differential partisan turnout is often a major problem and source of uncertainty in lower turnout midterm and special elections.**
  - Running turnout scenarios may help better express this uncertainty.
  - Having a dialogue about partisan assumptions with campaign team is important so everyone understands the context behind the numbers they are seeing.

- **We've been talking mostly about weighting. Sampling, research design, and mode are all important considerations as well that are also a focus of attention.**
  - There was no "magic mode" in 2020 – all modes, including multi-modal polling, experienced similar issues. However, there is more to learn and explore.
  - Changing how we sample on front end may also be an important piece of the solution, though we first need to better understand non-response bias.
  - Thinking about who is calling and who is said to be sponsoring polls may also be an important piece of the solution given the relationship between responsiveness and trust.

Thank You