Pollfish

FUTURE OF SURVEYS:

Random Device Engagement (RDE) with Organic Samples

Dr. David Rothschild
Economist at Microsoft Research

Dr. Tobias Konitzer
C.S.O. and co-founder of PredictWise

August 17, 2018

CONTENTS:

1. Introduction

2015 and 2016 saw high-profile polling failures throughout the world.

In the summer of 2015, before Brexit and the 2016 US election, The New York Times asked, somewhat rhetorically:

What is the matter with political polling? Implying that there was already a crisis of confidence in polling. Then in 2016, the United Kingdom stunned the world by voting in favor of Brexit, a referendum on the United Kingdom leaving the European Union, despite opinion polls shifting towards remain in the last few days.[1] A few months later, despite polling showing Democratic candidate Hillary Clinton winning in enough states to win the US election, and poll aggregators confidently pointing to a Clinton victory, Republican candidate Donald Trump won a fairly comfortable Electoral College victory (but, still lost the popular vote).[2] While there is some nuance to the label of failure, the popular vote was forecast spectacularly well by polling aggregators, and "failure" was really a local phenomenon boiling down to a number of state-level polls in the Rust Belt (and applied to the presidential election only, and not congressional elections): the public perception was that of "failure in broad and absolute terms.[3] As is now well known, this failure (or at least either perception of failure or partial failure) led to a reckoning with the status-quo modus operandi of polling; the whole industry faced a market-threatening question of where they were going.

[1.] https://www.huffingtonpost.com/entry/brexit-polls-missed_us_ 576cb63fe4b017b379f58610
[2] https://www.quantamagazine.org/why-nate-silver-sam-wang-and-everyone-else-were-wrong-20161109
[3] https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx

Culprits were readily identified and one target was Random Digit Dialing (RDD) polling samples, the gold standard of high-quality polling in recent decades, which has undergone a massive shift in recent years. RDD response rates have decreased from 36% in 1997 to single digits in the 2010s. And, as Gelman et al. (2016) shows this non-response is coupled to political attitudes: today, traditional polls, RDD with a mix of landlines and cellphones, have a hard time reaching those with lower levels of education and lower levels of political knowledge. Thus, polls in 2016, especially the crucial state-level polls in the contiguous states of the Rust Belt, that neglected to weight on education had a huge problem. Similarly, RDD has a hard time reaching White blue-collar voters, dubbed Bowling Alone Voters, especially mobile blue-collar voters ("Truck Driver phenomenon), as a Post Mortem by Civis Analytics has pointed out.[4] This is even harder to control with traditional analytics.

But, even more serious than its current problems: even if RDD can still work, it is doomed in next few years. Do you have a landline? Do you answer unknown (or suppressed) numbers on your cell phone? Will you have a cell phone in 10 years?
Will the platform for reaching you be a phone number or a user ID? These are serious questions that further jeopardize the future of random digit dialing: by definition it is impossible without phones!

[4.] https://www.wired.com/2016/11/pollsters-missed-bowling-alone-voters-handed-trump-presidency/

As with all discussions around polling, it is critical to delineate two distinct things: data (or sample) collection and analytics. Data collection is how respondents are gathered. Data analytics is how the collected data is turned into market intelligence. Nothing prevents the most advanced analytics from being used on any data collection, although different analytics will provide various levels of benefit to different samples. For this paper we will stick to data collection, but refer to several previous papers exploring data analytics (Goel, Obeng and Rothschild 2015).

As is the case with all innovation, some innovation is good and scientifically sound, some innovation is snake oil, with little or no effect, and some innovation is flat-out dangerous. In this paper we shed light on three such innovations competing to replace RDD: Online (non-)probability panels, Assisted Crowdsouring, and Random Device Engagement (RDE). All innovations come with strengths and weaknesses. But, as we spell out here, one is the clear winner: RDE, which is why RDE is at the core of the PredictWise methodology.

2. Sample Methods

2.1 Random Digit Dialing (RDD)

Random digit dialing is exactly as the name says: building a sample calling random people on the phone.

The first step is to identify a cluster of phone numbers that have reasonable demographic and geographic representation. Then, start calling those numbers at random, trigger a response, and collect poll answers over the phone. The mode is confined, by definition, to a telephone, but it has recently expanded to both landline and cell phones. The mode has high coverage (in that most people have either or both a landline and cell phone), but coverage becomes harder to assess while landline penetration is dropping as cell phone penetration is rising. This makes it hard for survey researchers to map the population in either group or any individuals inclusion in either group. Response rates are oftentimes in the single digits.

2.2 Online Panels

Online panels collect responses either via a fully opt-in structure, including a signup page, or start with an RDD-telephone (and/or supplemented with cell phone) or mail recruitment.

Panelists are then recruited to participate in specific surveys, for example via email invitation to the page of the panel provider. The mode is a mix of desktop, tablet, and smartphones, depending on the device of choice from which the invitation is opened. The mode has very low coverage (very few people opt-in to panels), but RDD-based panels, which start out with random methods of recruitment, have better coverage. Response rates, although generally decent from panelists, are low when one considers the low degree of opt-in to the panel. This makes them hard to compute accurately.

This has a number of advantages:

1. **Panels provide repeated and connected users:**
   Over-time trends can be analyzed, and any custom polling built on top of baseline tracking can be guided by priors derived from data a serious innovation.

2. **Relatively cheap and fast:**
   Marginal polling is relatively inexpensive and can be done faster than traditional random digit dialing.

Curating panels comes with a number of serious disadvantages:

1. **Locked into one model of data collection:**
   Polling firms that are locked into a specific mode of data collection will be hit with tremendous costs because the old infrastructure will have to be dismantled as technology shifts over time. And, no one can predict how long online panels will be a viable mode of data collection as web usages shifts to mobile and beyond (yes, you are reading this right: we want you to think virtual reality here). And, many companies that build their polling around this form of panel are locked into non-transferable unique identifiers of each respondent. This has some short-term benefits, but it will make it very costly when the companies need to shift data collection as technology evolves.

2. **Panel fatigue:**
   A myriad of research has documented that repeated participation in polls of panelists can lead to panel fatigue, resulting in non-response error or measurement error (Porter, Whitcomb and Weitzer 2004; Kasprzyk 2005). The applied scenario: respondents might be eager to fill out surveys correctly and with care, but this willingness declines the more respondents are invited to participate in surveys, especially if respondents are at risk to lose panel status. Instead of providing meaningful answers, respondents then click random answer options, or gravitate toward Dont know.

3. Panel effects/Panel conditioning:

Slightly different from panel fatigue are panel effects, or panel conditioning (Sturgis, Allum and Brunton-Smith 2009; Halpern-Manners, Warren and Torche 2017). Even if panels recruit a sample that looks like the perfect cross-section of the desired target population at the time of recruitment, the demand to answer political surveys turns these initially representative panelists into a bunch of very politically aware citizens. Panel conditioning has plagued a number of panels or panel-like setups. In the worst case scenario, all panelists will have acquired a base degree of political sophistication as a consequence of being professional political survey takers. In that case, even the most advanced bias correction algorithms will fail because of sharp separation: Among the panelists, no one (read: zero) who mimics the stratum with low levels of political sophistication is left.

4. Mix of web and mobile not clean:

web panels tend to engage respondents either desktop, or on their mobile devices, but the infrastructure may, or may not, be very adaptive. Either way the users are engaging in different experiences conditional on the device of engagement, which is hard to control for.

5. Non-Organic:

In panels, respondents are not engaged in their natural (read: organic) environment (Zaller et al. 1992). Instead, an alternative digital environment is created, with the potential of introducing measurement error. As respondents are taken out of their normal routine, thought processes can deviate from those in more natural environments, leading to artificial considerations that can unduly influence item response.

Online panels have the ability to track public sentiment over time more easily than RDD, and are able to leverage the longitudinal panel structure of the data to parse out true swings from artificial movements. In addition, clients of custom polls can be guided by a plethora of prior baseline data when writing the poll. But, reliance on data collection methods and dangers of panel fatigue and panel conditioning mean that insights can be seriously biased, especially if the panel exists for a longer period of time (and panels, as a class, exist for a longer period of time) and it is getting harder to recruit a fresh replacement sample.

2.3 Assisted Crowdsourcing

Assisted crowdsourcing polling relies on social networks with massive penetration, and data on their users, to supply respondents.

(read Facebook: while it can be done on other display or search ad platforms, the massive penetration/coverage and availability of background demographic data mean that Facebook is really one of the few alternatives).

First, the researcher creates a set of demographic quotas (i.e., the number of respondents they want with any combination of demographics). She then submits these quotas to a social media platform, along with an ad to invite respondents to participate in the survey. The social network then serve this content to a targeted group of users, and the polling firm surveys respondents who click on the ad and go to survey site. The mode is mainly desktop, but could be tablet or mobile as well. This method has very high coverage, but low response rates.

There are some advantages with this sampling method:

1. Speed and Targeting

The main advantage here is that due to the penetration and reach of Facebook, polling can be done at granular areas (think state legislative districts), at a somewhat cheaper cost (by our estimates, respondents will run at about $5). Thus, a polling firm engaging in assisted crowdsourcing could sell a poll of N 1,000 for about $8,000-$10,000, slightly cheaper than traditional polls (but with a similar costs to online panels), and, due to Facebooks reach, faster. In summary, good depth, speed, and relatively good costs.

2. Organic Sample

Facebook is an organic location for getting opinions. Instead of curating professional survey takers who answer many political polls akin to a (side-)job, assisted crowdsourcing reaches respondents where they spend time organically. That is to say, people live on Facebook, get their information on Facebook, share their thoughts on Facebook; assisted crowdsourcing gathers opinions in that natural environment.

There are BIGGER disadvantages:

1. Quota Sampling is bad:

Quota sampling has long been shunned by highquality polls, and for good reasons: The debacle in the 1948 election laid bare the dangers in quota sampling (i.e., Dewey did not beat Truman).[5] If respondents are "recruited to fill demographic buckets, pollsters are going to recruit respondents in that bucket who are easiest to reach. You need to recruit 10 non-college educated Whites? Great, you have interacted with representatives of that demographic bucket in the past, why not simply recruit these folks? While this is done in practice, hitting the same respondents over and over again is problematic. More important, the ability to reach someone within a bucket is likely correlated with the respondents level of political engagement, partisan affiliation, and political knowledge: the same things you are trying to measure. Specifically, respondents of certain demographic strata who are easy to reach have abnormally high levels of political engagement, knowledge, etc., leading to a sizable bias that cannot easily be corrected.

[5] https://en.wikipedia.org/wiki/Dewey_Defeats_Truman

Contact: sales@pollfish.com

2. Quota sampling on social networks is worse:

If you are dealing with social networks, the quota sampling problem discussed above gets much worse. Facebook algorithms are designed to expose the cheapest respondent to the ad, i.e. the respondent that is most likely to maximize click-through rates (see for example this discussion about Facebook targeting algorithms in this recent PNAS letter (Eckles, Gordon and Johnson 2018)). Hence, it makes sense to show ads to participate in a political survey, especially those that have a political cue, to users who are more likely to click on political content for example users who declare a self-reported ideology as part of their profile, or like a lot of political content. If polling firms relying on assisted crowdsourcing target, say, non-college educated Whites, chances are that those non-college educated Whites who are exposed to the ad because of their high likelihood to click on political content exhibit unusually high levels of political engagement. To make matters worse, the characteristics most predictive of that non-representativeness, behavioral metrics from Facebook such as Likes of political content, are not available to polling firms for bias correction. And, in expectation, Facebooks machine learning algorithms get better at predicting who clicks on ads to participate in political polls, and who does not, over time. This means that (a) biases exacerbate the longer polling firms recruit respondents on Facebook, and (b) the number of fresh respondents diminishes, in effect leading to a panel structure bringing with it concerns of measurement error due to panel fatigue, or panel conditioning effects, meaning a change in underlying attitudes as a direct consequence of membership in a panel-like structure.

3. Assisted crowdsourcing is at the mercy of social networks:

Simply, any survey tool on a social network is reliant on the legal framework surrounding social networks with high penetration (and, there are really only two or three to speak of). Much like online panels, assisted crowdsourcing lacks agility with technology and adaptability to new audiences. Should any preemptive legislative strike result in the social networks withdrawal from the political ad market (or a dramatic shift in costs or types of exposure), a possible scenario amidst the recent turmoil surrounding the data breach leveraged by the now defunct right-wing analytics firm Cambridge Analytica, the respondent market and methodology fine-tuned to the idiosyncrasies of respondents drawn from the social network in question, can become obsolete in a matter of minutes.

Polling companies relying on Assisted crowdsourcing have the ability to poll every political race from presidential elections to state legislative elections, and that is commendable. But, biases introduced by quota sampling, exacerbated by fine-tuned targeting algorithms of social networks, mean that severe and uncorrectable sample bias can lead to serious polling error. In addition, the nature and extent of respondent supply is completely dependent on a legal framework polling firms have no influence over.

2.4 Random Device Engagement

Many of the tenants of RDD are commendable: calling respondents in their homes means that respondents are picked up in an organic location for getting opinions. Pollsters reach respondents where they spend time organically.

That is to say, people engage in their quotidian tasks at home, get information at home, and interact with friends and family. In short, RDD gathers opinions in that natural environment. Can we fix what is broken with RDD while maintaining its strengths? Let us introduce Random Device Engagement (RDE); it is the natural successor of RDD, in terms of orthography, philosophy and quality.

Random device engagement (RDE) polling relies on advertising networks, or other portals on devices, to engage random people where they are. One of the most common versions of this is within advertising modules on smart phones, but it can easily be placed in gaming, virtual reality, etc. Respondents are asked to participate in a poll in exchange for an incentive token that stays true to the philosophy of the app in which they are engaged: For example, respondents contacted via the popular mobile gaming App Harry Potter: Hogwarts Mystery can be reimbursed for survey participation with energy points, a crucial currency of the game. Direct monetary incentives are also possible, such as the chance to win an Amazon gift certificate.

The key here is that by being able to monitor the unique identifier of the device world ad IDs survey firms can prevent fraud originating from SUMAs (single users, multiple accounts). And, RDE samples are both random and organic. This is the natural successor to random digit dialing, which aims to randomly engage with landline (and now cell) phones. In many ways, it is just making RDD generic for the future: random, device (rather than phone), engagement (rather than dialing). It addresses RDDs greatest problem: technology is always changing. It solves for this by targeting a respondents unique ID that can be tracked across changing devices, as the future of phones is uncertain. In addition, RDE brings a plethora of telemetry or para data to the table that is amenable to bias correction, from location history to application usage.

This method has a number of advantages:

1. Fast:

RDE can be extremely fast. RDD takes days (and weeks in some cases). Using social networks (assisted crowdsourcing) can be done a little faster, but still lacks speed compared to RDE. Using online panels is comparable in speed, if you pay for extra respondents from a merged panel (online panels will charge extra to get respondents from other panels to increase speed).

2. Cost-effective:

RDE is extremely inexpensive compared with other sampling 12 options. The major RDE providers, like Pollfish, Dalia or Tap Research, charge 10% the cost of RDD, 20% the cost of using assisted crowdsourcing, and 25% the cost of online panels.

3. Coverage is good and growing:

Accuracy is good, because coverage is good. The major RDE providers mentioned easily reach 5,000,000 unique respondents, in the US market alone. And, while RDE is still behind RDD in coverage at this time, it will reach parity soon.[6] Coverage is similar to social media-based assisted crowd source polling, and much better than with online panels. Online panels have a very small footprint, which also affects their ability to get depth in population.

4. Response rate is solid:

Pollfish reports a reasonable response rate (much higher than RDD), conditional on being targeted for a poll (to completion of the survey, that is). Online panels have low sign-up rates and high drop out, but do not post comparable response rates. Social media-based polling, in assisted crowdsourcing, is reliant on ads that suffer from very low click-through.

5. Flexible:

RDE is meant to be flexible with the growth of devices. It should provide a seamless experience across device types. RDD is stuck with telephones, by definition. And, RDD is subject to interviewer effects (albeit to a smaller extent than in-person surveys), meaning that tone of voice can influence considerations of the respondent, or trigger undesired interviewer-respondent interactions, ultimately introducing measurement error. RDE, with its streamlined experience, is not subject to this kind of error. (Tucker 1983; West and Blom 2017)

[6.] http://www.pewinternet.org/fact-sheet/mobile/

6. Telemetry data:

RDE is able to supplement collected attitudinal data with a rich array of para or telemetry data. As we know, people who answer surveys are fundamentally different than people who do not. As the progressive analytics shop CIVIS has argued recently, a battery of nearly 30 additional demographic, attitudinal, and lifestyle questions that get at notions of social trust and cosmopolitanism is necessary to be able to weight and correct for all the ways in which survey respondents are unusual. As Konitzer, Eckman and Rothschild (2016) argues, telemetry data is a much more cost-effective (and unobtrusive) way to collect these variables. Home and work location, commuting or mobility patterns or the political makeup of ones neighborhood or social network, derived from satellite-based (read: extremely accurate) longitudinal location-coordinate data predict demographic variables well, such as race and income. And, applications on the device can more accurately describe political traits prone to erroneous self-report, such as frequency of political discussion, political engagement or knowledge.

7. RDE will get stronger in the future:

Penetration of devices will further increase in the future, increasing reach of RDE in the US, and making RDE the only viable alternatives in less developed markets. Take Africa: the smartphone penetration rate is projected to grow at 52.9% year- on-year. Currently, there are 293 million smartphone users across the continent, meaning that taking into account current growth rates, there will be 929.9 million smartphones

by the year 2021 in Africa.[7] But the rosy future for RDE is not just about penetration. Advances in bridging Ad IDs with other known identifiers in the American market, such as voter file IDs, Experian Gold IDs, etc., mean that individual targeting based on financial history or credit card spending patterns will be possible. And, RDE will be able to adopt list-based polling, in which political survey firms poll directly from the voter file, large-scale administrative data detailing the turnout and registration history of 250,000,000 Americans.

8. River sampling is different, as devices are unknown:

River sampling can either mean banner-ad based polling or engagement with respondents via legacy websites or similar places RDE recruits from. In contrast to RDE, devices are unknown to river samplers: River sampling usually does not have access to the Ad ID, introducing two huge disadvantages: River samples have no way to address SUMA it is possible for fraudsters to engage with the same poll twice to increase chances to win the price for participation, especially if it comes in the form of financial incentives. And, any degree of demographic/geographic (not to mention individual) targeting is virtually impossible. In addition, banner ads themselves, similar to social-media ads, suffer from disastrous response rates. Good RDE polling is done with cooperation of the publisher, providing a native experience, while banners ads are pushed through the ad-network. This degraded user experience depresses response rates and can introduce serious measurement error.

[7] https://www.idgconnect.com/blog-abstract/23175/africa-2017-smartphone-penetration-open-data-online-freedom

Second, ad-networks optimize their delivery in a way that fights against the random sample. The users are chosen because they are more likely to respond, due to unobserved variables (at least to the survey researcher), that are correlated with how they will respond. As this underlying data is never shared, it is impossible to correct for by the survey researcher.

This method has some disadvantages:

Just like every other modern sample method RDD, assisted crowdsourcing, online panels RDE is non-probability. There is no sample method (anymore) that has perfect coverage and known probabilities for any respondent. This is one of the reasons we have developed analytics to overcome known biases. And, RDE has bias that we understand and can overcome, and additional data points that add to the power of correcting bias, such as telemetry data that is not available to RDD. While RDD has shifting and shrinking coverage, online panels suffer from panel fatigue and panel conditioning, and assisted crowdsourcing has bias introduced by efficient but to the polling firm nontransparent targeting algorithms that cannot be addressed, RDE is our method of choice, and the future, in the ever-changing market of polling.

3. Examples

Here we review work published in both Goel, Obeng and Rothschild (2015) and Konitzer, Corbett-Davies and Rothschild (N.d.) to showcase how effective RDE samples can be. And, add examples from the 2017-2018 special congressional elections.

Example 1:

(Goel, Obeng and Rothschild 2015) shows how RDE, through Pollfish, is able to closely match gold-standard polling such as the General Social Survey. This gold-standard uses yet another method: house-calls. This is unaffordable for most research, so we have left it off of this paper, but it provides a useful benchmark.

Example 2:

(Konitzer, Corbett-Davies and Rothschild N.d.) shows how RDE, utilizing the Pollfish platform, is able to closely match RDD polling in the 2016 election (actually doing slightly better). This is an example of using RDE samples with an anaylitics methods call Dynamic MRP. The anayltics methods is detailed in their paper.

When (Konitzer, Corbett-Davies and Rothschild N.d.) quantifies their state-bystate errors, they show that their predictions based on a single poll are not significantly worse than the predictions from poll aggregators. They compare their state-by-state estimates against the actual outcome. Compared to poll aggregator Huffington Post Pollster, their Root Mean Squared Error (RMSE) is only slightly higher: 4.24 percentage points vs. 3.62 percentage points (for 50 states excluding DC).

When they focus on the 15 closest states, they predictive accuracy is even higher. The RMSE is 2.89 percentage points, compared to 2.57 percentage points of Huffington Post Pollster. Overall, besides binary accuracy the RDE-based polling predictions also have low error in the precise percentage value. This is illustrated in Figure 1.
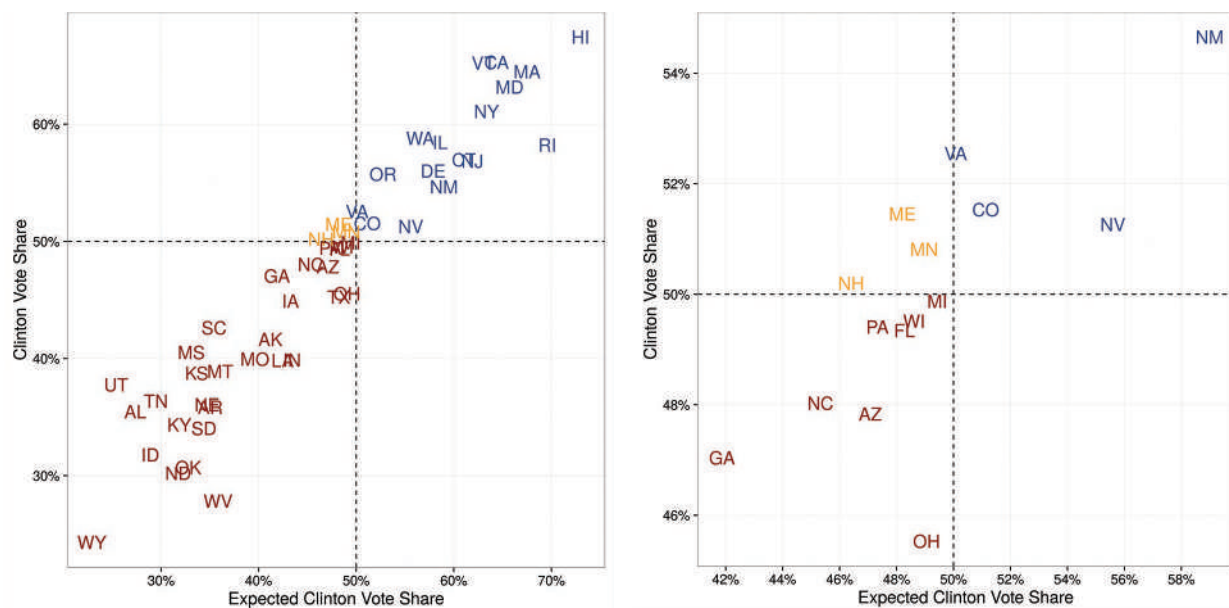

Figure 1: Predicted State-by-State Outcome of Presidential Election 2016 vs. actual State-by-State Outcome of Presidential Election 2016 for all states except DC (left panel); and for 15 closest States (right panel)

Not only are RDE-based polling state-by-state estimations fairly accurate, they also add meaningful signal to the poll aggregations. The left panel of Figure 2 displays the correlation between state-by-state errors of our predictions and the state-by-state errors of Huffington Post Pollster, and the right panel compares the distribution of errors across their approach and Huffington Post Pollster. At the very least, using RDE has significant potential to increase the quality of aggregators, as we discuss more below.
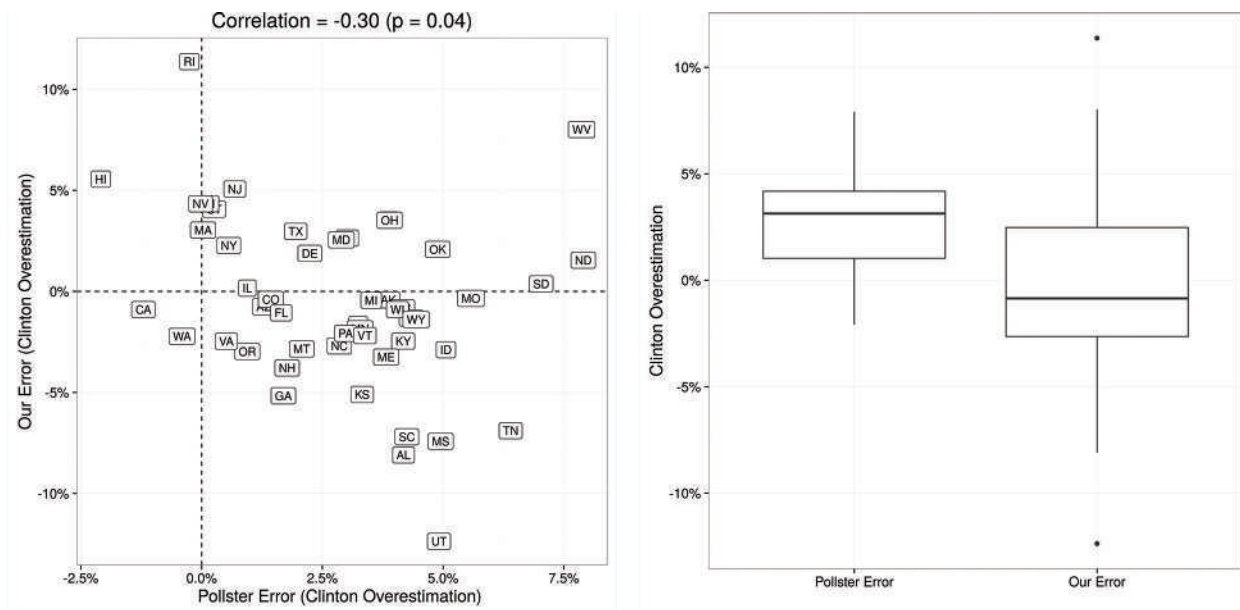
Figure 2: Correlation of Errors in Konitzer's State-by-State Prediction and Huffington Post Pollster's Estimates (left panel); Distribution of Errors in State-by-State Predictions in Konitzer's Approach and Huffington Post Pollster's Estimates (right panel)

Example 3:

During the course of 2017 and 2018 polling firms have employed all three new methods in predicting Congressional election outcomes: RDE comes out way above the other two.

4. Discussion

In this paper we outlined four methods of data collection for surveys. The first method, Random Digit Dialing (RDD), is the traditional method, working fine, but it is doomed in the next few years. Thus, the paper is really about which of the new innovations will replace it: online panels, Assisted Crowdsourcing, or Random Device Engagment (RDE). We believe strongly that RDE is the future.

5. References

Eckles, Dean, Brett R Gordon and Garrett A Johnson. 2018.
"Field studies of psychologically targeted ads face threats to internal validity."
Proceedings of the National Academy of Sciences p. 201805363.

Gelman, Andrew, Sharad Goel, Douglas Rivers, David Rothschild et al. 2016.
"The mythical swing voter."
Quarterly Journal of Political Science 11(1):103–130.

Goel, Sharad, Adam Obeng and David Rothschild. 2015.
"Non-representative surveys: Fast, cheap, and mostly accurate. In Working Paper."

Halpern-Manners, Andrew, John Robert Warren and Florencia Torche. 2017.
"Panel Conditioning in the General Social Survey."
Sociological Methods & Research
46(1):103–124. PMID: 28025587.
URL: https://doi.org/10.1177/0049124114532445

Kasprzyk, Daniel. 2005.
"Measurement error in household surveys: sources and measurement."
Technical report Mathematica Policy Research.

Konitzer, Tobias, Sam Corbett-Davies and David Rothschild. N.d.
"NonRepresentative Surveys: Modes, Dynamics, Party, and Likely Voter Space."
Forthcoming.
URL: http://researchdmr.com/Methods2016.pdf

Konitzer, Tobias, Stephanie Eckman and David Rothschild. 2016.
"Mobile as Survey Mode."

Porter, Stephen R, Michael E Whitcomb and William H Weitzer. 2004.
"Multiple surveys of students and survey fatigue."
New Directions for Institutional Research 2004(121):63–73.

Sturgis, Patrick, Nick Allum and Ian Brunton-Smith. 2009.
"Attitudes over time: The psychology of panel conditioning."
Methodology of longitudinal surveys 113:126.

Tucker, Clyde. 1983.
"Interviewer effects in telephone surveys."
Public Opinion Quarterly 47(1):84–95.

West, Brady T and Annelies G Blom. 2017.
"Explaining interviewer effects: A research synthesis."
Journal of Survey Statistics and Methodology 5(2):175–211.

Zaller, John R et al. 1992.
"The nature and origins of mass opinion."
Cambridge university press.

Contact: **sales@pollfish.com**

Thank you