USC Dornsife / Los Angeles Times 2016 Election "Daybreak" Poll

Data Documentation

Draft

Version: Wednesday, March 1, 2017

USC Dornsife
Dana and David Dornsife College of Letters, Arts and Sciences
Los Angeles Times

2016 Presidential Election "Daybreak" Poll

Table of Contents

USC Dornsife
Dana and David Dornsife College of Letters, Arts and Sciences
Los Angeles Times

2016 Presidential Election "Daybreak" Poll

Introduction

The 2016 USC Dornsife / LA Times Presidential Election "Daybreak" poll represents a pioneering approach to tracking changes in Americans' opinions throughout the 2016 campaign for the White House. The poll used non-traditional probabilistic polling methods rather than traditional polling's categorical vote questions and likely voter screens.

The Daybreak Poll's probabilistic polling approach is based on work done at RAND in 2008 (Delavande & Manski, 2010). Four years ago, the team responsible for the Daybreak poll developed the tracking poll methods we used in this election, to conduct the successful 2012 RAND Continuous Presidential Election Poll (Gutsche, Kapteyn, Meijer, & Weerman, 2014; Kapteyn, Meijer, & Weerman, 2012).

This document summarizes the various data files, methodological reports and other programs and documentation associated with the 2016 Daybreak poll. It includes question text, overviews and links to more detailed information regarding the poll's weighting and estimation schemes, and links to the Daybreak Poll's data and information. These data and documents, including this document, are available for download at https://uasdata.usc.edu/data/election-data.

Daybreak Poll-related news stories, op-eds, blog posts, and reports are listed on our poll coverage page.

The USC Dornsife/LA Times Presidential Election "Daybreak" Poll is one of many projects associated of the ongoing Understanding America Study (UAS) at the University of Southern California's (USC) Dornsife Center for Economic and Social Research (CESR), in partnership with the Jesse M. Unruh Institute of Politics and the Los Angeles Times. The UAS is an ongoing probability-based internet poll of about 6,000 individuals whose household addresses were randomly selected from zip codes across the United States. More information about the UAS is provided in this document, and online.

CESR's Daybreak Poll team included the Center's Executive Director Arie Kapteyn, Managing Director Tania Gutsche, Survey Director Jill E. Darling, Senior Economist Erik Meijer, and IT Director Bas Weerman.

Understanding America Study
Dornsife Center for Economic and Social Research
University of Southern California

PO Box 77902
Los Angeles, CA 90007-9983

Contact us at uas-l@mymaillists.usc.edu

To look at other survey data sets, visit https://uasdata.usc.edu

Find out more about our cost effective quarterly Omnibus Survey!

To get notices when there are new datasets,

Follow us on twitter

Probabalistic Polling Questions

Each day, we invited one-seventh of the members of the UAS election panel to answer three predictive questions. As shown in Figure 1 (click here for a larger view) those questions asked:

What is the percent chance that…

1. you will vote in the presidential election?
2. you will vote for Clinton, and for Trump, and for someone else (answers sum to 100%)
3. Clinton, Trump or someone else will win (answers sum to 100%)

Figure 1: Screenshot of the Daybreak Election Poll vote questions (order of candidates was randomized)

To find out more about what lay behind the vote, each week, following the three vote questions, we asked respondents one or two questions about their preferences and values. The data and documentation associated with these topical questions are described on page 9. The data were released on March 1, 2017.

Sampling and Recruitment

This section describes methods we used to conduct the Daybreak Poll in the Understanding America Study panel. We link to more detailed or technical documents available online.

Understanding America Study

The Daybreak Poll is based in the UAS, a probability-based internet panel survey that was established in 2014. UAS members are recruited from among all households in the United States and are a representative sample of U.S. residents. UAS members take about one to three surveys a month and are paid an incentive of about $.66 per question for each survey they complete. Recruited households that did not have internet access were provided with a tablet and internet service to ensure coverage of includes all ages and socioeconomic statuses.

The UAS panel grew over the course of the election season from about 4000 to about 6000. Membership in the election panel grew proportionally from about 3000 to 5000 (Figure 2).

Detailed information about the overall Understanding America Study's sampling and recruitment procedures is in the UAS Panel Sample and Recruitment document, available online.

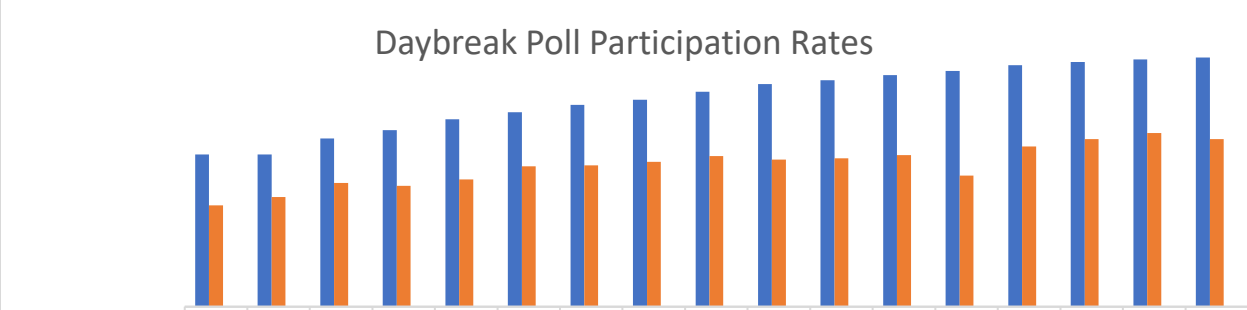
Additional information, including our recruitment protocol and detailed data on recruitment rates at each phase of UAS panel recruitment is available here.

Daybreak Poll

The Daybreak Poll started on July 4, 2016, and ran through November 7th, 2016.

Daybreak poll's election panel members were recruited via a UAS survey of their 2012 vote history (UAS 47). We invited a total of 5706 eligible UAS members to participate between May 16 and November 1, 2016. Of those, 5007 completed the survey (88%), and 4,857 consented to participate in the election panel (85%). Figure 2 shows a weekly summary of number of participants invited, participated, and response rate.

Figure 2: Number of election poll panel members invited, and participated, by week



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
invited	2957	2957	3270	3434	3641	3779	3922	4019	4180	4328	4403	4500	4583	4692	4758	4805	4842
participated	1968	2130	2404	2351	2471	2726	2748	2816	2926	2862	2882	2943	2545	3114	3254	3375	3260
%	66.6%	65.1%	70.0%	64.6%	65.4%	69.5%	68.4%	67.4%	67.6%	65.0%	64.0%	64.2%	54.2%	65.4%	67.7%	69.7%	67.3%

We assigned each Daybreak poll member a specific participation day of the week, so that 1/7th of the sample were invited to participate each day. Panel members could respond any day or time that was convenient before their next participation day.

Each night, the daybreak poll statistician weighted the Daybreak Poll results from the previous seven days to match demographic characteristics (race/ethnicity, gender, age, educational attainment) from the U.S. Census Current Population Survey and aligned self-reported vote in 2012 to the 2012 presidential election outcome. He posted the latest estimates based on that data online at election.usc.edu and on the LATimes.com Politics site shortly after midnight.

Immediately after the election, we conducted a post-election poll to determine how our participants actually voted (if they voted). We conducted several other post-election surveys on election-related topics. These surveys are linked on the Daybreak poll election data site at

Weighting and Estimation in the Daybreak Poll

We weighted the Daybreak poll sample to ensure that the poll's sample matched the population of interest on a number of characteristics: sex, age, race-ethnicity, education, household size, family income, and voting in the previous presidential election (2012). We used data from the Current Population Survey (CPS), restricted to U.S. citizens age 18 or older to create the reference benchmarks.

Weights were constructed using a raking method, which ensures that the distributions are exactly the same in the weighted poll data as in the reference population (e.g., Lu & Gelman, 2003; Valliant, Dever, & Kreuter, 2013, pp. 358–361; Kolenikov, 2014).

For more detailed information, please consult the document posted by our statistician online during the election: Weighting the Daybreak Poll.

To obtain the values shown in the election forecast chart we weighted each respondent's likelihood of voting for a candidate with their likelihood of voting in the presidential election. We then calculated the mean of that number for all respondents during the last 7 days, taking into account respondent-level weights based on demographics and past voting behavior. This created the estimated fraction of the population that we forecasted would vote for the candidate.

Detailed information about the Daybreak poll's estimation procedures is available online in the document: Sample selection and estimation in the Daybreak Poll.

Daybreak Poll Data files and Documentation

This section describes the data files compiled during the course of the election campaign, and the updated versions that we have posted with minor corrections. Data files and documentation, including this document, are available from https://uasdata.usc.edu/data/election-data. Detailed codebooks will also be made available at that location.

During the course of the election, the Daybreak Poll site's "detailed data" tab provided the day's chart data point in csv format. Detailed information about the chart files, which continue to be available on the site, is here: http://cesrusc.org/election/readme.txt.

Note that data from other UAS surveys can be linked to election survey data. The list of data files and topics is available at https://uasdata.uas.edu .

The poll generated two sets of microdata files: *fulldata* and *polldata*. The *fulldata* file included the cumulative microdata from all participants over the course of the election. Each record of the *polldata* file represents a daily estimate of the vote. The combined *polldata* file has the cumulative set of 7-day aggregations over each day of the election season.

On the Daybreak poll data site, we have provided both the original microdata files that were available during the election, a "corrected" fulldata file that reflects minor error fixes in variable definitions, and an aggregated *polldata* file. All changes away from the original are documented in the sections below.

Daybreak Poll Cumulative Microdata: the "fulldata" file

These files include the poll's cumulative set of individual responses over the course of the election. They are provided in stata format, zipped. A full codebook is available here.

Filenames	fulldata.dta (original fulldata file)
	full2016corrected.dta (updated file)
File locations	https://uasdata.usc.edu/election/fullraw/1
	https://uasdata.usc.edu/election/full/1
Observations	50249
Unique observations	4509

Unique identifiers for each observation

uasid	individual identifier
ts	time stamp of the observation

Panel variables

invite	Date invited to participate
pollwave/pollweek	Week invited to participate, and week the poll was completed
polldate	Date the poll was completed
polldayno	Sequential day number survey was completed (see Appendix, pg 11)

Daybreak Poll Daily Estimates: the "polldata" file

We have retained the final polldata file produced during the election, and a file that provides all of the Daybreak poll vote estimates calculated for the daily tracking chart from July 4 to November 7th, 2016. These are provided in the poll2016combined file. Each record represents a daily estimate of the forecast vote.

Filename	polldata.dta (original file) and poll2016combined.dta (combined file)
File location	https://uasdata.usc.edu/election/pollraw/1 (zipped file)
Observations	304496

Unique identifiers for each observation:

targetdate	date for which the analysis is used (last day of 7-day window)
uasid	individual identifier
ts	time stamp of the observation

Variable change from original: In this final data file, the "samp" variable which was included in the *polldata.dta* files posted online during the election season has been replaced by the targetdate variable.

Because most observations are used for 7 samples, i.e. each observation is included in seven daily tracking estimates, most (uasid,ts) observations occur 7 times, with different weights.

Weight variables

wgt_main	main weight
wgt_1--wgt_80	replication weights

Important information about weighting and analysis in the combined polldata file

The data have been declared and saved as survey data, with the proper weights, replication weights, and cluster info. However, the replication weights were developed for separate analyses of the different target samples, and in particular do not take dependence of observations across target samples into account.

For this reason, it is inappropriate to use this file for analyses that involve more than one target date. Hence, all analyses should be done with either ... if (targetdate == ...) or bysort targetdate: ... Any analyses (e.g., panel data analyses) that use samples that differ in some way from the target samples (e.g., use subsets or data that span more than 7 days) will need to compute dedicated weights. For analysis like these it is better to start from the fulldata file.

Also note that is difficult to impossible to recreate the polldata files from the fulldata file, because inclusion in any given day depends on the order of the observations in the raw dailypoll/fulldata file (which differed from day to day). It also varies by the Stata version used (July 10-15 were run in Stata 14, which uses a different random number generator, which affects the bootstrap replication weights). This should not matter for the estimates (which use wgt_main) but it slightly affects standard errors.

To use weights but do not want standard errors, you only need the main weight.

If you want standard errors but don't use Stata and don't know how to use replication weights, you can act as if the weights are design weights and compute robust s.e.'s using these; the calculation will typically be conservative.

Election Topics File

Each week, we asked election poll respondents to answer a very few (1 to 3) election-related questions, directly after the vote questions. Some questions were authored by the CESR team, others were asked by non-CESR researchers. These microdata were embargoed until February 28th, 2017 and released on the Daybreak data site on March 1.

We have provided a basic Stata "do" file with full question wording and code for processing and cleaning which data users may freely use, reference or adapt.

Each weekly set of questions was answered only once by each respondent who participated in the poll that week. Some sets of questions were repeated at intervals. These data may be merged with the *fulldata* file to compare respondents' responses to their candidate preference and turnout predictions from that week.

Unique identifiers for each observation:

uasid	individual identifier
ts	time stamp of the observation

Panel variables

pollwave/pollweek	Week invited to participate, and week the poll was completed
polldate	Date the poll was completed
polldayno	Sequential day number survey was completed (see Appendix, pg 11)

Data file location: https://uasdata.usc.edu/election/daybreaktopics/1
Data process file program location: http://cesrusc.org/election/daybreaktopics_proc.do

Weighting the election topics file: these data can be weighted as they are, or combined with fulldata.dta before constructing weights specific to that week's sample.

Demographics, party affiliation, and other useful analytic measures

Daybreak's data files include basic demographics. Measures such as party registration, party affiliation, political ideology, religiosity, candidate job ratings, etc. were collected at various points in time during the election season, and in several post-election polls, but are not included in *polldata* or *fulldata*.

To create these measures, several data sources must be combined. For example, aggregating party affiliation for most of the UAS panel requires combining three or more sources. We have provided example code for constructing such a variable on the data website: https://uasdata.usc.edu/data/election-data.

References

Delavande, Adeline, and Charles F. Manski. 2010. Probabilistic polling and voting in the 2008 presidential election: Evidence from the American Life Panel. Public Opinion Quarterly 74:433–459. doi: 10.1093/poq/nfq019

Gutsche, T. L., Kapteyn, A., Meijer, E., & Weerman, B. (2014). The RAND Continuous 2012 Presidential Election Poll. Public Opinion Quarterly, 78, 233–254. doi: 10.1093/poq/nfu009

Kapteyn, A., Meijer, E., & Weerman, B. (2012). Methodology of the RAND Continuous 2012 Presidential Election Poll (Working Paper No. WR-961). RAND Corporation. doi: 10.2139/ssrn.2146149

Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). Stata Journal,14,22–59.

Lu, H., & Gelman, A. (2003). A method for estimating design-based sampling variances for surveys with weighting, poststratification, and raking. Journal of Official Statistics, 19, 133–151.

Valliant,R.,Dever,J.A.,&Kreuter,F. (2013). Practicaltoolsfordesigningandweightingsurvey samples. NewYork,NY:Springer.

Appendix: Correspondence of of polldayno and calendar dates

pollday	date	pollday	date	pollday	date	pollday	date	pollday	date
1	4-Jul	29	1-Aug	60	1-Sep	90	1-Oct	121	1-Nov
2	5-Jul	30	2-Aug	61	2-Sep	91	2-Oct	122	2-Nov
3	6-Jul	31	3-Aug	62	3-Sep	92	3-Oct	123	3-Nov
4	7-Jul	32	4-Aug	63	4-Sep	93	4-Oct	124	4-Nov
5	8-Jul	33	5-Aug	64	5-Sep	94	5-Oct	125	5-Nov
6	9-Jul	34	6-Aug	65	6-Sep	95	6-Oct	126	6-Nov
7	10-Jul	35	7-Aug	66	7-Sep	96	7-Oct	127	7-Nov
8	11-Jul	36	8-Aug	67	8-Sep	97	8-Oct		
9	12-Jul	37	9-Aug	68	9-Sep	98	9-Oct		
10	13-Jul	38	10-Aug	69	10-Sep	99	10-Oct	Polling weeks	
11	14-Jul	39	11-Aug	70	11-Sep	100	11-Oct	week	Start date
12	15-Jul	40	12-Aug	71	12-Sep	101	12-Oct	1	4-Jul
13	16-Jul	41	13-Aug	72	13-Sep	102	13-Oct	2	11-Jul
14	17-Jul	42	14-Aug	73	14-Sep	103	14-Oct	3	18-Jul
15	18-Jul	43	15-Aug	74	15-Sep	104	15-Oct	4	25-Jul
16	19-Jul	44	16-Aug	75	16-Sep	105	16-Oct	5	1-Aug
17	20-Jul	45	17-Aug	76	17-Sep	106	17-Oct	6	8-Aug
18	21-Jul	46	18-Aug	77	18-Sep	107	18-Oct	7	15-Aug
19	22-Jul	47	19-Aug	78	19-Sep	108	19-Oct	8	22-Aug
20	23-Jul	48	20-Aug	79	20-Sep	109	20-Oct	9	29-Aug
21	24-Jul	49	21-Aug	80	21-Sep	110	21-Oct	10	5-Sep
22	25-Jul	50	22-Aug	81	22-Sep	111	22-Oct	11	12-Sep
23	26-Jul	51	23-Aug	82	23-Sep	112	23-Oct	12	19-Sep
24	27-Jul	52	24-Aug	83	24-Sep	113	24-Oct	13	26-Sep
25	28-Jul	53	25-Aug	84	25-Sep	114	25-Oct	14	3-Oct
26	29-Jul	54	26-Aug	85	26-Sep	115	26-Oct	15	10-Oct
27	30-Jul	55	27-Aug	86	27-Sep	116	27-Oct	16	17-Oct
28	31-Jul	56	28-Aug	87	28-Sep	117	28-Oct	17	24-Oct
		57	29-Aug	88	29-Sep	118	29-Oct	18	31-Oct
		58	30-Aug	89	30-Sep	119	30-Oct		
		59	31-Aug			120	31-Oct		