SMS Nonresponse Experiment Findings

Introduction

One of the major challenges facing polling today is the dramatic decline in response rates, which have the potential to exacerbate nonresponse biases in survey responses if left unmitigated. Nonresponse bias in surveys is present when a core assumption of ignorability or conditional independence is violated whereby survey response is found to be correlated or dependent on survey outcomes after controls via respondent weighting.

We conducted an experiment to identify whether our current demographic and ideological controls are enough to adjust for factors that influence both inclusion in the poll and the survey outcome of interest. We sought to quantify the influence of partisan factors, if any, on inclusion and thereby understand the significance of resulting bias in survey outcomes.

Our nonresponse experiment sought to achieve two main goals:

1. Identify characteristics of high-intensity survey respondents to inversely determine profiles of likely nonrespondents.
2. Analyze the correlation between nonresponse factors and survey outcome when applying current demographic controls.

Experimental Design

We sought to identify high-intensity respondents by recontacting sampled voters multiple times and observing response frequency. We studied the relationship between a variety of political engagement indicators, survey responses, personality profiles, and demographic characteristics with high response propensity to inversely identify profiles of nonrespondents.

Sample Design

We ran the nonresponse experiment on a **random national sample of 98,176 registered voters stratified by gender, race, education, age, urbanicity, and level of regional support for Biden** in the 2020 presidential election. Further details on the sample composition can be found in the Appendix.

The experiment employed the use of SMS to facilitate recontacting efforts and improve likelihood of matching to a national voter file. Each sampled person was contacted five times via an SMS with a short message and an external link to the survey detailed in the Appendix. SMS recipients had the opportunity to opt out of receiving further messages at anytime and could choose to stop taking a survey at any time without completing all the questions.

Survey Design

The surveys in each round were identical in content, with a maximum of 44 questions depending on the display logic, and took the median respondent less than 9 minutes to complete. The questions ranged from our regular demographic and political ideology question topics to items related to work, personality traits, and other attitudes beyond our typical scope of inquiries.

On each round, we contacted the same people, with the exception of those who opted out. To minimize the risk of being flagged as spam, we changed both the content of the message and the link to the survey in each round. The only change in the surveys were due to response randomization to minimize acquiescence bias and inattention.

Data Summary

From the 98,176 sampled, we received 1,001 unique respondents with at least one fully complete survey. We consider three populations of respondents:

1. **Out-of-Survey Completion Population:** population of sampled individuals regardless of response. No data cleaning is applied to this population.
2. **In-Survey Full Survey Completion Population:** population of survey respondents that fully completed the survey with data cleaning applied.
3. **In-Survey Partial Survey Completion Population:** population of survey respondents with partial survey completion with data cleaning applied. Only respondents with 29% progress are included, ensuring responses to at least the first partisan question.

The following table describes the in-survey population distribution by responder type before and after data cleaning. We see a small fraction respond more than one time throughout the entire experiment, and we see the presence of late participants who respond to a survey only after multiple recontacts. In each round we mostly get unique respondents, although the overall response rate declines with each pass through the sampling frame.

	Count Completion (No Data Cleaning)	Count Completion After Cleaning Process	Count After Cleaning for Partial Progress
Complete Only in Round 1	330	317	394
Complete Only in Round 2	192	188	226
Complete Only in Round 3	151	139	177
Complete Only in Round 4	130	121	148
Complete Only in Round 5	129	119	140
Two Completes	58	55	126
Three Completes	6	5	20
Four Completes	4	3	2
Five Completes	1	2	4
Total	**1,001**	**949**	**1,238**

The experiment had 15,465 people who opted out, or about 18% of the sample, while response rates neared 1.6%. Opting out gives us information about a much larger proportion of our sampling pool than response does.

	Count	Percentage
Opted Out, 0 Completes	15,293	16%
Did Not Opt Out, 0 Completes	79,124	83%
Opted Out After at Least One Complete	172	0.18%
Did Not Opt Out, One Complete	764	0.8%
Did Not Opt Out, Multiple Completes	65	0.06%

People who gave us a full complete and then opted out later also tend to mostly be single respondents. However, there were some who gave us two, three and even four completes before calling finally opting out of receiving further SMS.

	Opt-Out Count
Complete Only in Round 1	102
Complete Only in Round 2	42
Complete Only in Round 3	21
Complete Only in Round 4	3
Complete Only in Round 5	0
Two Completes	3
Three Completes	1
Four Completes	1
Five Completes	0

Considering partial responses at different completion percentage thresholds, we see that there are a lot of examples of drop-off for all levels except the most extreme multi-respondents.

	Count of Capcha or Further Partials	Count of 2020 Recall or Further Partials	Count of GCB or Further Partials	Count of >75% Partials	Count of Full Completes
One	2,480	1,351	1,172	1,032	934
Two	262	106	93	73	58
Three	37	16	12	10	6
Four	5	3	4	4	4
Five	2	2	1	1	1

Taken altogether, the respondents we see some differences in response behaviors by mode. While in web panels it appears that respondents have an unlimited appetite for survey taking, when we compare with SMS we found that these super responders must account for only an incredibly small fraction of the population. Building a panel selects strongly on interest in taking surveys, and there also could be something about the empaneling process that potentially increases people's interest in taking surveys.

Data Cleaning

Over the five rounds of texting, we received a total of 3,262 survey responses from 2,851 different people in our sample. To analyze the data for survey responses, we applied some basic cleaning that mirrors our current internal process for publishing surveys. We ask a couple of questions that are intended to catch inattentive or poor-quality respondents who answer with falsehoods. We describe the steps and relevant counts that resulted from the process in the following table.

Filter	Full Survey Completion Population	Partial Survey Completion Population
Completion Filter	1,110	2,919

Filter	Full Survey Completion Population	Partial Survey Completion Population
Deduplication	1,019	2,851
Dropping Liars	949	1,238

In-Survey Data

In this section we provide some definitions and basic summary statistics for the questions that are unique to the experiment and outside the typical scope of our polling. We included these questions with the aim of identifying qualities in our respondents that we hypothesized might be related to survey response propensity.

Occupational Requirements

We measure the level of physicality involved in respondents' jobs using a paired-down version of the U.S. Bureau of Labor Statistics (BLS) measure. These occupational questions are only asked to the subset of respondents who state that they work full- or part-time jobs when asked about their employment status. This represents 681 (55.0%) of partial survey respondents and 563 (59.3%) of respondents who fully complete the survey.

We ask a series of questions regarding the portion of the work day spent standing, carrying or lifting different amounts of weight, and group respondents' occupational strength level into five categories: sedentary, light, medium, heavy, or very heavy work using the the breakdown described in the following table.

Strength Level	**Duration of Lifting or Carrying**		
	Less than 1/3 of the work day	**Between 1/3 to 2/3 of the work day**	**Between 2/3 to the full work day**
Sedentary work	I do not lift or carry any weight	I do not lift or carry any weight	I do not lift or carry any weight
Light work	Any weight	-	-
Medium work	More than 50 pounds	Less than 50 pounds	Less than 50 pounds
Heavy work	-	More than 50 pounds	-
Very heavy work	-	-	More than 50 pounds

The following table shows the distribution by occupational strength requirements of our survey respondents by population. In the fourth column, we include how BLS measures compare for the U.S. civilian worker population. It is important to highlight that the BLS population is not directly comparable to our two in-survey populations for the following reasons:

1. Our sample comes from a registered voter environment while the BLS data universe expands to the wider adult population.

2. We used a more diluted set of questions to categorize the physicality of work to reduce the number of questions and subsequent burden on respondents.

Occupational Strength Requirements	Full Survey Completion	Partial Survey Completion	BLS Civilian Workers
Sedentary	320 (56.8%)	363 (55.2%)	29.1%
Light	135 (24.0%)	158 (24.0%)	32.8%
Medium	73 (13.0%)	89 (13.5%)	28.4%
Heavy	26 (4.6%)	33 (5.0%)	8.7%
Very Heavy	9 (1.6%)	14 (2.1%)	1.0%
NA	0	24 (3.7%)	-
Total	563	657	100%

Our experiment also included questions related to how much time a person spends on the computer, their managerial level, income, and industry. These questions aim to capture how daily work conditions relate to people's interest or salience in taking surveys.

Loneliness Score

For the nonresponse experiment we employed UCLA's loneliness score methodology, which uses a series of questions to calculate a respondent's level of loneliness. We know that this metric is reliable and consistent and allows us to assign a score to each response and create an **average loneliness score**. We asked a series of three questions regarding the frequency with which respondents express feeling isolated, left-out, or lacking companionship. Responses of *Hardly ever or never* received a score of 1, *Some of the time* a score of 2, and *Often* a score of 3.

Final loneliness scores range between 3 (highest loneliness level) and 1 (lowest loneliness level). Note that for the population of Partial Survey Completion, not all observed respondents had completed the necessary questions to calculate a loneliness score; as such, these folks have missing loneliness scores for the purposes of this analysis.

Loneliness Score	Full Survey Completion	Partial Survey Completion
1-1.4	545 (57.4%)	564 (45.6%)
1.5-1.9	137 (14.4%)	140 (11.3%)
2-2.4	203 (21.4%)	209 (16.9%)
2.5-3	64 (6.7%)	67 (5.4%)
NA	0	258 (20.8%)
Total	**949**	**1,238**

Online Concern Scores

We hypothesized that an aversion to spam or clicking links from unknown sources might deter sampled people from participating in our SMS polling. We designed two questions regarding online

privacy concerns and trust in online identity using the guidance in this comprehensive review of instruments for measuring privacy concerns. The results are scored as shown in the following table and averaged for a final online concern score.

Response	Score	Privacy Online	Identity Online
Not at all	1	88	142
Somewhat	2	452	449
Very much	3	409	358

We saw that the vast majority of our respondents claim to be somewhat or very much concerned with online privacy and identity. This was also reflected in open-end responses, where respondents often requested more information about the surveyor, the purposes for the information gathered, and some even mentioned only responding after looking up our information online after multiple contacts.

Personality Scores

The TIPI (Ten-Item Personality Inventory) uses a series of 10 questions to score respondents on each of five personality dimensions: *Openness to experience, Conscientiousness, Agreeableness, Extraversion,* and *Emotional Stability*. Studies have shown that these five personality traits can predict a range of attitudes and behaviors, some of which are particularly relevant in the world of SMS survey response. For example, personality is associated with distinct tastes for social versus solitary activities, such as an online survey without direct personal contact. In a web panel experiment, researchers found that unit nonresponse was higher among those low in conscientiousness and high in openness to experience, while those with higher extraversion and agreeableness scores were positively correlated with joining an online panel.

Some of the Big Five traits have also been identified with differences in political preference with those who are highly open to new experiences tending towards more liberal views. As such, personality traits as described by the Big Five model have the potential to impact response propensity, as well as survey outcome.

Time Use

We consolidated some of the main groups from the BLS American Time Use Survey to understand how our survey respondents spend their time. We asked generally about time spent in the span of a normal week on work, education, household and care activities, and leisure.

Political Preferences and Attitudes

We also asked our regular political ideology and preferences questions to track survey outcome bias due to nonresponse biases. We asked the following questions to gauge political identity, attitudes, and preferences in the current political climate.

1. **Generic Candidate Favorability:** We asked respondents whether they have a favorable or unfavorable opinion of Joe Biden and Donald Trump on a five-point Likert scale ranging from Very Favorable, Somewhat Favorable, Somewhat Unfavorable, Very Favorable, and finally Haven't Heard Enough to Say. We binarize responses into favorable and unfavorable for ease of analysis.

2. **Generic Congressional Ballot:** In this question we asked respondents to consider how they would vote in a congressional election if there were to be an election held today. Response options include The Democrat, The Republican, and Don't Know.
3. **Partisanship:** This question is similar to the GCB question, but asked about the respondent's personal identity regarding the two major parties. We asked if they see themselves as a Democrat, Republican, Independent, or Something Else.

Outside Data Sources

In addition to survey data that we collect in-survey, we relied on national voter file and commercial data available for the full sample regardless of response. This data can be broadly categorized as indicators for political engagement.

- **Recent Donation Data:** We consolidate political donation data from several sources and match them back to individual registered voters in our sample. We consider only political donations made to a party after Jan. 1, 2020.
- **Historic Election Participation:** The voter file has an accurate history of registered voters' participation in each election. We consider the following measures of vote history:
    - **Recent primary voter:** Here again, recent is defined as having voted in a primary in the 2020 cycle or later.
    - **General Election Participation:** Number of times an individual has voted in the last 4 federal general elections, namely 2022, 2020, 2018, and 2016.
- **Historic Survey Participation:** We have accumulated years of cross-modal survey data and matched voter registration data to survey response to the best of our abilities. We use this consolidated survey response history as a unique indicator of political engagement.
    - **Past SMS respondent.** This includes individuals who had completed a survey via our SMS program prior to this experiment.
    - **Past non-SMS respondent.** This includes individuals who had completed a survey by us using a mode other than SMS prior to this experiment. These are mostly web respondents whom we had matched back to the voter file.
- **Opt-Out From the SMS Survey Experiment:** We offer folks the opportunity to opt out of receiving further messages from us.

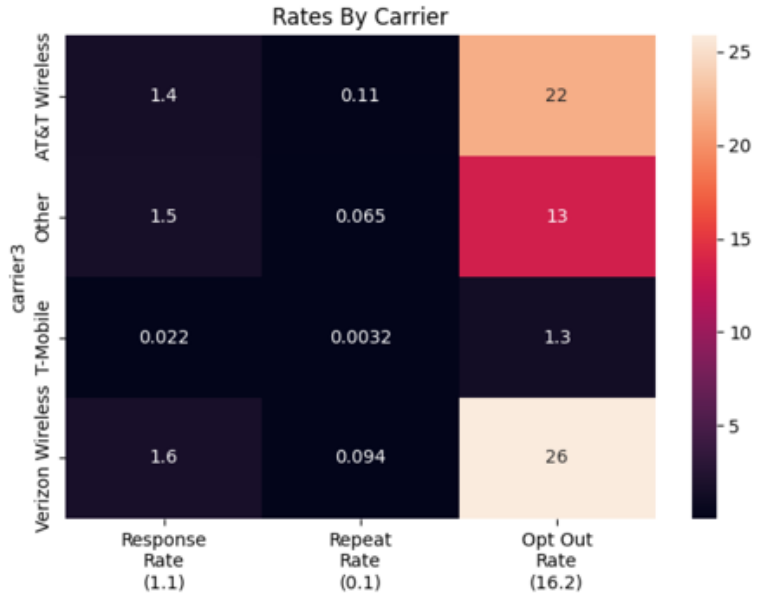
Additional Data Considerations

Finally, it is important to note that we found some SMS deliverability issues that inevitably affect how frequently sampled people were able to be contacted. By the fifth round of contact, we appeared to have delivered 81,324 SMS messages, representing about 82.9% of our original sample due to invalid phone numbers, opting out, and other deliverability issues.

Our sample is represented by three main carriers: Verizon, T-Mobile, and AT&T Wireless at about an even split as shown in the table below.

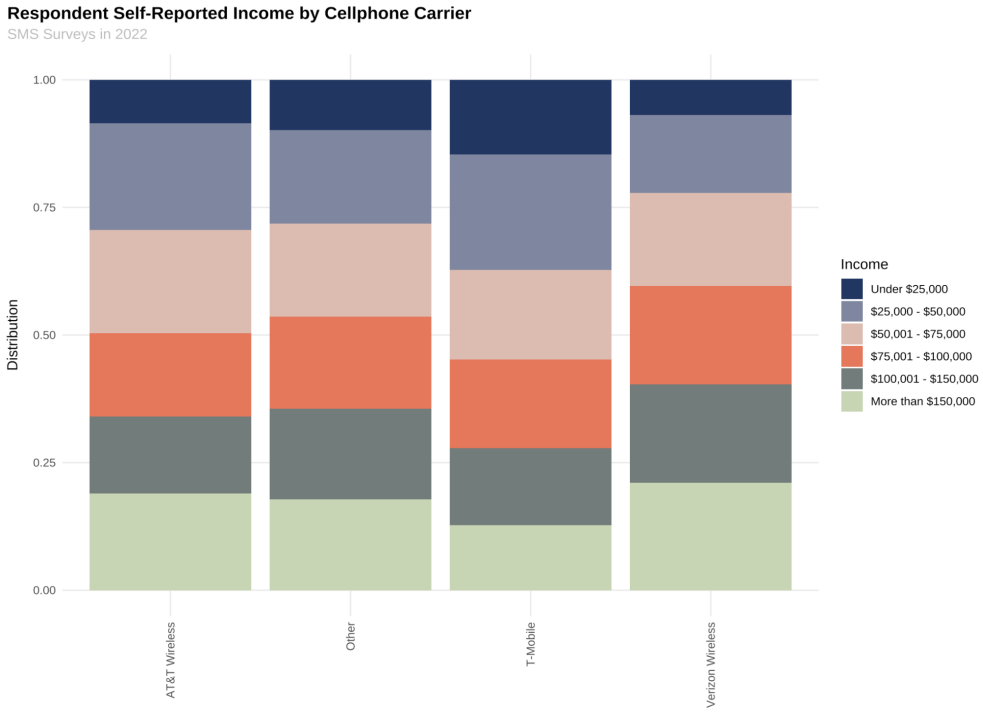
Carrier	Number Sampled	Proportion of Sample
Verizon Wireless	33,981	34.6%
T-Mobile	32,184	32.8%
AT&T Wireless	28,869	29.4%
Other	3,128	3.2%

However, looking at the response rates within the survey experiment, repeat survey rate in our historical surveys, and opt-out rates for the experiment, we found extremely low rates uniquely among our sample with the T-Mobile carrier.

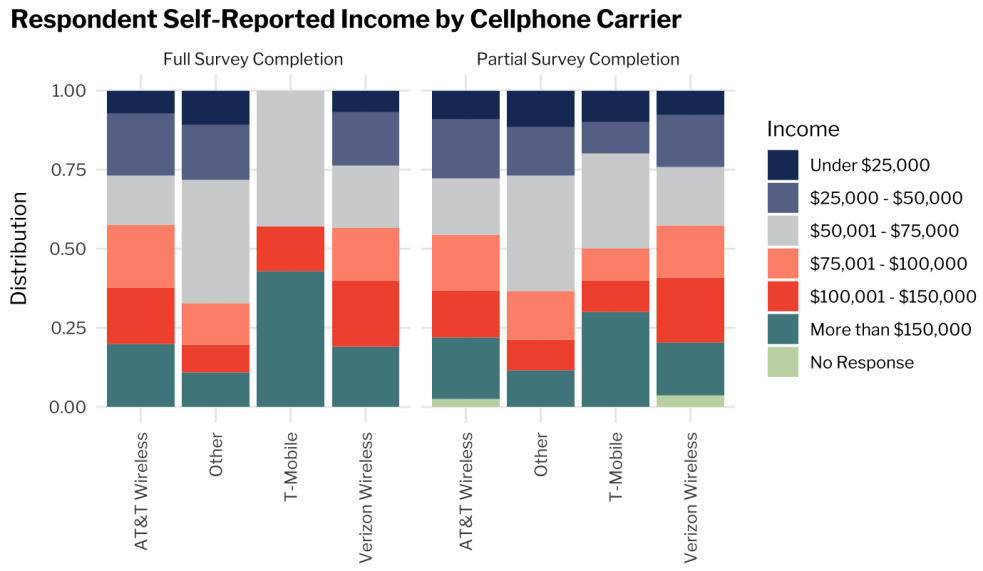


We registered low interaction across several measures with T-Mobile users even as deliverability rates for all carriers were about even. At the time the messages were sent, we were contacting respondents using a local phone number, for which T-Mobile applies the most conservative policies for blocking delivery of messages among carriers. For future polling, we have switched our phone number to a toll-free number and seen improvement.

Furthermore, carriers have an underlying bias to consider: T-Mobile users represent a lower socioeconomic status than users of other carriers. In the following graph we see the distribution of self-reported income for SMS respondents in national surveys in 2022.



When we compare the same breakdown among respondents within the experiment, T-Mobile respondents are different than the rest of the carriers and historic T-Mobile respondents by income. We see very low rates of T-Mobile response and higher self-reported income than respondents from other carriers.



Results

This section of the report is structured as follows. First, in the Out-of-Survey analysis, we examine how response rates vary along observable dimensions in the voter file, as well as along proxies for unobservable response interest. These proxies are whether or not an individual completes the survey multiple times, as well as an instrumental variable approach, where we alter the criteria for inclusion in the survey based on whether an individual responded to the first invitation to the poll or to one of the four subsequent invitations. This section does not make use of any in-survey items and is solely aimed at quantifying variation in response rates. T-Mobile users were excluded from this analysis due to the carrier's previously detailed deliverability issues, which influence response rates.

In the In-Survey nonresponse section, we examine how variation in response interest, as proxied by the number of times a respondent completed a survey, is correlated with various in-survey items related to work, personality traits, and other attitudes that are not typically used as weighting variables. The goal of this analysis to describe the characteristics which differentiate higher vs. lower response-interest respondents. This descriptive understanding may be useful in devising strategies for adjusting for nonresponse by identifying potential variables and concepts worth attempting to adjust for.

In the outcome analysis, we show how variation in response rates, both along observed and unobserved dimensions, is correlated with the target in-survey outcome variables of Biden approval, Trump approval, and Generic Congressional Ballot.

The key findings are:

- Unobservable and likely highly transient factors appear to be the primary driver of survey response.
    - Nearly all groups of individuals, including those whom we would consider very highly politically engaged, have opt-out rates vastly in excess of response rates.
    - Past response is by far the most influential predictor of response, and the only thing that meaningfully predicts repeat responses. However, despite high response rates relative to the average, past respondents have response rates that are not very high in absolute terms and which are only at parity with opt-out rate.
    - Both the IV approach and the repeat response approach give us a plausible way to quantify variation in underlying response interest beyond standard observables.
- In-survey nonresponse analysis does not yield especially strong signals we can use to describe or explain nonresponse in generalizable ways.
- Outcome analysis indicates non-negligible but highly complex relationships between unobservable drivers of response and all three of our outcomes of interest, even after accounting for standard weighting items such as demographics and 2020 vote recall.

- In particular, we find some evidence to suggest that Biden 2020 voters who are dissatisfied with Biden as well as Trump 2020 voters who are dissatisfied with Trump may be over-represented in SMS surveys.
- Copartisan voters in the generic ballot, meaning Biden 2020 voters who prefer Democrats in the GCB and vice versa for Republicans, also appear overrepresented in SMS surveys
- The second result points to classical theories about partisan nonresponse where bias is produced by hard partisan's enthusiasm for survey response. However, the first suggests something different, where response is associated with in-party dissatisfaction as opposed to in-party fervor. It is plausible that both of these dynamics could coexist.

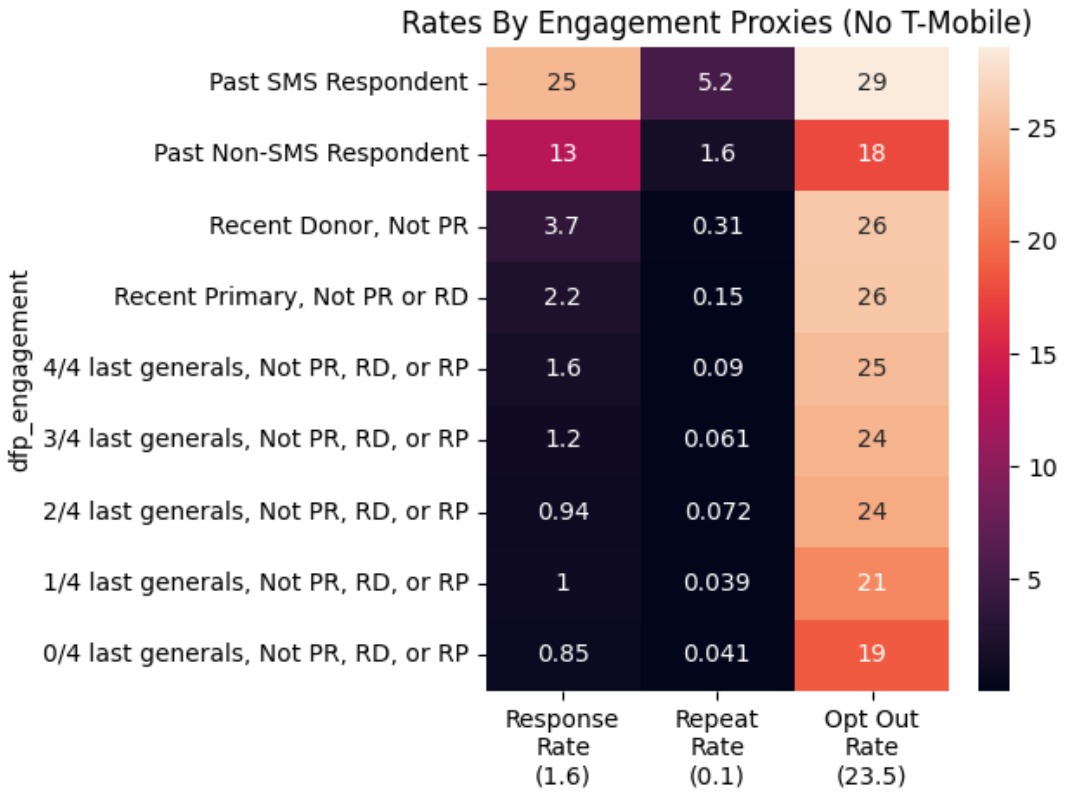
Nonresponse

Out-of-Survey Nonresponse Analysis

Response rates along observable dimensions

At a high level we can summarize the results of the response analysis using the engagement variables previously defined.

Response rate (defined as an individual giving at least one complete from five invitations to the survey), repeat rate (defined as an individual giving more than one complete from five invitations), and opt-out rate are shown below, averaged by this engagement scale.



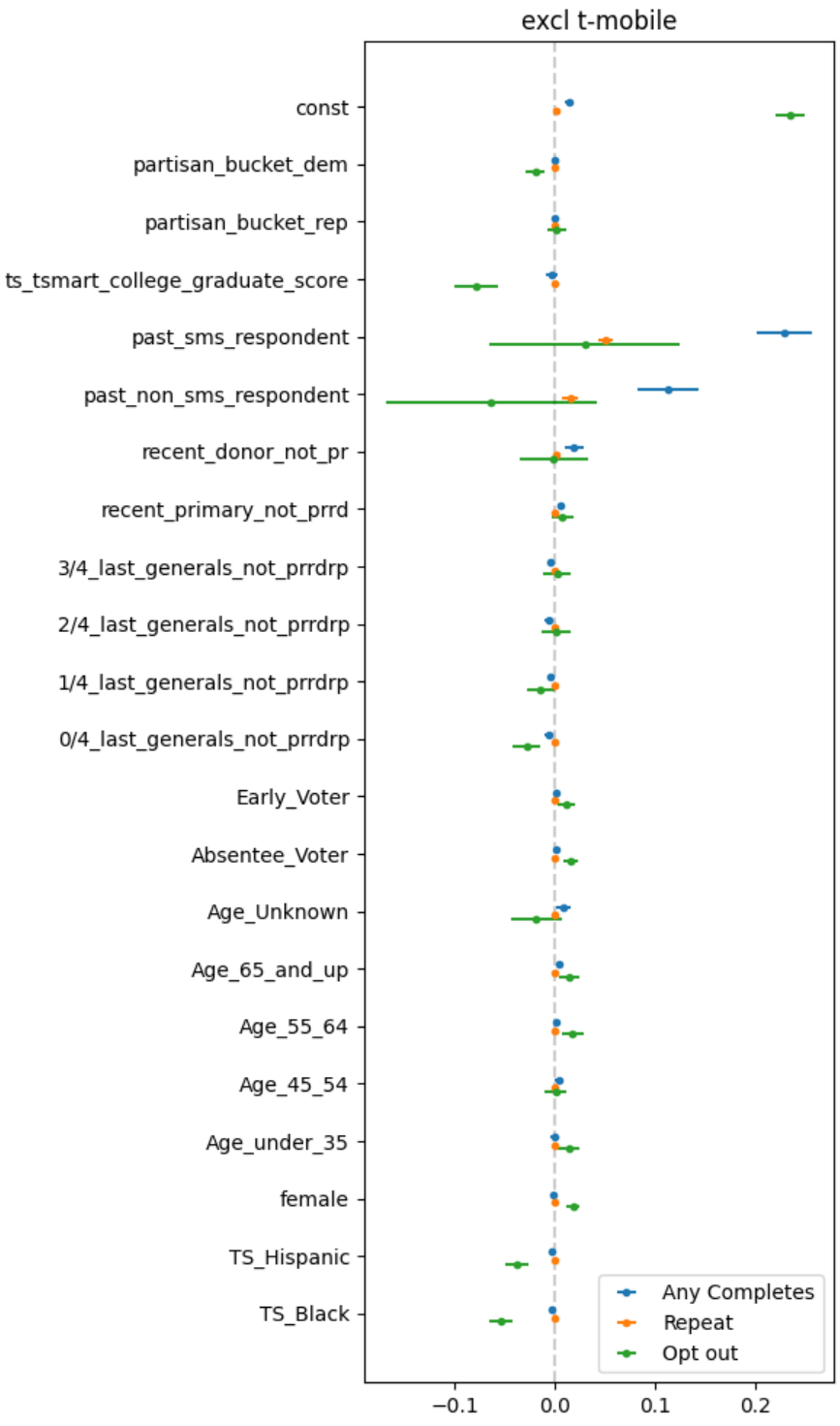
We can readily glean several key results from this table. First, the engagement scale reveals that a person with 4/4 of the last 4 generals but none of the indicators of high engagement has a response rate roughly in line with the global average of 1.6%. Moving up on the engagement scale we see response rates increase sharply, in relative terms, with rates for recent primary voters, recent donors, past non-SMS respondents, and past SMS respondents being ~1.4X, ~2X, ~8X, and ~15X the average rate, respectively. For repeating, the relative rates appear even more extreme, at ~1.5X, ~3X, ~16X, and ~50X for the four high engagement categories, respectively.

This result points to a significantly skewed population distribution of response interest, where a relatively small number of individuals have response propensities significantly in excess of average. Opt-out rates drive the point home further. Even for groups we would consider very high propensity, response is very rare and opting out is more common by a factor of 10 or even more. Among past respondents, response is only roughly at parity with opting out. While it may be true that survey respondents are characterized by indicators of higher than average political interest, the typical highly politically engaged individual has very little interest in responding to a poll. Low absolute response rates and high opt-out rates suggest that there is likely a sizeable portion of the target population who are “never-responders” or individuals with a functionally zero probability to ever responding to an SMS invitation to a poll.

Aside from an extreme population distribution of unobservable response interest, we can also glean that response interest is a highly transient phenomenon. Past response is the best predictor of future

response, but it is far from a guarantee. Behaviors like voting are generally understood to be very durable habits that last a lifetime once a person starts engaging in them, which makes vote history a very powerful predictor of future voter turnout. Participation in political surveys, on the other hand, does not appear to be a durable habit for most people, a concept which is also supported by high attrition and churn rates reported by panel providers.

Below we use OLS as an estimating equation for coefficients for response, repeat, and opt-out rates for a larger set of variables, including bucketed TargetSmart partisanship score, TargetSmart college score, age buckets, voter file gender, early and absentee voter status, and TargetSmart race indicators.



There are a few subtleties in these results not captured in our simple table of response rates by engagement strata. Though again we find that response propensity follows a skewed distribution primarily along the engagement scale, and response history is the only variable which meaningfully predicts repeat response. Opt-out rates are much higher than response rates for nearly everyone. It

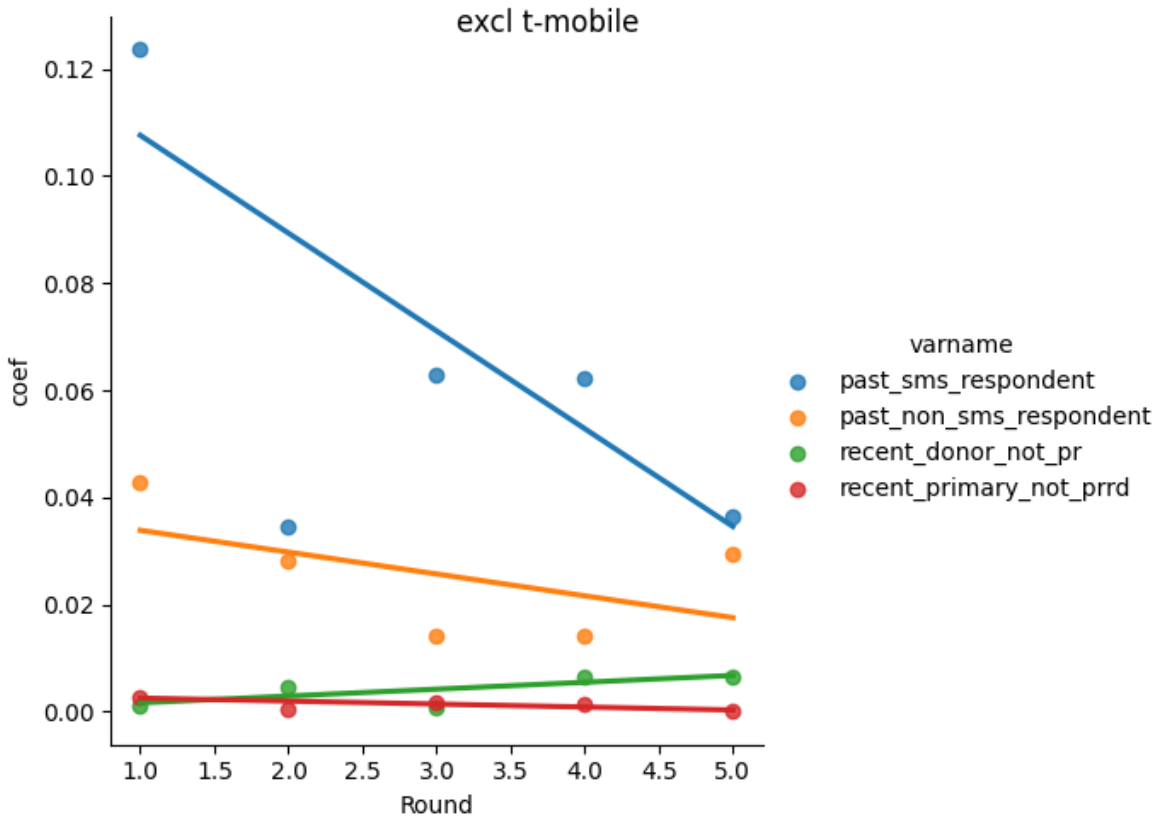
might be the case that some lower-response propensity groups such as Black voters, Hispanic voters, and voters with infrequent turnout histories have lower than average opt-out rates, but it is also possible this is due to some silent non-delivery mechanism, such as having a phone number on file that is no longer valid.

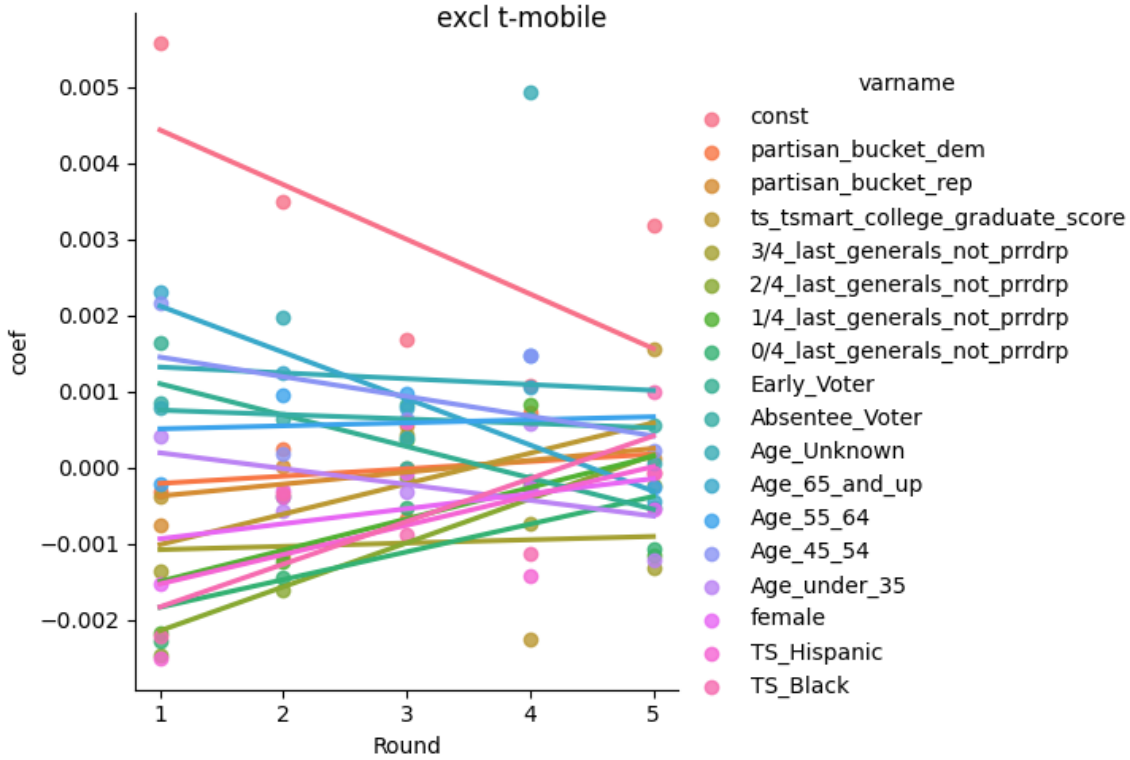
With average response rates declining, we could imagine two extreme scenarios describing the population distribution of response interest. In the first, response interest is close to uniform, perhaps with some variation along observable dimensions, but every individual simply has a low chance of responding if contacted. In the second, response interest is highly skewed where there are a small number of individuals with well above average response interest along with a large number of individuals with low or practically zero response interest.

The results above point mostly toward the latter case, but by examining response rates by each individual sampling round, it seems plausible that there is some subsegment of the population where the uniform but low response probability condition is a more reasonable assumption than it is for the population at large. In each round of sampling we mostly get unique respondents, and overall response rate increases meaningfully with each round, as shown in the table below.

Sampling Round	Cumulative Response Rate	Gain From Last Round
1	0.56%	0.56%
2	0.89%	0.33%
3	1.1%	0.24%
4	1.3%	0.21%
5	1.6%	0.20%

Further, using OLS to estimate coefficients for completes by round, we see that in later rounds non-uniformity in response generally attenuates along observable dimensions, and often significantly so. Note, the figures below are split in two since the Y axis is about a factor of 10 bigger for the variables in the first chart compared to the second.


excl t-mobile
coef
Round
varname
past_sms_respondent
past_non_sms_respondent
recent_donor_not_pr
recent_primary_not_prrd


excl t-mobile
coef
Round
varname
const
partisan_bucket_dem
partisan_bucket_rep
ts_tsmart_college_graduate_score
3/4_last_generals_not_prrdrp
2/4_last_generals_not_prrdrp
1/4_last_generals_not_prrdrp
0/4_last_generals_not_prrdrp
Early_Voter
Absentee_Voter
Age_Unknown
Age_65_and_up
Age_55_64
Age_45_54
Age_under_35
female
TS_Hispanic
TS_Black

This is promising because it suggests that while SMS surveys are dominated by high-interest people, using a little persistence with sending invitations may be a viable way to start to tap into lower response interest groups and mitigate some of the observable bias in response. We will explore this further in the instrumental variable section.

Quantifying Unobservable Response Drivers with Repeat Survey Taking

While responding is rare, repeat responding is even more rare, and is an indicator of exceptionally high response interest. In our original proposal, we set out to quantify this by estimating essentially a fixed-effects type model for response considering a normal population distribution of unobserved individual response propensity driving binomial response. Thus far, we have not been able to implement a model of this type that appears useful. Our response analysis also points to fairly serious defects in the underlying assumptions of such a model.

First, we see that the underlying population distribution of response propensity can not be a normal distribution or perhaps even a single distribution, as opposed to a mixture. Since the incidence of repeat response is so low, the assumption of normal or even long tailed distribution such as Laplace for the unobservable response propensity results in excessive regularization of predicted response propensities.

Second, an implicit assumption of this sort of model is that response propensity does not change much for an individual over time, while we find lots of reason to believe this is not a good assumption. A fair number of single and even repeat respondents ended up opting out at some point in the experiment, and past respondents also have high opt-out rates. Even within the relative short time window of this experiment, we can see what is essentially the attrition, churn, and fatigue dynamics observed over longer periods of time in panels.

With these limitations granted, using repeating or complete count itself as a proxy for high response interest appears very useful. It may be possible to incorporate this into a more formal estimate of response propensity with individual effects given many months worth of panel data, but in the random-contact context, such formal estimates do not appear feasible. It is, however, relatively easy and very inexpensive to collect data on repeat respondents by recanvassing individuals who responded in the first round, which still makes it an attractive option for understanding nonresponse in random contact surveys.

Quantifying Unobservable Response Drivers With an Instrumental Variable

Not long after our experiment was fielded, an article was published advocating for the use of randomized response instruments for quantifying and adjusting for nonresponse bias. A response instrument can be any protocol that can be assigned randomly to potential respondents, and which creates variation in response while not directly influencing outcome. Examples can be a high-incentive vs. low-incentive group, a high-contact effort vs low-contact effort group, “attrition” questions that more or less induce some individuals to opt out of the survey, and a few others.

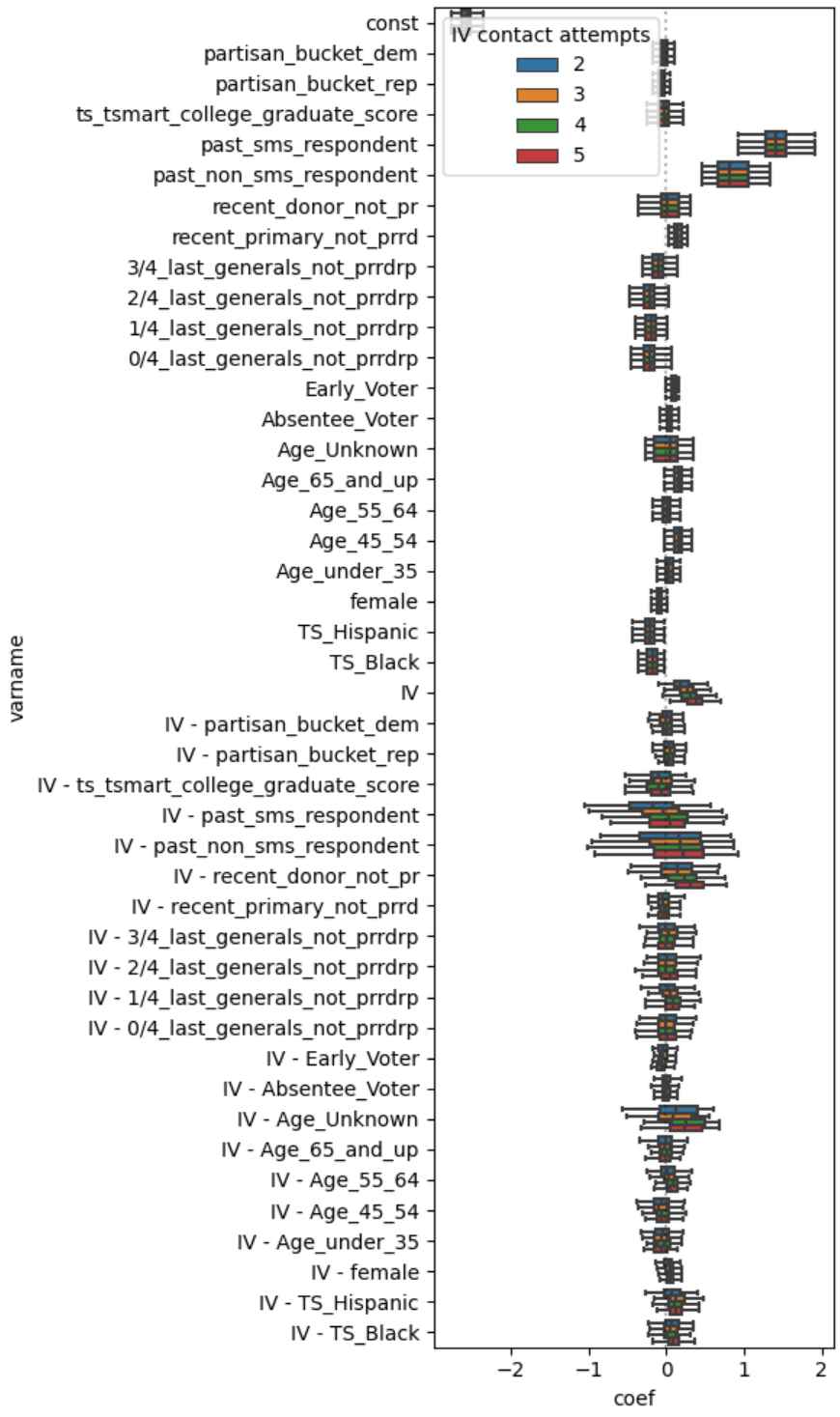
Since we contacted everyone in the sampling frame five times, we can use this as a response instrument. This particular instrument is interesting because, unlike varying incentives or inducing opt-

outs, we can actually observe the counterfactual where different individuals had been randomly assigned different levels of contact effort. Specifically, someone who responds on the first round is considered a respondent at any level of effort. Someone who responds on the second round or later but not the first is a respondent at the two-invitation level of effort or greater.

This allows us to use a bootstrap resampling technique to quantify variation as follows. After defining a maximum contact effort threshold of 2, 3, 4, or 5 contacts, in each bootstrap realization we:

1. Randomly assign everyone in the sampling frame to a high level of contact effort or low level of contact effort group with 50-50 probability.
2. Define the sample with the following inclusion criteria:
    a. For the low level of contact effort, a person is selected as a respondent only if they responded in the first round, any/all responses after that are ignored.
    b. For the high level of contact effort, a person is selected as a respondent if they responded in any round up to the prescribed effort threshold.
3. Fit a probit model for the response indicator given by 2, accounting for voter file covariates, the IV, and interactions thereof.

Below we show the distribution of the probit coefficients over 100 bootstrap realizations, for each of the 4 possible contact thresholds. For all cases, the first set of base coefficients just reflect the first-round response rates, and so are unchanged by the effort threshold defining the IV. The main effect for the IV does indeed increase response rate, and the amount of increase goes monotonically with effort threshold. Variance is pretty high with the interaction terms, and so we can neither rule in nor rule out the idea that increased contact effort may disproportionately increase response rates among already high engagement groups. Since this is potentially important to account for, we use this fully interacted model for outcome analysis along with the bootstrap.



The IV approach is attractive for several reasons. First, it is relatively easy to implement, and could be implemented with any random-contact mode such as live caller or even mail. Cost-wise, it is not cheap, since the minimal version of it would involve contacting the entire sampling frame a minimum of two times, though two contacts with SMS is still likely to be cheaper than live caller or mail. Also,

while there are diminishing returns to additional contacts, given a fixed budget the feasible number of contacts could be scoped from the start.

More importantly, it does appear to give us a plausible instrument which gives us a snapshot of unobservable response interest at a given moment in time. The bootstrap construction is a useful way to quantify the resulting uncertainty. In the outcome analysis section, we look at how this bootstrapped response score with the IV, as well as repeating indicates non-ignorable nonresponse.

In-Survey Nonresponse Analysis

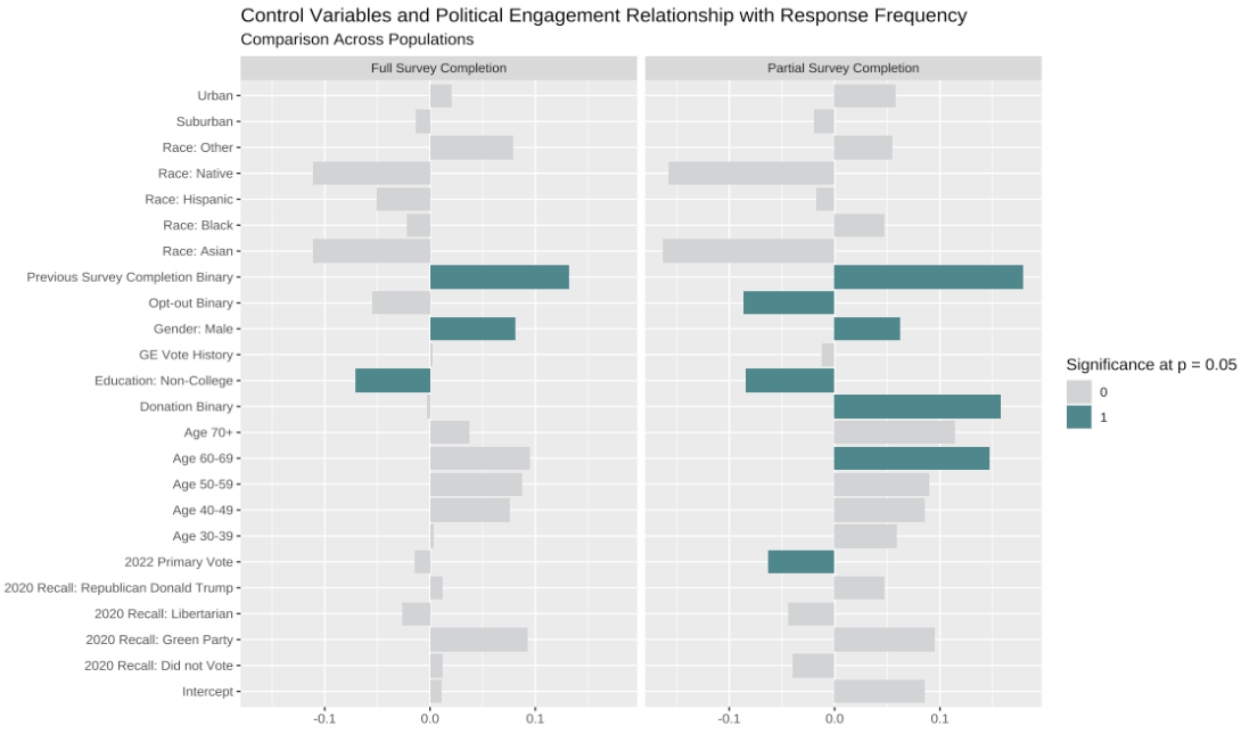
In the in-survey nonresponse analysis we model the number of responses given by each respondent in the sample who responded to the survey at least one time. We use a combination of survey response data, demographic data, and political engagement indicators to learn characteristics of repeated intensity.

We studied effects on response frequency of each individual across the full survey completion and partial survey completion populations.

We found that the response frequency data was underdispersed, violating the assumption of equal mean and variance required for a Poisson regression. As such we applied a quasi-Poisson regression that controls for underdispersion by incorporating a dispersion parameter.

Completion Definition	Average Response Count	Variance of Response Count
Full Survey Completion	1.09	0.15
Partial Survey Completion	1.16	0.24

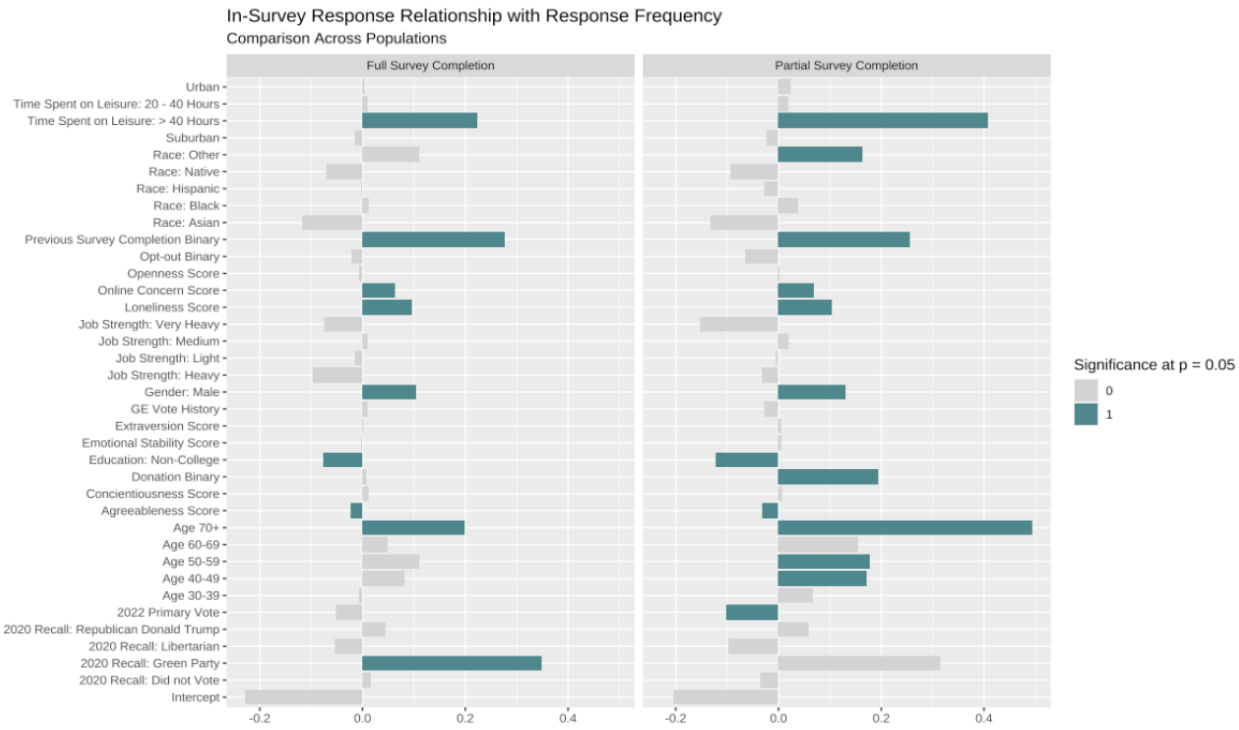
We incorporated our traditional survey weighting control as well as the aforementioned political engagement variables in our in-survey analysis and found that the act of previously completing a survey has a significant and positive relationship with response frequency.



Recent engagement indicators showed significant relationships to survey response frequency in the partial survey completion population. Turnout in the 2022 primary election, and opting out of the experiment had a negative relationship with response frequency, and recent political donations had a positive relationship with response propensity.

Even controlling for these additional *engagement controls,* we still find that there is a higher response frequency among men and college-educated respondents in both populations.

In the following analysis, we look at the individual responses to some of the ideological, psychological, and lifestyle questions that we described.



We find significant positive correlations between a respondent's time spent in leisure activities and response frequency. We also find that folks with higher loneliness scores have higher response frequency within the experiment. Loneliness and amount of leisure time are not correlated but tend to point to the same notion of respondents with more of a flexible time availability for survey completion — either through lack of formal work or social activity.

We also see that low agreeableness is associated with a higher response frequency. Agreeableness relates to an individual's tendency toward trust, conformity, and concern for social harmony. Our survey prompted respondents to click a link to share their views, and some of the first questions are related to political leaning. It is possible that folks who seek conformity or harmony are not interested in engaging with politics or find sharing their opinions less appealing.

Interestingly, we see that individuals with higher concern about online privacy have higher response response frequencies. This contradicts our expectations — whereby those less averse to spam or online piracy would be less likely to be concerned or put off by an SMS asking for personal information.

Covariate Correlations

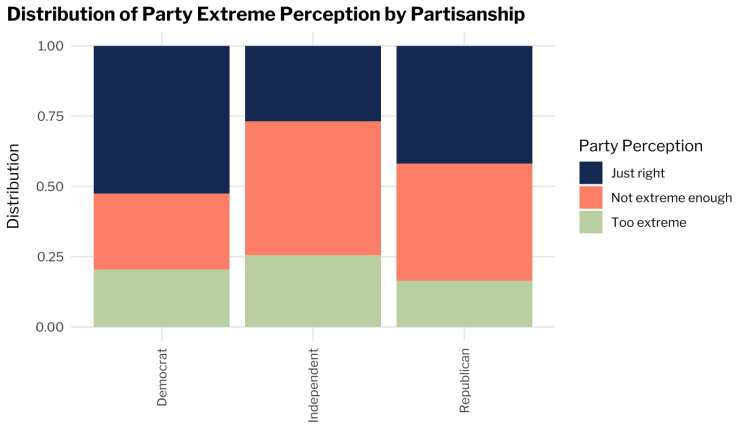
To reduce the influence of multicollinearity in our analysis, we detail covariate correlation analysis we underwent to define our model specifications in the Appendix. From this analysis we uncovered some incongruous relationships between correlated covariates, and their relationships with response frequency. We identified variables that are significantly correlated to pair down factors included in the regression analysis. However interchanging the related variables would on occasion significantly

change final results indicating that the in-survey characteristic conclusions are largely unstable and unreliable.

Party Dissatisfaction

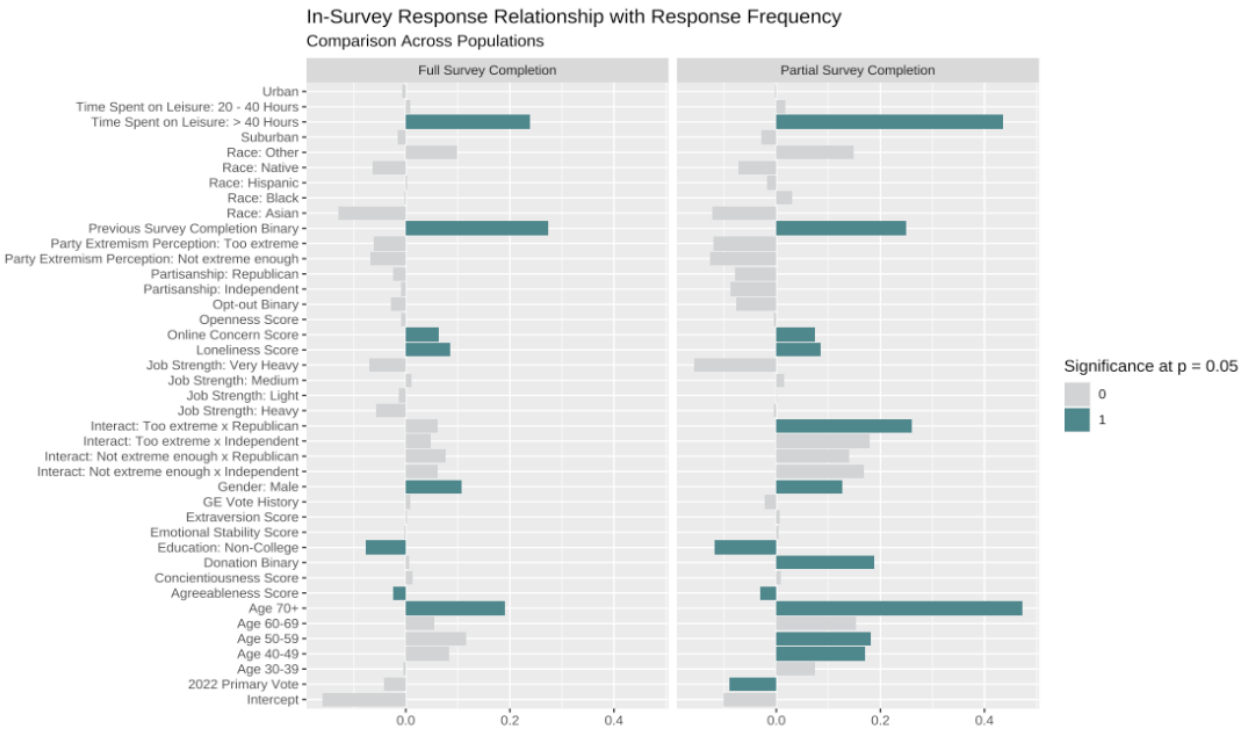
There is a growing narrative in polling regarding the 2024 election concerned with generalized dissatisfaction among voters. The 2024 election represents a “rematch” of the 2020 election, and tracking shifts in each candidate’s base will be key to polling accuracy in the coming election. We also know that the 2020 election polling saw overresponse from activist or hard-partisans that biased respondent pools and ultimately final predictions, so it is important to understand the more extreme representatives of each party.

With this in mind, we asked respondents how they perceived their party’s alignment on the political spectrum, ranging from *Too Liberal, About Right, Too Conservative, Not Conservative Enough*, and *Not Liberal Enough*. We subsequently bucketed Democrats who viewed their party as *Too Liberal* or *Not Conservative Enough* as perceiving their party as *Too Extreme*, and vice versa for Republican respondents.



At a high level we can see that Republican respondents tended to think their party was *Not Extreme Enough* more so than Democratic respondents. On the other hand, Democratic respondents tended to respond that they viewed their party correctly aligned on the political spectrum, or *Too Extreme* at higher rates than their Republican counterparts.

On its own, it is not clear if these tendencies result from compositional differences in each party’s base, or if it is a result of biases in SMS survey response habits. We incorporate an interaction effect between partisanship and party perception to the response frequency regression to analyze if these factions of each party responded at different rates to our surveys.



There does not appear to be partisanship and party perception relationship with response frequency. However, among respondents with partial survey completion, we did see a positive and significant relationship in response frequency among Republicans who view their party as *Too Extreme.*

Conclusion on In-Survey Nonresponse

Although our in-survey analysis showed a fair amount of instability in the results regarding response frequency, we can come up with a general framework for the characteristics higher likelihood respondents. Some demographic factors for which we currently control in our surveys were strong indicators of nonresponse. For example, we found that college-educated men consistently demonstrated significantly higher response frequencies than women and non-college educated survey participants. We also found some correlations with age, whereby older respondents on occasion appear to have higher response propensities; this aligns with the notion that retired folks and those with more leisure time are all higher-propensity responders.

We found that loneliness is related to response frequency with a certain degree of instability across populations. However, we found incongruous relationships between covariates and subsequent relationship with the response frequency predictor which point to a potentially existent but weak relationship between folks' willingness to respond to SMS surveys and personality and lifestyle attributes.

Higher likelihood respondents' concern with online privacy is also unstable across populations, and contradictory to what common sense would dictate. We also found that agreeable respondents are less likely to participate in our surveys. Further research into how we signal the political nature of the content of the survey may provide a path forward on alleviating this bias.

The only significant relationship between occupational characteristics and survey response we found was that of free time and retirement. We were able to identify some relationships between income, occupational physical requirements, and time spent at a computer that seem to indicate an approach to divide blue- and white-collar workers, but no measure had any consistently significant effect on response frequency.

Finally, and most importantly, we find that the in-survey and out-of-survey analysis point toward a similar conclusion: that the most important indicator of survey response is past survey response. Participation in surveys identifies a strong and unique quality in respondents that other personality and lifestyle attributes can not explain with the same certainty.

Survey Outcome Bias

The previous sections show two important things. First, the extremely skewed distribution of response interest in the population implies significant downside risk of nonresponse bias, since even small correlations between response interest and outcomes after accounting for standard weighting variables can . A skewed distribution of response interest is not a guarantee or requirement of bias, but the assumption that a poll is free of nonresponse bias becomes less and less tenable as this skew increases.

Second, it is very difficult to come up with theories about the drivers of nonresponse beyond the usual demographic factors. Work characteristics, personality traits, convenience, and so on do not appear to give us much of a story to tell about why nonresponse happens.

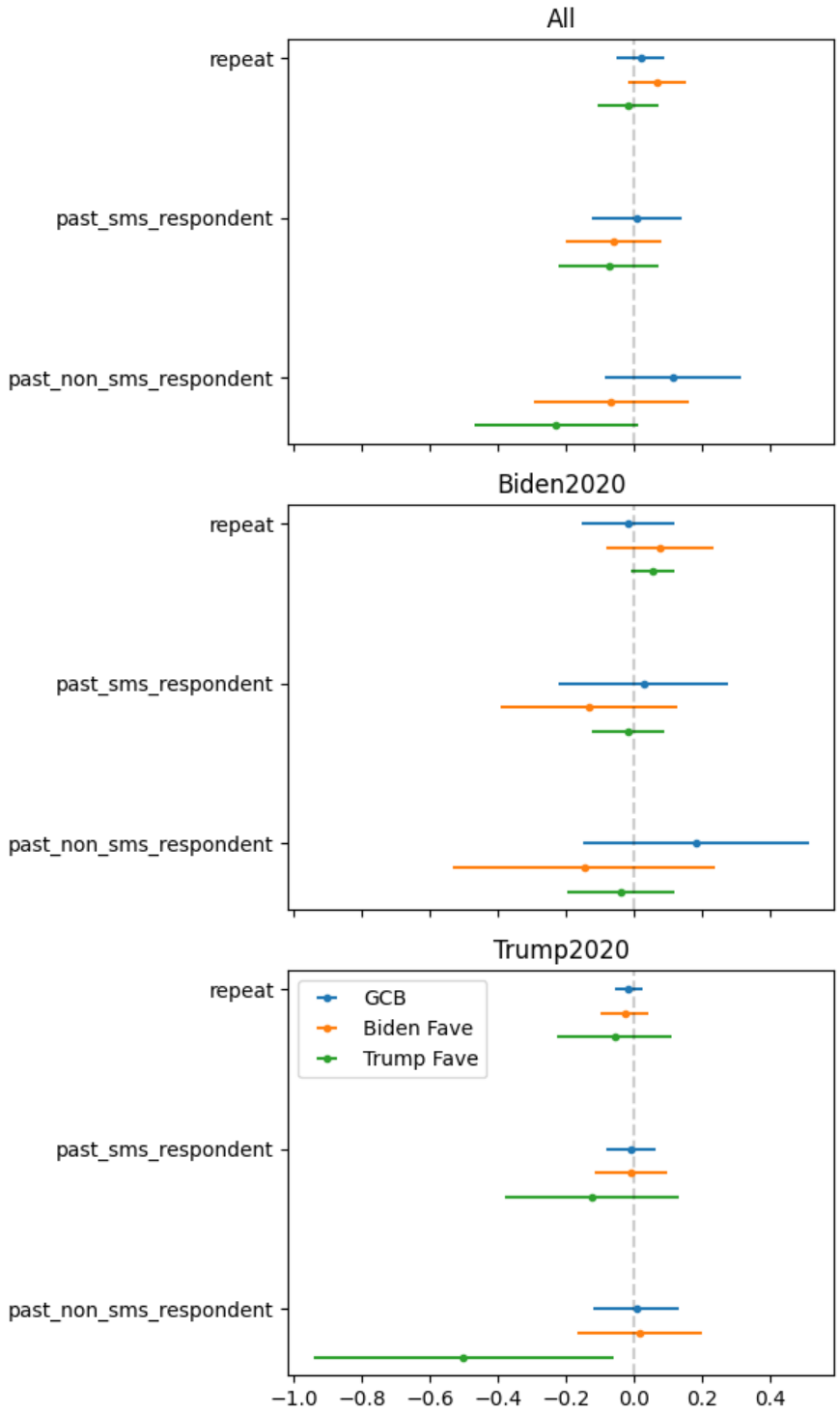
However, this does not limit our ability to detect nonresponse nor to examine its potential role in biasing outcomes of interest. We know response is rare, and repeat response even more so. Both are predominately associated with past response. We also know that response is sensitive to changes in the data collection process, whether those are large changes such as comparing completely different modes, or if they are more marginal changes such as varying contact effort within a single mode. These facts give us a toolkit for coping with nonresponse even though we don’t have much ability to explain it.

The particular outcomes we used in this study are Generic Congressional Ballot, Trump favorability, and Biden favorability. For each of these we run an OLS regression with a binarized outcome. For GCB, 1 indicates a preference for Democrats, and 0 a don’t know or Republican preference. For Biden and Trump favorability, 1 indicates somewhat or strongly favorable.

We include all of our typical weighting variables such as age, race, gender, education, and 2020 recall as controls. We consider two approaches for examining the role of nonresponse bias beyond these factors. First, we include a binary flag for repeat survey taking, meaning a respondent who has taken the current survey multiple times. Additionally, we include binary flags for past response, broken out by mode of response. Past non-SMS respondents are overwhelmingly web respondents.

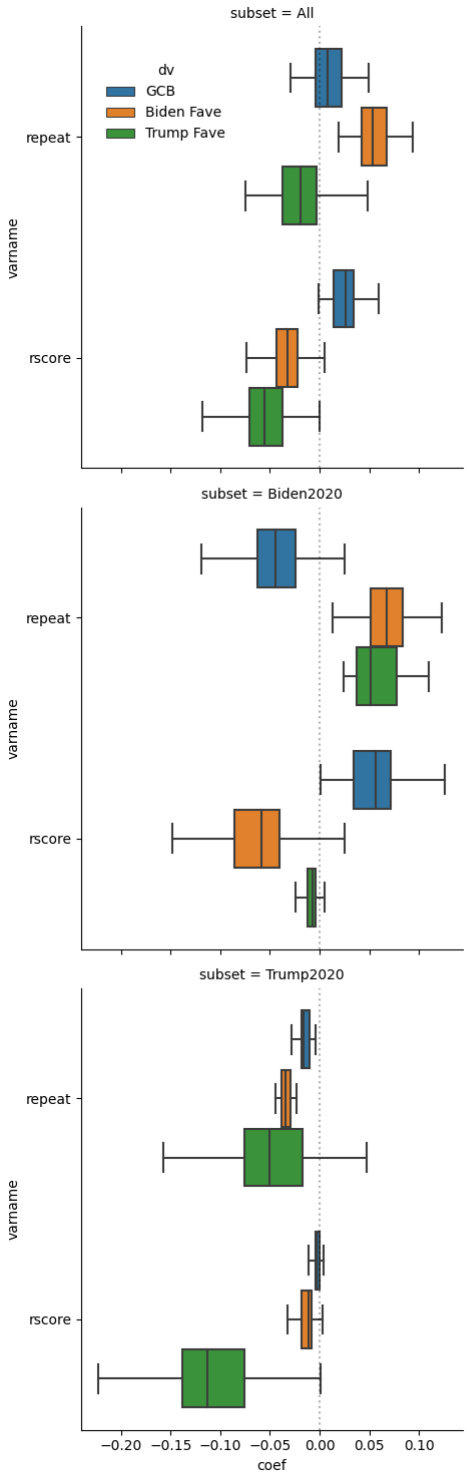
It appears that there are numerous potentially credible interactions between response proxies and standard weighting variables in these outcome models, and so we show three versions of each model. The first contains all respondents, and then two more are subset to Biden 2020 voters and

Trump 2020 voters. The results are shown below with coefficients for standard weighting variables excluded to reduce clutter.



Effects are not especially well estimated here, but point to complex patterns between non-ignorable response and these outcomes. A striking result is that past non-SMS respondents, who are primarily web respondents, are more anti-Trump than we would predict given base weighting variables. This is especially true among Trump 2020 voters. We also see that these web respondents have a higher preference for Democrats in the GCB than expected. Anti-Trump and pro-Democrat bias are both trends we have independently identified in our retrospective analysis of web panel polling from the 2020-2022 cycles.

A second approach is to run the same OLS regressions, but instead of including response history binaries, we use the response score estimated from probit regression using the instrumental variable approach with the bootstrap. Below we have distributions of the outcome regression for 100 bootstrap realizations.

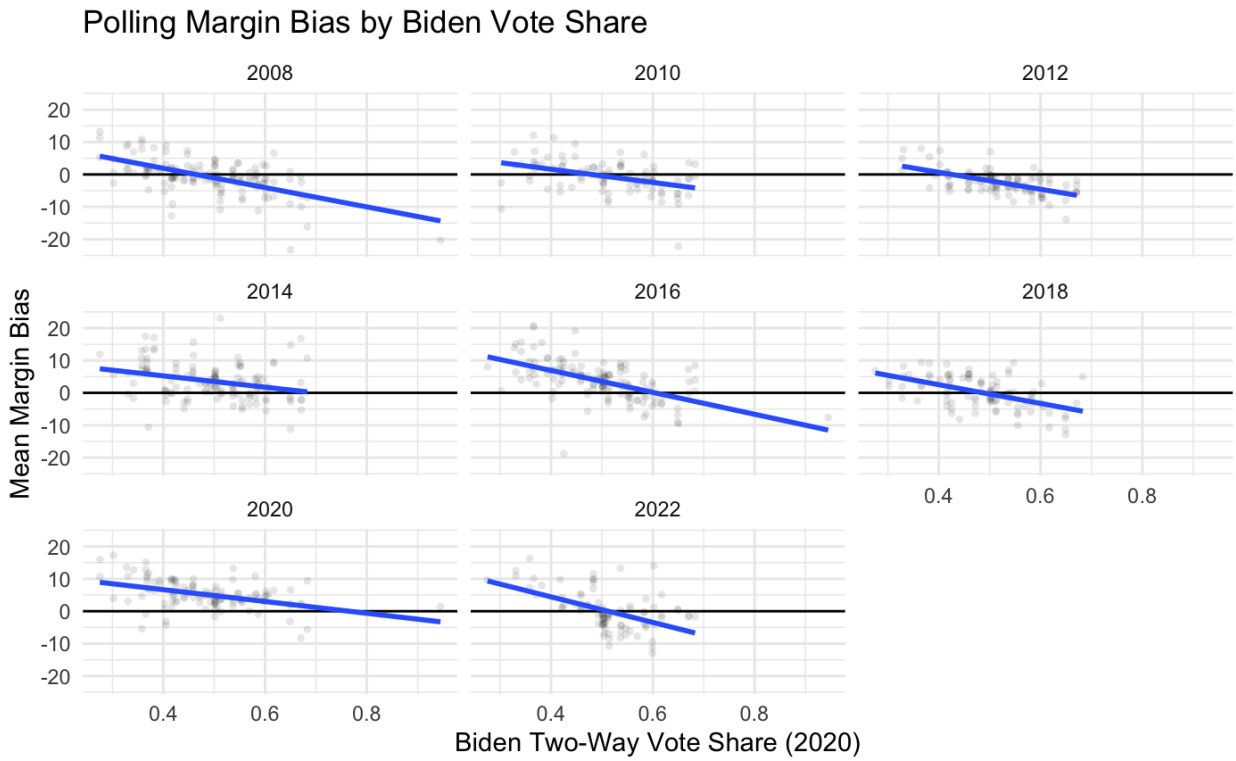


This approach leveraging the IV appears to give us a little more clarity, though again uncertainty here is quite high. Among all voters, higher response score is associated with higher support for Democrats in the GCB, but lower favorability for *both* Trump and Biden.

Among Biden 2020 voters this same overall pattern appears to hold, while for Trump 2020 voters, higher response score is associated with lower Democratic support in GCB but lower favorability for both Biden and Trump. This highlights another very striking and, to our knowledge, novel result about nonresponse. In contrast to more classical theories of partisan nonresponse, which can be summarized roughly as saying that nonresponse is mostly a snapshot of which side's partisans are cheering the loudest at a given moment, here we find evidence to suggest that higher response is associated with dissatisfaction of the leaders of each party. Among both Trump 2020 and Biden 2020 voters, higher response propensity is associated with stronger rates of copartisan voting in the generic ballot, but lower rates of support for Biden and Trump themselves.

This suggests that partisan nonresponse is the net result of both especially enthusiastic partisans that continue to support their party and especially dissatisfied partisans on either side. This paints a much more complex picture than the traditional interpretation that intense survey takers were necessarily enthusiastic hard partisans or activists.

Perhaps the only regular pattern we find in historical polling error across years, firms, and modes is that generally, polls taken in more conservative areas tend to overstate Democrats' chances, while polls taken in more liberal areas tend to overstate Republicans' chances. Below we show this by comparing polling error from public polls vs. the Biden 2020 share of the geography where the poll was taken, which we use as a proxy for baseline partisanship.



There is significant noise here, which limits the predictive value of the trend, but what is interesting is that the trend consistently slopes down. If dissatisfied partisans are overrepresented in polls, this is

the sort of trend we would expect to see in error. Dissatisfied Republicans in GOP strongholds would inflate Democratic margins in polls taken there and vice versa. Our experience has shown that it is difficult to find broadly applicable theories about the causes and effects of nonresponse, so we should note a bit of caution here. Also in this study we found copartisan dissatisfaction with Biden and Trump specifically associated with nonignorable nonresponse, as opposed to dissatisfaction with the two parties more generally, but this does seem like an avenue worth pursuing further.

Regardless of whether or not we can explain response, we can still adjust for it with this framework. At a high level, response score could be incorporated into a principled statistical adjustment by using it as a prior weight for raking, or as a sample weight in an outcome model-based approach. In our outcome regressions here, we see that inclusion of response score does not drive the coefficient for repeat respondents to null, as we would expect to be the case when we have fully captured the relationships between response and outcome. This could be explained by three non-mutually exclusive explanations, which which we are exploring further.

1. The response model may be misspecified and biased due to missing variables. Since we know response history is very important and our coverage for this variable is extremely limited, it is likely this is playing some role. We also use a probit model for response in line with the classical Heckman approach, but this assumes a normal distribution of errors, while in reality our error distribution may have heavy tails. Robust alternatives are worth exploring, at the very least to determine whether or not nonresponse analysis is sensitive to these distributional assumptions.
2. The outcome model may be misspecified due to potential nonlinear relationship between response and outcome and/or interaction terms with our standard weighting variables. This also is plausible given our current analysis as well as our past analysis of nonresponse in web panels, where complex and heterogeneous relationships between response and outcome seem like the norm.
3. The repeat coefficients may not be meaningful in a practical sense, since they represent such a far extreme of a skewed distribution of response that it might be better to simply drop them than to pay them much mind.

These are three items we are currently working to get more clarity on, since this will shape the details of how this method is used to apply actual statistical adjustments.

These results point to the fact that we can indeed do better than what we are doing now to identify and potentially adjust for nonresponse, but there are also some hard limitations. High opt-out rates point to the existence of a large group of never-responders, whose opinions cannot be measured with SMS surveys. This highlights the fact that while we may be able to use repeating and IV techniques to quantify variation in response interest and detect correlations with outcomes *among respondents*, we can't guarantee that this will allow us to properly extrapolate to the target population. An example of such an extrapolation failure can be found on page 66 of these slides. Nonetheless, this analysis shows that the standard ignorability assumptions are violated, and we should try as best we can to fix that, even if we can't get any guarantees.

Conclusions

We find that overall, unobservable and highly transient factors appear to be the primary driver of survey response, with past response being the most influential predictor of response and the only thing that meaningfully predicts repeat responses. Political engagement factors offer clear stratas of response propensity among sampled registered voters and point to recent political activity as a strong indicator for current survey response.

While the in-survey nonresponse analysis does not yield especially strong signals, we can continue to work on a general narrative of nonresponse at the margins. With the pervasiveness of nonresponse and large opt-out rates signaling large portions of our sampling pool as never-responders, we know that these characteristics can only point to minimal changes in response. For example, while we understand that some measure of time availability is a factor for survey response, the shifts in availability required to move response rates currently sitting at around 1% would require more hours in the day for most registered voters. Ultimately response rates are currently too low for any of these marginal adjustments to make a meaningful difference.

The survey outcome analysis provides some novel findings whereby unobservable drivers of response have non-negligible but highly complex relationships with GCB, Biden favorability, and Trump favorability, even when controlling for standard weighting items. There is evidence to suggest that both hard partisans and dissatisfied Trump and Biden 2020 voters are overrepresented in SMS surveys. This theory can potentially explain some of the trends we have seen in the directionality of polling error, with polling tending to overestimate Democratic support in conservative areas, and vice versa.

While the concept of enthusiastic voter bases for each party being overrepresented in polling aligns with classical polling theories, the dissatisfaction component of the findings suggests a novel signal. It is possible that these two theories coexist, and that the rates at which these two groups overrespond may shift over time as the election nears. We know that temporality in survey response exists, and there is reason to believe that temporal trends would not be the same for these two factions.

Next Steps

The findings from this experiment offer several avenues for further testing and a roadmap for adjusting some of the tendencies we found in nonrespondents.

The engagement factors identified split our sample pool into strata that represent different response propensities. Probability sampling that accounts for engagement strata may facilitate oversampling unlikely responders.

In addition, we saw new respondents crop up at every round and found that in later rounds the observable factors associated with survey completion propensity became less definitive in predicting response. This suggests that we may be able to mitigate some of the observable factors related to nonresponse by incorporating a degree of recontacting as a part of our sampling practices.

The temporality of survey response habits suggests room for further study on the life cycle of survey response and opting out, or becoming a never-responder. Our findings regarding the relationship between survey response and party dissatisfaction may shape further research into how to track and adjust for shifts in approval from hard partisans.

We also noted a difference in the nature between static or consistent voting habits, and the irregularity of survey response. A deeper dive into the relationship between historical voting habits, modeled turnout scores, and survey response engagement should help us clarify the how SMS respondents compare in low- and high-turnout environments such as in primary versus general elections.

Appendix

Appendix

Sample Summary

The following tables contain the distributions of the stratification variables for the final sample for the SMS nonresponse experiment. We see a sufficient representation in the sample from the necessary subgroups.

Partisanship Score Quantile	Count	Proportion
0-20 (Furthest to the right)	23,761	24.2%
21-40	10,047	10.2%
41-60	9,477	9.7%
61-80	10,584	10.8%
81-100 (Furthest to the left)	43,382	44.2%
Unknown	925	0.9%
Total	**98,176**	**100%**

Gender	Count	Proportion
Male	44,372	45.2%
Female	50,564	51.5%
Unknown	3,240	3.3%
Total	**98,176**	**100%**

Zip Code 2020 Biden Friendliness Quantile	Count	Proportion
0-20 (Vote share least supportive of Biden)	14,3098	14.6%
21-40	13,037	13.3%
41-60	10,692	10.9%
61-80	12,817	13.1%
81-100 (Vote share most supportive of Biden)	22,578	23.0%
Unknown	24,743	25.2%
Total	**98,176**	**100%**

Age	Count	Proportion
18-29	9,912	10.1%
30-39	18,362	18.7%
40-49	17,471	17.8%
50-59	16,683	17.0%
60-69	16,500	16.8%
70+	17,087	17.4%
Unknown	2,161	2.2%
Total	**98,176**	**100%**

Race	Count	Proportion
Black	13,776	14.0%
Hispanic	11,958	12.2%
White	68,624	69.9%
Other-Unkown	3,818	69.9%0.4%
Total	**98,176**	**100%**

Urbanicity	Count	Proportion
Rural	27,611	28.1%
Suburban	35,451	36.1%
Urban	35,114	35.8%
Total	**98,176**	**100%**

Message Schedule and Content

Message Number	Date Sent	Message Content
1	June 14, 2023	Hi [first_name] We're reaching out to see how you feel about the issues that affect you. Please take our short survey and make sure we hear from you! survey link Text STOP to quit
2	June 17, 2023 **(3-day gap)**	Hello again! We're conducting a short survey to understand how people feel about the issues affecting them, and your opinion matters. Even if you recently took a survey with us, we still want your feedback to see how views change over time. Click here to participate: survey link Text STOP to quit
3	June 20, 2023 **(3-day gap)**	Hello again! We are continuing to send out our short survey to understand how people feel about the issues affecting them, and your opinion matters.

Message Number	Date Sent	Message Content
		Even if you just completed the survey, we are still very interested in understanding how your views change over time. Click here to participate: survey link Text STOP to quit
4	June 27, 2023 **(7-day gap)**	Hello again! We are continuing to send out our short survey and we would really appreciate your feedback to understand how views change over time. You can find the link to participate here: survey link Text STOP to quit
5	July 18, 2023 **(21-day gap)**	Hello! This is the last time we'll be reaching out with our short survey! We'd like to hear about how people feel about their work, lives, and the issues affecting them, and how those views change over time. Click here for your final chance to participate: survey link Text STOP to quit

Covariate Relationships

Age Correlations

We found that respondents that are retired are significantly more likely to respond at higher rates in the experiment. However, retirement is naturally correlated with the amount of time dedicated to leisure. The results from our regression analysis show that those that are 70+ in age respond at higher frequencies, which is directly related to retirement status as well.

Loneliness Correlations

While we find no relationship between retirement or age and loneliness, we do find a correlation between loneliness and how respondents compare themselves to their neighbors ideologically.

We also tested to see if the retirement characteristic was correlated with loneliness by applying a t-test to test the null hypothesis that there is no difference in the loneliness scores between retired and non-retired people.

t	df	p-value	CI
2.4864	526.56	0.01321	(0.02046138,0.17448750)

We cannot reject the hypothesis, and conclude that there is no significant difference in the loneliness scores between retired and non-retired respondents.

	Not-Retired	Retired
Av. Loneliness Score	1.539	1.44

We find a significant relationship between loneliness and perception of how their ideology compares with that of their neighbors. Folks who respond that they don't know about how their neighbors' political ideology compares with their own had higher loneliness scores on average.

Perception of Neighbors' Ideology	Av. Loneliness Score
More conservative than neighbors	1.42

Perception of Neighbors' Ideology	Av. Loneliness Score
Similar ideology as neighbors	1.43
More liberal than neighbors	1.57
Don't know	1.61

We confirm this using an ANOVA to look at the relationship between the two variables and find statistical significance.

	Df	Sum of Squares	Mean Square	F Value	P-value
Perceptions of Neighbors Ideology	1	3.96	3.962	12.76	0.000372***

To reduce any multicollinearity concerns in our models, we included only the loneliness scores in our analysis.

Occupational Correlations

A similar story occurs in the variety of questions related to occupational characteristics and socioeconomic status. We found that variables related to the occupational strength requirements, income, and time spent on a computer throughout the day were all highly correlated among each other.

Occupational Strength Requirement	Low Income (Under $25k - $50k)	Middle Income ($50k-$100k)	High Income ($100k-$150k+)
Sedentary	38 (36.2%)	110 (57.0%)	172 (64.9%)
Light	33 (31.4%)	40 (20.7%)	62 (23.4%)
Medium	23 (21.9%)	29 (15.0%)	21 (7.9%)
Heavy	10 (9.5%)	12 (6.2%)	4 (1.5%)
Very Heavy	1 (1.0%)	2 (1.0%)	6 (2.3%)
Total	105 (100%)	193 (100%)	265 (100%)

With the contingency table we can begin to see a trend where a as income increases, stated physical labor intensity of occupation decreases. We test this theory with a chi-square test and find that effectively there is a significant relationship between income and occupational strength requirements.

$$\mathrm{X}^2(8, N = 563) = 40.004, p = 3.198e-06.$$

Similarly, we use a one-way ANOVA to compare the effect of occupational strength requirements on time spent on a computer throughout the day and find that the two are significantly related.

	Df	Sum of Squares	Mean Square	F Value	P-value

Occupational Strength Requirement	4	549	137.32	7.413	8.03e-06***

To avoid multicollinearity in our analysis we therefore use only one of the three variables.

Time Spent on Leisure	Not Retired	Retired	Share Retired
Less than 20 Hours Weekly	537	162	23.2%
Between 20-40 Hours Weekly	120	86	41.7%
More than 40 Hours Weekly	20	24	54.5%

When we apply a chi-test we find a significant relationship between retirement and time spent on leisure.

$$\mathrm{X}^2(2, N = 949) = 41.957, p = 7.746e - 10.$$