

# PredictWise

## Methods for PredictWise Audiences

We gather samples of respondents using [Random Device Engagement](#) (RDE). RDE means targeting Ad IDs, unique device identifiers of mobile devices in-app, randomly on the platform unique users behind the Ad ID spend their time. That is to say, we reach potential respondents where they spend time organically: when they engage in their daily tasks at home, consume information, and interact with friends and family. Coverage is reasonably high, insofar as smart phones (as a key device type today) have very high penetration (~ 70%), reaching about 7 Million unique Ad IDs in the US, and quality of data is good because surveys and incentive to participate are embedded into the feel and design of the application where we pick up respondents. Response rates are much higher than the most comparable method, Random Digit Dialing (RDD), which is extremely similar, but tied to calling people on their phone. We collect about 5,000 survey respondents every week, and have access to large-scale ambient data: rich and accurate (cell-tower-based) latitude/longitude coordinates and application usage data for many Million Americans through existing relationships and data-share agreements.

We then create estimates of public opinion, capturing 250+ attitudes exhaustively mapping the US political space, using the most bleeding edge analytics of modeling and post-stratification (affectionately known as MRP+) - a technique that allows us to distill disparate streams of non-representative data into representative and accurate estimates of public sentiment. Compared to other market intelligence companies, our approach is differentiated---and superior---in three important ways:

1. **Depth:** We can present estimates with unprecedented demographic and geographic granularity, e.g. White college-educated females in WI, or even down to the individual
2. **Speed:** We can integrate opinions on additional dimensions within hours
3. **Accurate:** It is also accurate. Our state-of-the-art method uses AI to convert the mix of polling and ambient data we collect into validated results (more below)

## PredictWise AI models

First, we model the raw responses to each question, given each respondent's age, gender, location, education level, race, marital status, party identification, income, family size, urbanicity, application usage behavior, neighborhood composition, and work environment via a Bayesian latent trait model. This information divides the population into Millions of demographic categories.

Each of these predictions is informed by all responses, including responses received in previous surveys. To achieve this, we have developed a complex dynamic model that allows us to parse out variance in sample composition from true swings over time.

Second, we project the model estimates onto the target universe, a process known as post-stratification. We have curated the target universe over years, centering on voter-file data (that we receive from TargetSmart via a data share agreement), and enriched with consumer- and application usage data taken from ambient data collected over years – this is a 250 Million-rows data-set we update regularly. This allows us to track all US adults on 250+ attitudes, week-in week-out. For a full white paper, click [here](#).

# PredictWise

## Validation: sampling of media usage of our polls

- *The New York Times*: See [here](#) (Correct 2016 Florida prediction) and [here](#)
- *Slate*: See [here](#) and [here](#)
- *Washington Post*: see [here](#) (Showing why Trump could win primary), [here](#), [here](#), [here](#), [here](#), [here](#), [here](#), [here](#)
- *Buzzfeed*: See [here](#), [here](#), [here](#), [here](#)

**Validation:** Peer-reviewed articles published in leading academic journals: See [here](#), [here](#), [here](#)

## Accuracy

- **2016 Presidential Election:** Correctly predicted 46 states and winner ([released publicly during the election cycle](#))

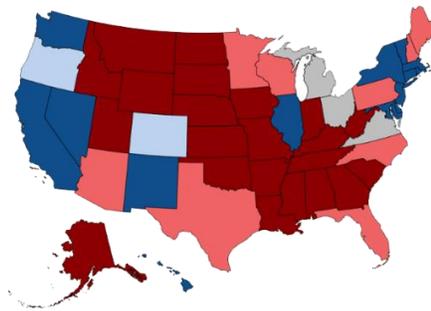


Figure 1: PredictWise electoral map from November 7, 2016, correctly predicting winner

- **2017/2018 Special Elections:** We have run and released polling for four of the special Congressional elections in 2017-2018. ***On average: our error is below one percentage point.***
  - 1) Republican candidate winning TX-27 by a margin of [34 percentage points](#): he won by 36 percentage points (i.e., we were off by 2.0 percentage points).
  - 2) Republican candidate winning Arizona's 8th Congressional District by [6 percentage points](#): she won by 5.2 percentage points (i.e., we were off by 0.8 percentage points).
  - 3) Democratic candidate winning Pennsylvania's 18th Congressional District by [1 percentage point](#): he won by 0.4 percentage points (i.e., we were off by 0.6 percentage points).
  - 4) We had an early big lead for the Republican candidate, but our same-day polling on Election Day showed it shrank to [5.5 percentage points after he assaulted a reporter](#): he won by 5.6 percentage points (i.e., were off by 0.1 percentage points).

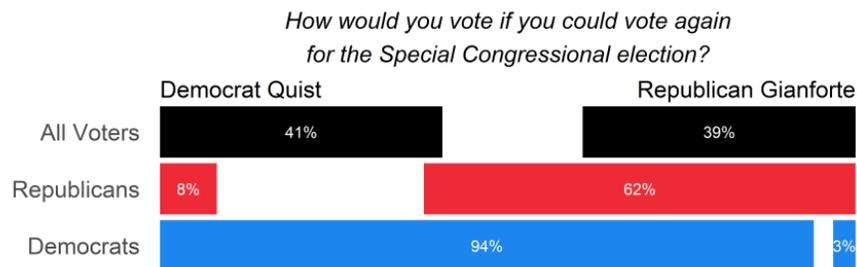


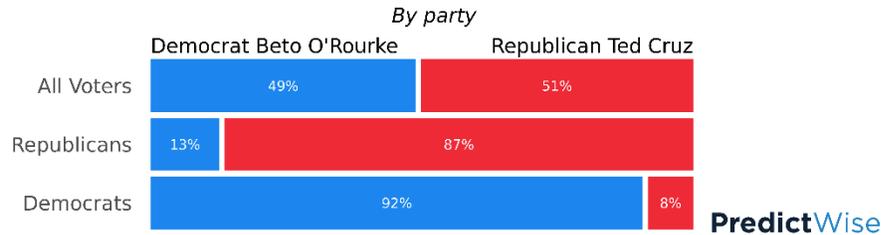
Figure 2: PredictWise election prediction for MT-AL special election, 05/25/2017

# PredictWise

- 2018 Midterm Elections:** We ran and released 3 pre-registered polls run through a novel, innovative set of analytics we dubbed [Calibrated Polling](#). On average, our error was below 2 percentage points.

1) **TX:** We had Beto loose by 2 ppt. And, Beto lost by 2 ppt. exactly

## TX: 11/04-11/05, Party, N=1,400



2) **NV:** We had Rosen at 51.3% and Heller at 48.7% The end result was 52.6% (Rosen) 47.4% (Heller)

## NV: 11/01-11/05, Party

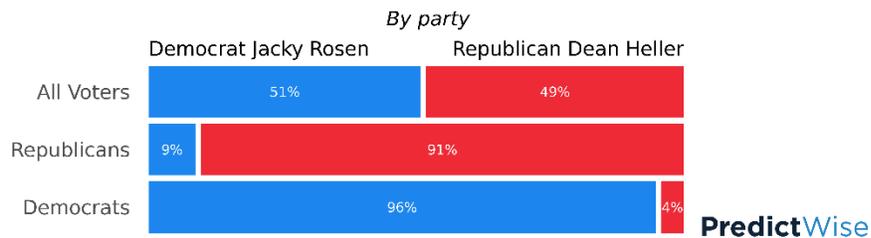
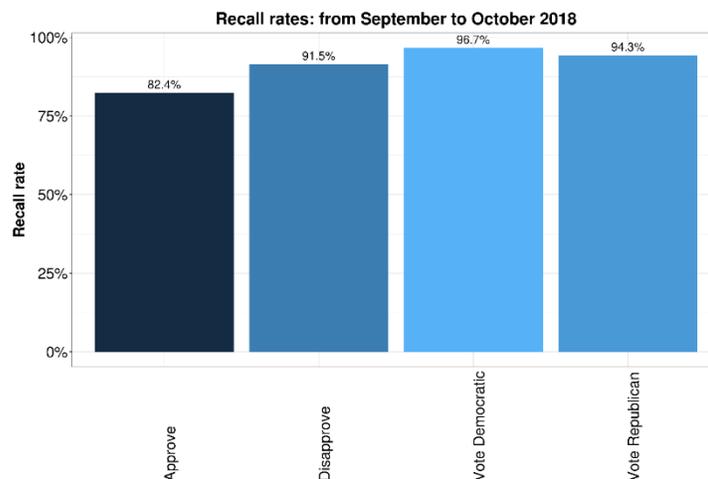


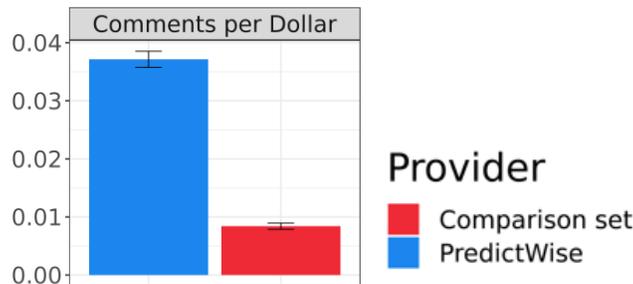
Figure 3: PredictWise calibrated polling in 2018

- Individual-level Data:** As noted, we use our AI to “score” all Americans on 250+ attitudes. The metric of quality of PredictWise Audiences is recall, i.e. how well PredictWise models perform when tested against ground truth. We can test this quality directly whenever we survey individuals for whom we have modeled scores, but who have not shown up to our polling data base before. Below are metrics from our recall test conducted in late 2018, comparing predicted outcomes based on models against actual data on approval (N=6,178, three categories, collapsed), and predicted outcomes based on models against actual data on vote choice (N=4,556; three categories). Our average recall is >91%.

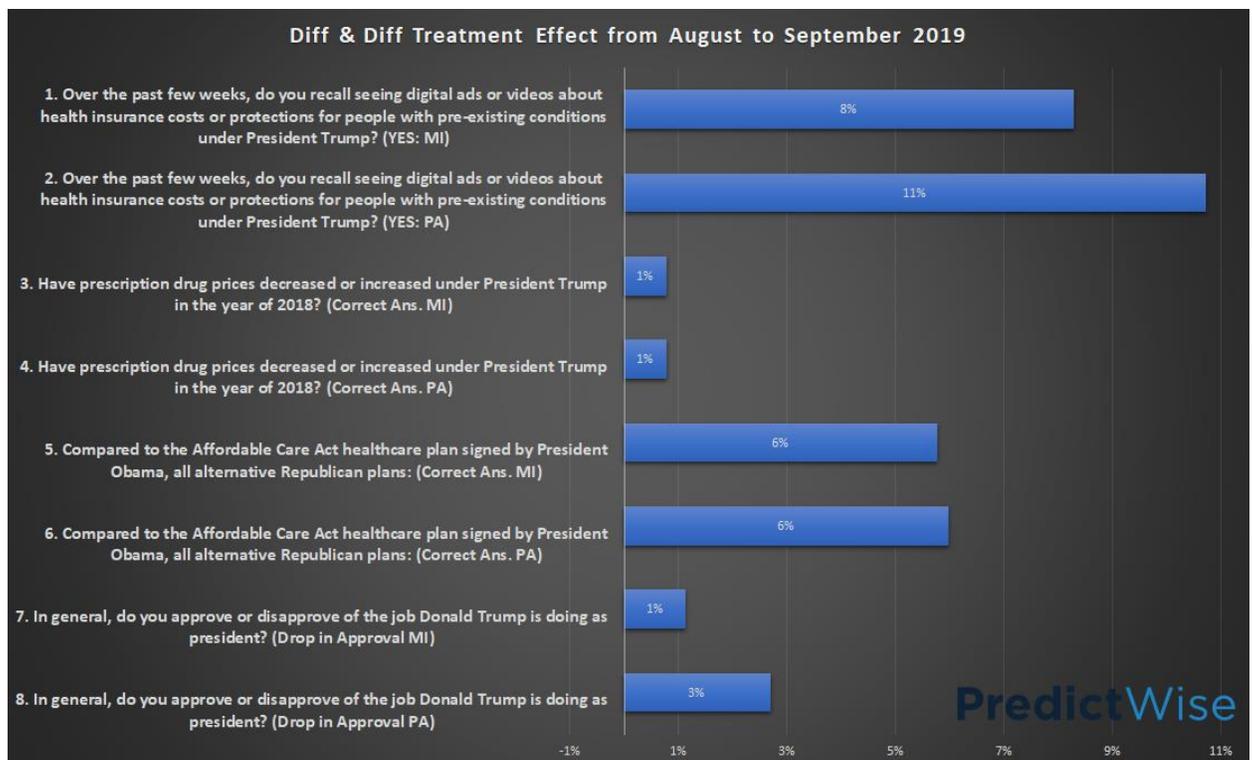


# PredictWise

- The CA-45 (Katie Porter) campaign needed a list of most persuadable voters to power its digital ad-buys on Facebook. A rival segmentation was built on custom data, with 10x our raw-cost. PredictWise, instead, created custom segmentations for each creative matching the primary message content of the creative in questions to likely Republican voters inhibiting a progressive position on that dimension, based on the PredictWise back-end data tracking 250+ attitudes for all Americans (no marginal cost), delivering a list of >100K individual targets. Our segmentations beat the rival segmentation handily, garnering 4x comments on Facebook.



- Together with ACRONYM, we targeted 100,000 Mobile Ad IDs (MAIDS) of persuadable Republicans on healthcare in Michigan and Pennsylvania with healthcare creatives. In a before-after design with control group (Randomized Control Trial), our audience technology coupled with ACRONYM spend resulted in massive increases of knowledge, and lowered Trump approval of our treatment universe by 3 ppt. in Pennsylvania.



: