Customizing Weighting Schemes for Online Opt-In Samples

AAPOR May 2023

The Harris Poll

Striving for Representativeness Beyond Demographics

Problem

- Differences between general population and online survey panels
  - Demographic
  - *Behavioral and Attitudinal*
- Minimize additional weighting while maximizing representativeness

Background

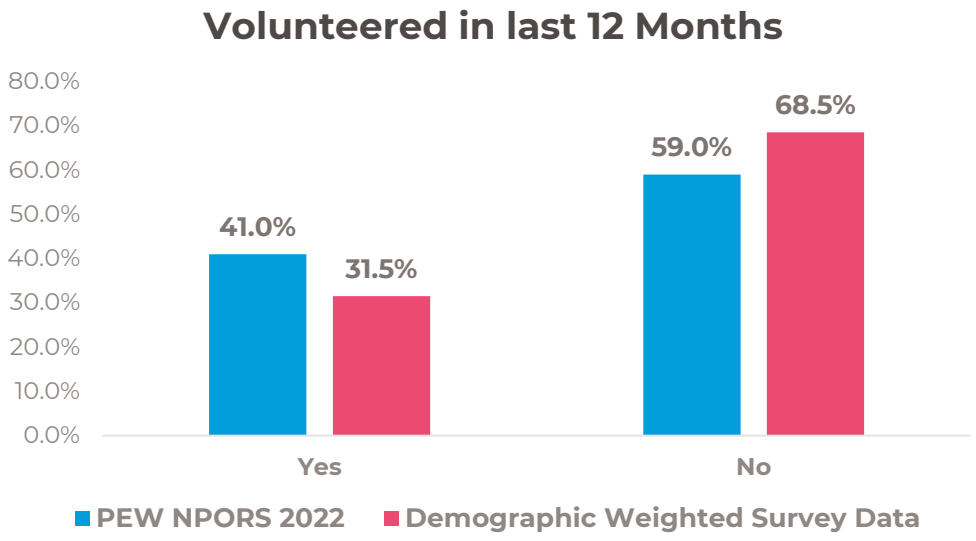
- Attitudinal and behavioral parameters in weighting schemes since the beginning of online survey research (Terhanian et al., 2000).
- Using behavioral and attitudinal measures to compensate for under-coverage and non-response (Fahimi, et al., 2015)
- Addition of politically focused variables reduced bias (Mercer et al., 2018)

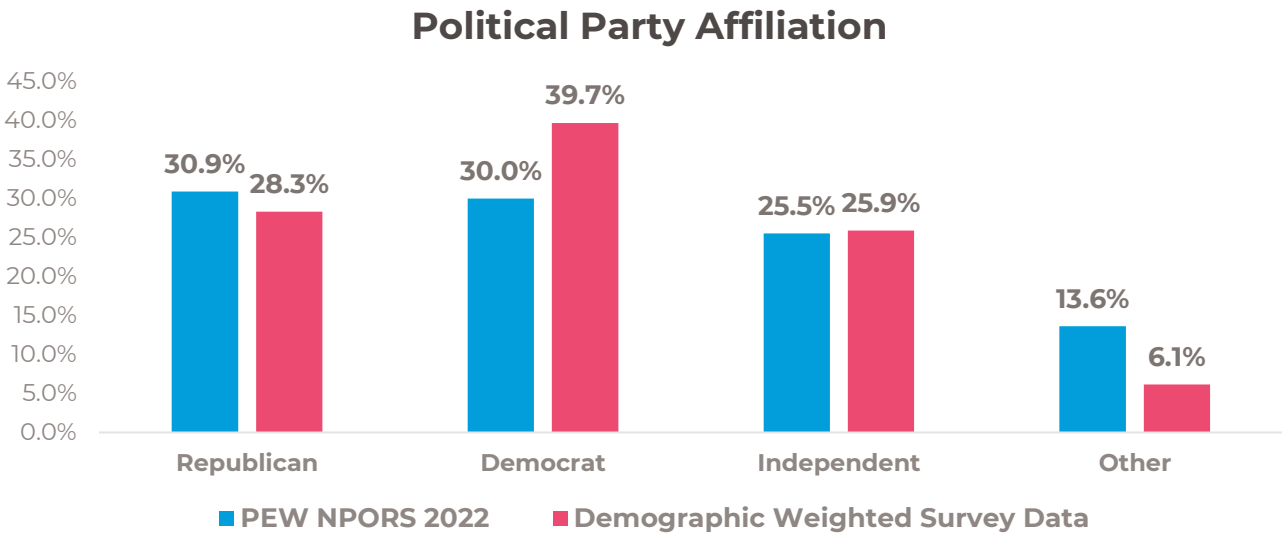
Let's check it out...

Will customizing weighting schemes based on survey topic help maximize bias reduction?

Identifying Attitudinal and Behavioral Benchmarks for Consideration

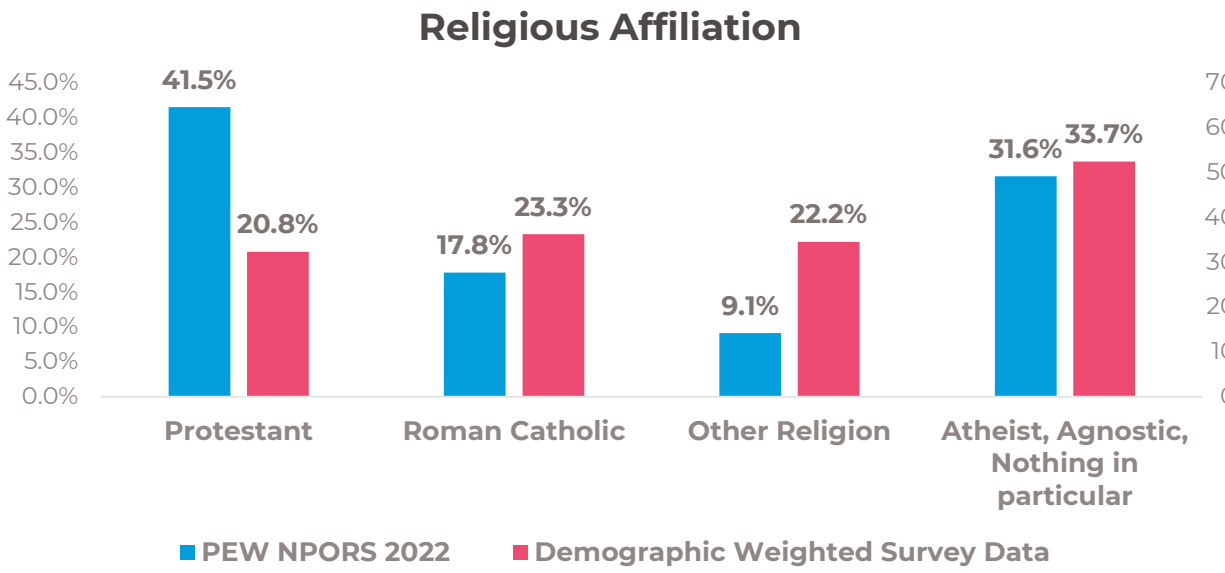
After demographic weighting, we still see a difference in representativeness of **Volunteerism** and **Political Party Affiliation** among the online opt-in panel sample compared to data from PEW NPORS 2022.

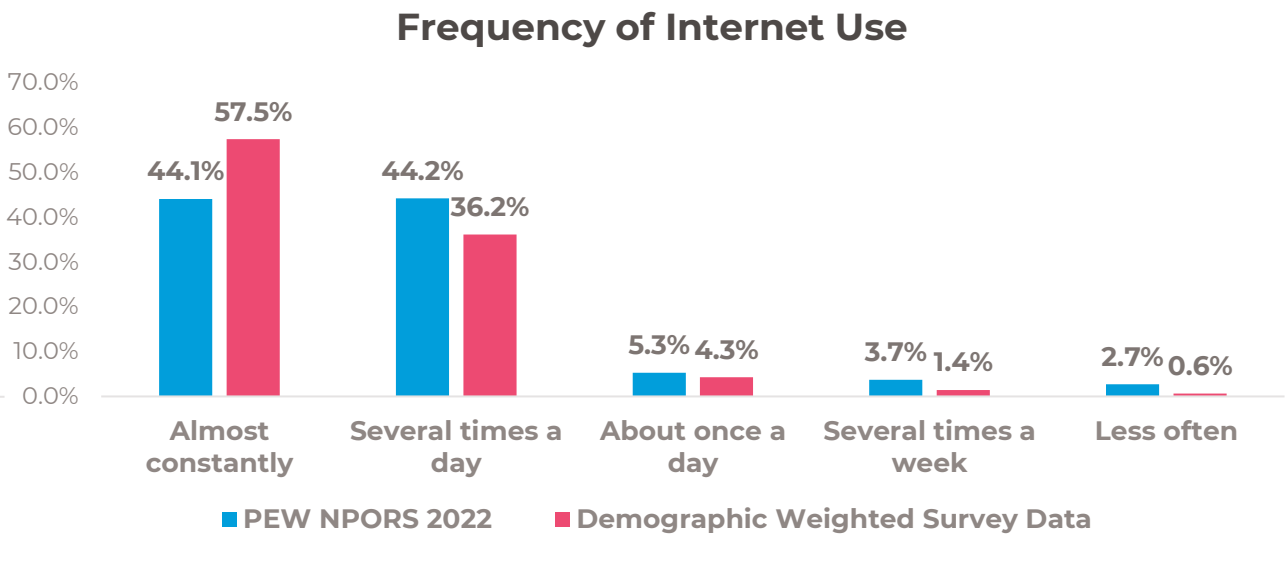




Identifying Attitudinal and Behavioral Benchmarks for Consideration

After demographic weighting, we still see a difference in representativeness of Religious Affiliation and Frequency of Internet Use among the online opt-in panel sample compared to data from PEW NPORS 2022.





If we add any or all of these variables to the weighting scheme, will it improve representativeness of measures across other topics?

Benchmark Questions for Testing

Source: PEW NPORS 2022

Volunteerism

In the past 12 months, did you spend any time volunteering for any organization or association? (This includes activities people may not think of, such as infrequent activities or for children's schools.)

- Yes
- No

Political Party Affiliation

Changing topics...Regardless of how you may vote, what do you usually consider yourself?

- Republican
- Democrat
- Independent
- Other

Religious Affiliation

What is your present religion, if any?

- Protestant
- Roman Catholic
- Mormon
- Orthodox
- Jewish
- Muslim
- Buddhist
- Hindu
- Atheist
- Agnostic
- Something else, Specify:_____
- Nothing in particular

Frequency of Internet Use

About how often do you use the internet?

- Almost constantly
- Several times a day
- About once a day
- Several times a week
- Less often

Measuring Error - Measures From Reputable Sources

Respondents were asked a range of questions on topics including:

- Sense of Security
- Sense of Community
- Altruism
- Political Views and Engagement
- News Consumption, Opinion Elite & Other Engagement
- Chronic Health Conditions
- Health Related Behaviors
- Internet & Social Media
- Demographics

External benchmark comparisons to the following sources were used for over 50 of the survey items:

US CENSUS

Census Pulse Week 48: July 27 - August 8, 2022

Census Voting and Registration Supplement 2020

Current Population Survey (CPS) 2021

CPS Volunteer Supplement 2019

PEW

PEW National Public Opinion Reference Survey (NPORS) 2021, 2022

PEW American Trends Panel (ATP) Wave 74 Online Harassment

PEW March 13-27 and April 4-18, 2017

PEW 2018 National Survey of Latinos

NHIS

NHIS 2018, NHIS 2019, NHIS 2020, NHIS Early Release Jul-Dec 2021

OTHER

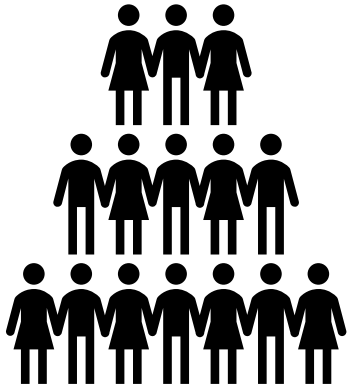
GSS 2021

CNN Politics

Experian

Gallup

Survey Design



We interviewed 8,524 US adults age 18+ across ten different online opt-in sample provider blends from August 16, 2022, through August 30, 2022.

The sample sizes for each ranged from n=850 to n=858.

Data were RIM/rake weighted in total to population proportions from the Current Population Survey (CPS) 2021 for:

- Education
- Age by Gender
- Race/Ethnicity
- Region
- Household Income
- Household Size
- Marital Status

Individual weights were capped at 5 and 0.2.

Weighting Scheme – Demo Weighting Only

Standard Demographics

Source: Current Population Survey (CPS) 2021

Age

11.3%	18-24
17.9%	25-34
16.6%	35-44
15.7%	45-54
16.5%	55-64
22.0%	65+

Household Income

18.7%	Less than $35,000
10.4%	$35,000-$49,999
16.3%	$50,000-$74,999
13.3%	$75,000-$99,999
18.2%	$100,000-$149,999
10.3%	$150,000-$199,999
12.8%	$200,000 or more
-	Prefer not to answer

Race/Ethnicity

16.9%	Hispanic
12.0%	Black only (not Hispanic)
6.1%	Asian only (not Hispanic)
65.0%	All Other (not Hispanic)

Education

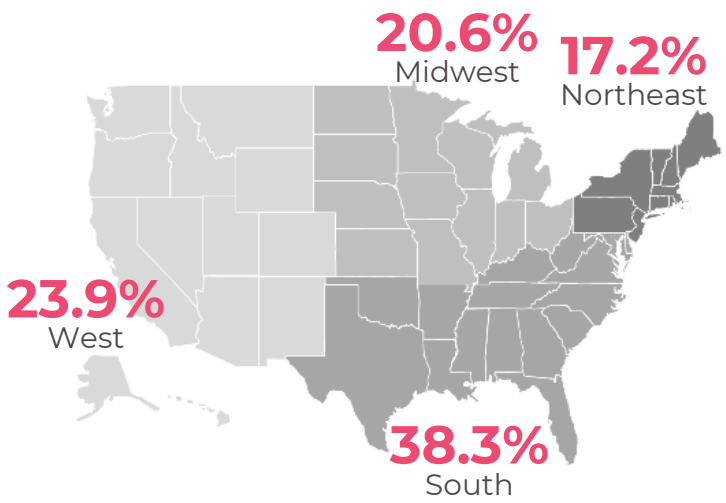
9.6%	Less than High School Degree
55.4%	High School Degree to Less than Bachelor Degree
35.0%	Bachelor Degree or Higher

Gender

48.4% male

51.6% female

Region



Marital Status

51.8%	Married / living with partner
30.2%	Never married
18.0%	Divorced / separated / widowed

Household Size

14.6%	1 HH Member
35.2%	2 HH Members
18.8%	3 HH Members
16.7%	4 HH Members
14.7%	5+ HH Members

Weighting Scheme – Additional Attitudinal and Behavioral Variables

Additional Attitudinal and Behavioral Variables

Source: PEW NPORS 2022

Volunteerism

41.0%	Yes, volunteered in past 12 months
59.0%	No, did not volunteer in past 12 months

Religious Affiliation

41.5%	Protestant
17.8%	Roman Catholic
9.1%	Other Religion
31.6%	Agnostic, Atheist, Nothing in Particular

Political Party Affiliation

30.9%	Republican
30.0%	Democrat
25.5%	Independent
13.6%	Other

Frequency of Internet Use

44.1%	Almost Constantly
44.2%	Several Times a Day
5.3%	About Once a Day
3.7%	Several Times a Week
2.7%	Less Often

Research Design - Testing of the Additional Variables

Control	Demo Only	
Test 1	Demo +	Volunteer
Test 2		Political Party
Test 3		Religion
Test 4		Internet Frequency
Test 5	Demo +	Volunteer + Political Party
Test 6		Volunteer + Religion
Test 7		Volunteer + Internet Frequency
Test 8		Political Party + Religion
Test 9		Political Party + Internet Frequency
Test 10		Religion + Internet Frequency
Test 11	Demo +	Volunteer + Political Party + Religion
Test 12		Volunteer + Political Party + Internet Frequency
Test 13		Volunteer + Religion + Internet Frequency
Test 14		Political Party + Religion + Internet Frequency
Test 15	Demo +	Volunteer + Political Party + Religion + Internet Frequency

Measuring Error - Mean Absolute Error (MAE)

Calculate Mean Absolute Error (MAE) across the variables in each topic area to determine representativeness

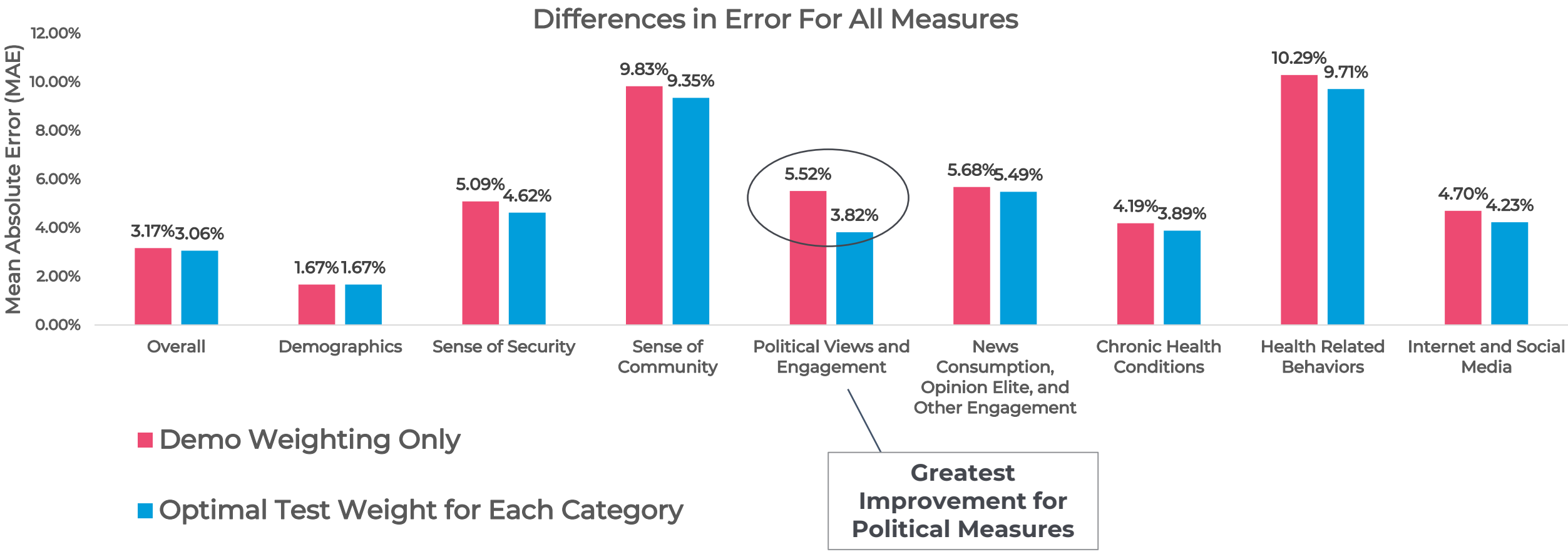
Example: Calculating the MAE for the Health Related Behaviors Category: $MAE = (1/n) * \Sigma|y_i - x_i|$

Variable	Response	Benchmark	Test 9 Weight (Party ID, Internet Freq)	Absolute Difference
Smoked at least 100 Cigarettes	Yes	35.2%	44.5%	9.3%
	No	64.8%	55.5%	9.3%
Currently Smoke	Every day	26.8%	39.5%	12.7%
	Some days	8.7%	16.6%	7.9%
	Not at all	64.5%	43.9%	20.6%
Smoking Status	Current every day smoker	9.4%	17.6%	8.2%
	Current some day smoker	3.0%	7.4%	4.4%
	Former smoker	22.7%	19.5%	3.2%
	Never smoker	64.9%	55.5%	9.4%
E-Cigarette Usage	Yes	17.4%	33.9%	16.5%
	No	82.6%	66.1%	16.5%
Flu Vaccination	Yes	47.9%	53.1%	5.2%
	No	52.1%	46.9%	5.2%
Covid-19 Vaccination	Yes	82.4%	75.6%	6.8%
	No	17.6%	24.4%	6.8%
Tested Positive for Covid-19	Yes	44.9%	34.9%	10.0%
	No	55.1%	65.1%	10.0%
Past 2 Week Sleep Challenges	Not at all	68.5%	38.9%	29.6%
	Several days	18.1%	36.5%	18.4%
	More than half the days	4.8%	12.9%	8.1%
	Nearly every day	8.6%	11.8%	3.2%
Covered by Health Insurance	Yes	89.3%	90.3%	1.0%
	No	10.7%	9.7%	1.0%
			Mean of the Absolute Difference:	9.7%

(n=23)

Improvement in Representativeness

We see the MAE decrease across categories when attitudinal and behavioral items are added to the Weighting Scheme.



Improvement in Representativeness

Optimal Test Weight for each Category (lowest MAE)

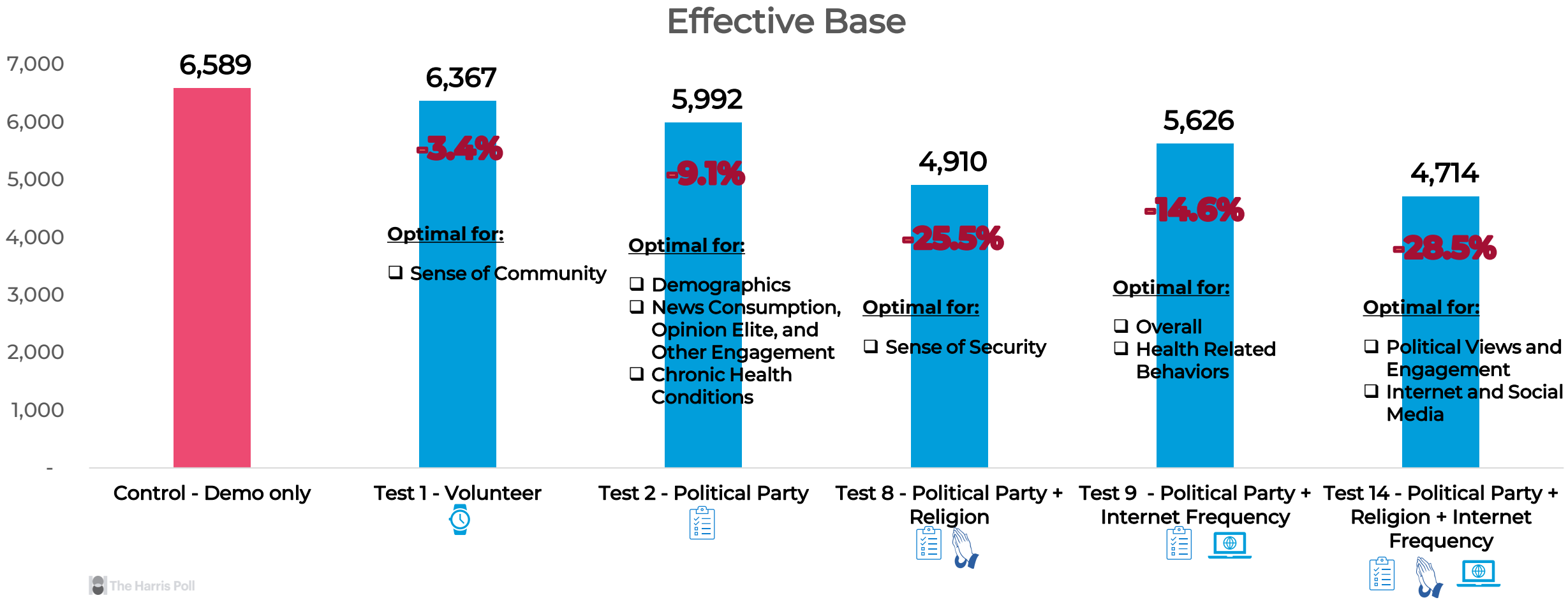
Question Topic Category	Optimal Test Weight
Overall	Test 9 - Political Party + Internet Frequency
Demographics	Test 2 - Political Party
Sense of Security	Test 8 - Political Party + Religion
Sense of Community	Test 1 - Volunteer
Political Views and Engagement	Test 14 - Political Party + Religion + Internet Frequency
News Consumption, Opinion Elite, and Other Engagement	Test 2 - Political Party
Chronic Health Conditions	Test 2 - Political Party
Health Related Behaviors	Test 9 - Political Party + Internet Frequency
Internet and Social Media	Test 14 - Political Party + Religion + Internet Frequency

Notably Political Party Affiliation is present in nearly all the Optimal Test Weights across topics

Consequence of Additional Weight Variable(s)

What are we sacrificing by including additional variable(s) in the weighting?

A drop in Effective Base…

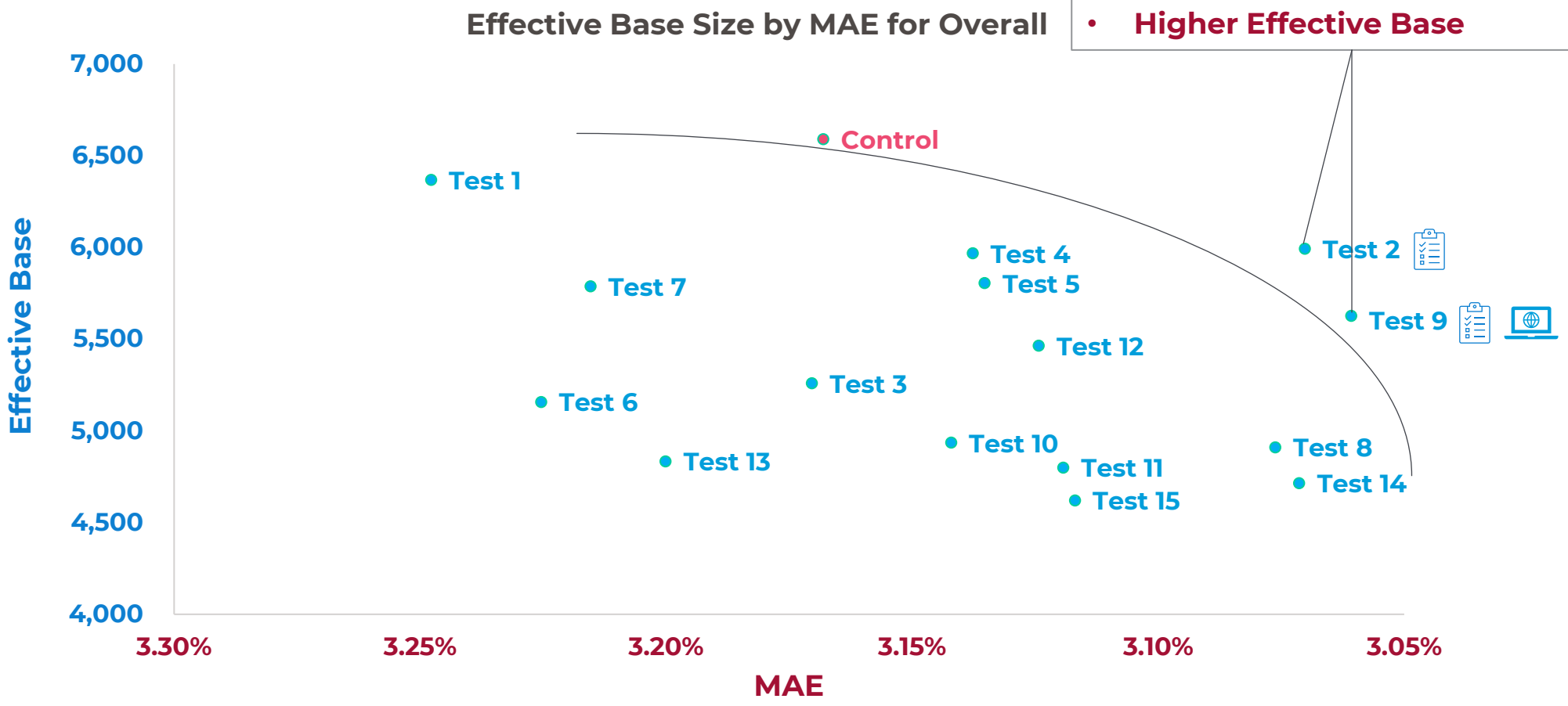


Finding the Right Balance

The Test weight with the lowest MAE might not always be the best choice

Test 2 (Political Party) might be better option than Test 9 (Political Party + Internet Frequency):

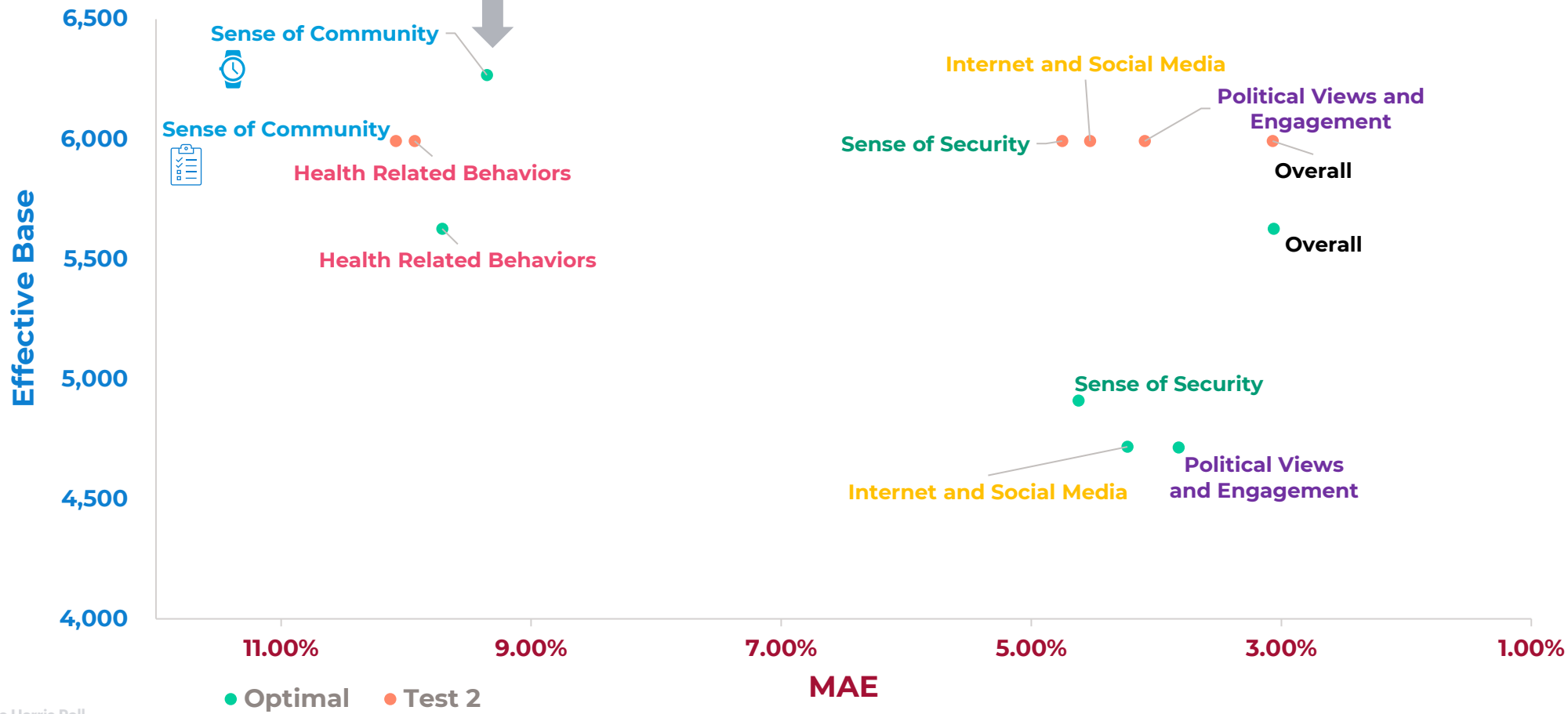
- **One versus Two Variables**
- **Higher Effective Base**



Finding the Right Balance

Test 2 Weight (Political Party) versus Optimal Test Weight by Category

Optimal Test Weight for Sense of Community (Volunteer) might still be best option over Test 2 Weight (Political Party)



Findings and Future Research

Findings

Adding attitudinal and behavioral variables had a positive impact on the representativeness of other measures and the optimal combination varied by topic

Future Research

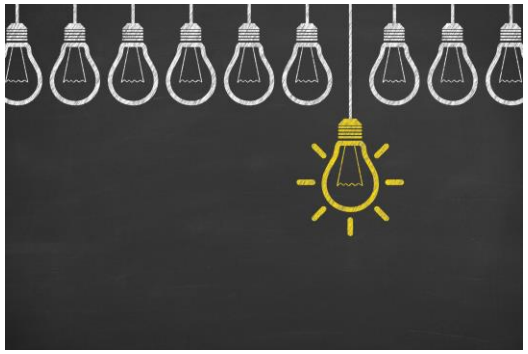
More work can be done pinpointing optimal Attitudinal and Behavioral variables to include in weighting schemes to bolster representativeness beyond demographics alone

Do optimal variable combinations vary by subgroups of the population?

Considerations

Consider trade-off of increased representativeness with respondent burden and lower effective base for statistical tests

Remember the benchmarks are estimates as well so error around those too



Thank You

For more information, visit

theharrispoll.com

Alyssa.Haskins@harrispoll.com

/harris-poll

/theharrispoll

@harrispoll

The Harris Poll