

# Methodology of the 2016 RAND Presidential Election Panel Survey (PEPS)

Michael Pollard, Joshua Mendelsohn

RAND Labor & Population

Sponsored by philanthropic contributions from RAND supporters, RAND, and UCLA



For more information on this publication, visit [www.rand.org/t/rr1460](http://www.rand.org/t/rr1460)

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2016 RAND Corporation

**RAND**® is a registered trademark.

#### Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions.html](http://www.rand.org/pubs/permissions.html).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

#### Support RAND

Make a tax-deductible charitable contribution at  
[www.rand.org/giving/contribute](http://www.rand.org/giving/contribute)

[www.rand.org](http://www.rand.org)

## Preface

---

This report describes the methodology to be used for the 2016 RAND Presidential Election Panel Survey (PEPS). Funding for this initiative was provided by philanthropic contributions from RAND supporters and income from operations and by the UCLA Departments of Political Science and Communication Studies. We thank Bonnie Ghosh-Dastidar and Rebecca Kilburn for their comments.

This research was undertaken within RAND Labor and Population, using the RAND American Life Panel. RAND Labor and Population has built an international reputation for conducting objective, high-quality, empirical research to support and improve policies and organizations around the world. Its work focuses on children and families, demographic behavior, education and training, labor markets, social welfare policy, immigration, international development, financial decisionmaking, and issues related to aging and retirement with a common aim of understanding how policy and social and economic forces affect individual decisionmaking and human well-being.

For more information on RAND Labor and Population, contact

Krishna Kumar  
Director  
RAND Labor and Population  
1776 Main Street, P.O. Box 2138  
Santa Monica, CA 90407-2138  
(310) 393-0411

or visit [www.rand.org/labor](http://www.rand.org/labor).

# Table of Contents

---

Preface.....	iii
Table of Contents.....	iv
Table and Figure.....	v
Abbreviations.....	vi
Methodology of the 2016 RAND Presidential Election Panel Survey.....	1
Sample.....	2
Weighting.....	3
Surveys.....	5
Predicting Election Results.....	6
Additional Analyses.....	7
References.....	8

## Table and Figure

---

### Table

1. PEPS Member Characteristics .....	3
--------------------------------------	---

### Figure

1. Key Election Dates and Timing of PEPS Survey .....	5
---	---

## Abbreviations

---

ALP	American Life Panel
CPS	2015 Current Population Survey (March Supplement)
PEPS	2016 Presidential Election Panel Survey

# Methodology of the 2016 RAND Presidential Election Panel Survey

---

The 2016 RAND Presidential Election Panel Survey (PEPS) uses a unique approach to study the evolution of public opinion, voting intentions, and voter behavior. Rather than surveying a new cross section of respondents, PEPS contacts the same respondents over the course of the election cycle. For many respondents, we can link responses to previously collected voting intentions, opinions, and behavior. We do this by fielding our surveys in the RAND American Life Panel (ALP).

The ALP consists of a panel of about 6,000 U.S. respondents ages 18 and older who regularly take surveys online. In contrast to most Internet panels, ALP respondents need not have Internet access when they are initially recruited; thus, the panel is a probability sample of the U.S. adult population. Since January 2006, the ALP has fielded surveys to its members on varied topics, including financial decisionmaking, the effect of political events on self-reported well-being, inflation expectations, joint retirement decisions, health decisionmaking, Social Security knowledge and expectations, measurement of health utility, voting preference in the presidential election, and more. All data collected by the ALP are publicly available to researchers.

The ALP is uniquely suited to identify changes over time because it is a panel survey, not a cross-sectional survey. When surveys sample a different group of people every time, it is difficult to distinguish whether people truly change their opinions over time or whether certain samples randomly included more people with a certain opinion. In contrast, the ALP asks the same questions to the same people over time, allowing for precise measurement of opinion change. In addition, a variety of information from ALP surveys dating to 2006 is available for many respondents and can be used to make a rich profile of individuals.

The PEPS includes a subsample of 3,000 ALP members and extends previous voting behavior work conducted with the ALP. We build on the findings of Delevande and Manski (2010) and Gutsche et al. (2014), which used a different approach to measuring voting intentions called *probabilistic polling*. Rather than asking respondents if they intended to vote and who they will vote for (single specific candidate), these two studies asked respondents about the probability of voting and the probability of voting for each candidate. These two papers also drew on surveys conducted in the ALP. Unlike these previous studies, the PEPS includes additional questions about a wide range of topics, such as opinions about political issues in the news, a variety of attitudes toward potential candidates, voting intentions and candidate preferences, underlying attitudes toward a range of groups, political affiliation and prior voting behavior, and perceived personality traits of candidates and the respondents themselves. Carman and Pollard (2014) also used a probabilistic polling approach, combined with a small set of questions on attitudes about current events for the 2014 midterm elections. Of the PEPS sample,

roughly one-third were enrolled during (and completed) the 2012 Continuous Presidential Election Poll surveys conducted in the ALP (Gutsche et al., 2014) and one-half participated in the Midterm 2014 Election Panel (Carman and Pollard, 2014). Gutsche et al. (2014) and Kapteyn, Meijer, and Weerman (2014) described in detail how an election poll can be implemented in the ALP. We based much of our methodology on theirs.

The PEPS methodology has two primary contributions relative to traditional polling. First, the use of a panel allows us to link respondents to their previous survey responses. This allows us to gain a better understanding of changes in voting intentions, behavior, and individual characteristics that predict or correlate with change. Not only will we investigate how voting intentions change leading up to the election, but for many respondents, we will be able to look at past voting and other behavior. Second, the use of intentions expressed in probability units can shed greater insight into voters in the middle, compared with binary (yes or no) responses about voting for a given candidate or voting at all. For those who are certain who they will vote for, simple questions can gauge their intentions; for undecided voters, probabilistic questions can provide more insight into what they may do. These benefits of the PEPS contributed to the accuracy of our predictions in the Continuous 2012 Presidential Election Poll, which was one of the most accurate predictors of the popular vote in 2012 (Silver, 2012).

## Sample

Survey participants for PEPS are drawn from the ALP. The ALP began surveying respondents in January 2006; since then, more than 400 surveys have been fielded to the panel, which has grown from about 1,000 respondents to more than 5,000. The ALP is a scientifically recruited, nationally representative Internet panel; panel members are recruited using address-based sampling and random digit dial sampling.<sup>1</sup> Unlike opt-in online surveys, Internet access is not required to participate; those who do not already have Internet access or computers are provided with laptops and Internet access. Provision of Internet access is an important feature for ensuring population representation in Internet panels (Couper et al., 2007; Schonlau et al., 2009). Over the history of the ALP, recruiting methods have evolved. Detailed information about the sample composition and the past recruiting methods can be found at the RAND ALP website (RAND American Life Panel, undated [a]).

We recruited 3,000 respondents for our baseline survey and will use this same sample throughout subsequent surveys. Because of the importance of the baseline data for assessing individual-level change over time, we will not include new respondents in subsequent waves. Our sample is limited to U.S. citizens, as we are examining presidential election voting behavior.

---

<sup>1</sup> The ALP also has several nonprobability-based subsamples. These respondents, including snowball samples and within-family referrals, are excluded from this research.



Table 1 provides summary statistics of the demographic characteristics of our baseline sample with and without weights, which are described in detail below.

**Table 1. PEPS Member Characteristics**

<b>Variable</b>	<b>Weighted to CPS</b>	<b>CPS (Weighted)</b>
Age	47.3	47.1
Male	48.1%	48.2%
Race		
White, non-Hispanic	64.9%	64.9%
Black, non-Hispanic	11.7%	11.7%
Hispanic	20.0%	15.5%
Other	3.4%	7.9%
Household size	2.6	3.0
Education		
Less than high school	7.9%	12.3%
High school	33.9%	29.4%
Some college	19.0%	19.5%
College	26.4%	28.2%
Advanced degree	12.8%	10.6%

## Weighting

As with other surveys, we apply weights to ensure that our results are representative of the overall population. Analysis of our results will be based on two approaches. For some, we will draw significantly on previous data collected in the ALP, either as part of this project or from previous years. At other times, we will focus on cross-sectional data, using results from one survey. Each approach will require a different weighting scheme.

With cross-sectional data, we will use a raking algorithm (DeBell and Krosnick, 2009) to match the characteristics of our sample to that of citizens ages 18 and older in the 2015 Current Population Survey (CPS).<sup>2</sup> We match on age, gender, race/ethnicity, education, household size, and income distribution, as well as the joint bivariate distributions of race and gender, education and gender, age and gender, and income and household size. In order to create weights, it is necessary to account for missing values of certain weighting variables for some observations in the PEPS data. We impute missing values sequentially, beginning with the more-basic (and less-

---

<sup>2</sup> The raking algorithm generates sample weights by matching the proportions of predefined strata in the PEPS to those in the CPS, using a series of two-way marginal distributions: gender  $\times$  age, gender  $\times$  ethnicity, gender  $\times$  education, gender  $\times$  household income, household income  $\times$  number of household members. These strata are defined such that none of them contains fewer than 5 percent of the PEPS sample.

frequently missing) demographic traits of gender, age, and citizenship, for which we replace missing values with the modes of each variable. The remaining missing variables are then imputed using linear regression for continuous variables and logistic regression for discrete variables (including multinomial logistic regression or ordinal multinomial logistic regression for discrete variables with more than two outcomes). Missing values on these characteristics are rare in the PEPS data, with fewer than 0.5 percent of values missing for each variable used in weighting.

A second stage of weighting is done separately for cross-sectional analysis. This consists of reweighting the sample using post-stratification, such that its recall of voting behavior in 2012 matches known population voting behavior in 2012; essentially the reported 2012 voting behavior of the PEPS is weighted to match the known voting behavior of the electorate in 2012 (DeBell and Krosnick, 2009). This is based on the premise that the best predictor of future voting behavior is past voting behavior and that any discrepancies in composition with respect to past voting behavior are likely to give biased predictions of voting behavior in the 2016 election.

Information about the voting behavior of PEPS members in 2012 is obtained from the answers to I1 and I2 in the baseline PEPS survey:

I1: Did you vote in the 2012 presidential election? (yes or no)

I2: Did you vote for Obama, Romney, or another candidate?

All respondents reported answers to I1.<sup>3</sup> The post-stratification weighting procedure was then identical to that used in the RAND Continuous 2012 Presidential Election Poll (see Kapteyn, Meijer, and Weerman, 2014).

When we use data from more than one survey (either multiple waves of PEPS or waves of PEPS linked to previous ALP surveys), we will use a two-stage weighting procedure. The first step matches that used for the cross-sectional approach. In the second step, we will use inverse probability weighting (Duncan and Hill, 1989) to account for factors that may predict participation in the second survey. Factors to be included in the inverse probability weights include demographic characteristics, past voting behavior, and/or party preference. For respondents who did not participate in the 2012 election polls, this information will be collected as part of the PEPS. Inverse probability weights are calculated using a regression model to predict who participates in the second survey. The results of these regressions may be of interest in and of themselves. Literature on political polling suggests that participation in polls may be biased by current events, with individual respondents' likelihood of participation varying

---

<sup>3</sup> Eight hundred and eighty-eight baseline respondents also completed a post-2012 presidential election survey that asked about voting behavior. The agreement between the 2012 and baseline surveys for these responses was more than 91 percent. During the 2014 midterm elections, ALP members were again asked if they voted in the 2012 presidential election; 1,442 of the PEPS baseline respondents had previously answered this question, and the agreement rate was 95 percent.

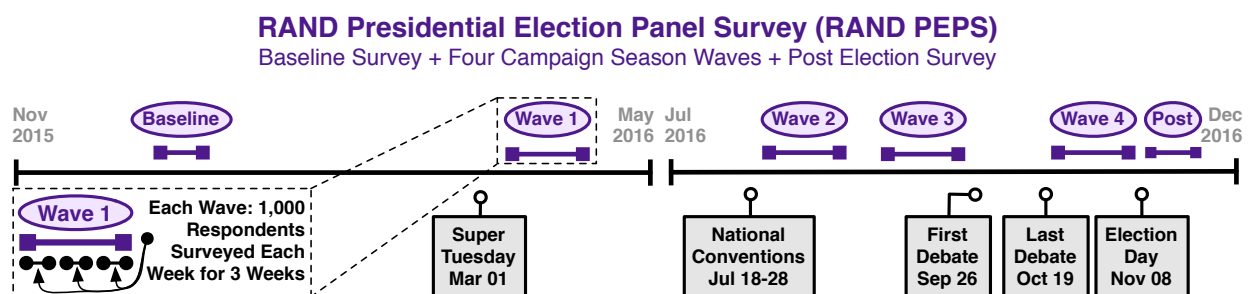
depending on current events. Our methodology will also allow us to look at whether the likelihood of participating in the surveys varies from wave to wave.

## Surveys

The baseline PEPS survey was opened to panel members at midnight on Sunday, December 13, 2015. All panel members were invited to participate, but they were also told the survey would only be open to a limited number of respondents for a limited time. By 11:59 p.m. on Thursday, December 31, 2015, the baseline survey had been completed by 2,818 respondents. The baseline survey was closed when the 3,000-respondent target was met, at approximately 9:45 a.m. on Wednesday, January 6, 2016; the total baseline sample was 3,037 at the end. Note that budget constraints limited a maximum number of 3,000 participants for all planned waves.

This baseline sample will be invited to take at least five additional surveys at key points throughout the 2016 election cycle, as shown in Figure 1.

**Figure 1. Key Election Dates and Timing of PEPS Survey**



For Waves 1–3, the baseline sample will be divided into three random subsamples of 1,000 members, who will be invited to participate in a short survey of ten or fewer questions on a rolling basis. For each wave, one-third of the sample will be invited during the first week, one-third will be invited during the second week, and one-third will be invited during the third week. Each survey will remain in the field for two weeks for each subsample, in effect creating a continuous poll over the course of one month. Each subsample of 1,000 is sufficiently large to power analysis independently, as well as when combined. This strategy has the added benefit of potentially capturing the effect of significant unexpected events on voter opinions. Wave 4 will be fielded to the entire baseline sample simultaneously because of the relatively short time between the final debate and election day. The post-election survey will be fielded immediately after the election.

We will assess voting intentions during the first four waves, in addition to the baseline. Voting intentions will be assessed with two questions. Our methodology, as well as question wording, are derived from that of Delevande and Manski (2010) and Gutsche et al. (2014). First, respondents will be asked the following:

Q1. We'd like you to ask you to think about the upcoming Presidential election in 2016.

What is the percent chance that you will vote in the Presidential election?  
\_\_\_\_\_ %

The percent chance can be thought of as the number of chances out of 100. You can use any number between 0 and 100. For example, numbers like 2 and 5 percent may be 'almost no chance', 20 percent or so may mean 'not much chance', a 45 or 55 percent chance may be a 'pretty even chance', 80 percent or so may mean a 'very good chance', and 95 or 98 percent chance may be 'almost certain'.

The description of probabilities included in question Q1 is similar to that included in other surveys as described in Manski (2004).

Q2. If you do vote in the election, what is the percent chance that you will vote for a Democrat? And for a Republican? And for someone else? Please provide percent chances in the table below.

Democrat	_____ %
Republican	_____ %
Someone else	_____ %
Total	_____ %

The order of the Democrat and Republican rows will be randomized. The total will automatically be summed as people type in their responses. If responses do not add up to 100, an error message is shown that says: "Your total does not add up to 100%, please return to the previous question and fix your answers." Only responses between zero and 100 will be allowed to avoid problems faced in the previous presidential election poll, in which respondents gave inappropriate responses such as "1100%" and "-1000%."

## Predicting Election Results

While predicting voter behavior is only a part of what these data will allow us to do, it is important to describe the methodology in detail. Here, we draw heavily on Kapteyn et al. (2014). Refer to that publication for additional details.

Likely voter turnout can be calculated as the weighted average of the reported probability of voting:

$$Turnout = \sum_i w_i Q1_i$$

where  $w_i$  is the weight and  $Q1_i$  is the response to question Q1 for individual  $i$ . Weights are calculated as described above.<sup>4</sup> The expected percentage of the vote for each party can be calculated as

$$\text{Popular Vote} = \sum_i w_i Q1_i \frac{QJ_2}{100}$$

where  $J$  represents political party (Democrat, Republican, or other).

Standard errors are calculated in accordance with those discussed in Kapteyn et al. (2014).

## Additional Analyses

In addition to examining voting intentions and behavior, the baseline survey collected in December 2015 contained a wide range of questions, including such topics as opinions about political issues in the news, a variety of attitudes toward potential candidates, candidate preferences, underlying attitudes toward a range of groups, political affiliation, prior voting behavior, and perceived personality traits of the respondents themselves. The specific voting intention questions are designed to more accurately capture the likely votes of a greater number of voters in the crucial “middle” (that is, not closely aligned with a candidate) and were used with great success in the 2012 presidential election.<sup>5</sup>

Looking beyond voter behavior, our data collection will also allow us to do a number of additional analyses, focused instead on associations between respondent and candidate characteristics, for example. In these cases, analyses will differ depending on the specific topic. These will often draw on historical data collected in the ALP but will also address new and timely topics.

Additional survey content in planned waves and additional survey waves are possible, given research community interest and support. The PEPS data will be used to produce a series of blogs published on RAND’s website throughout the election cycle. As with all RAND ALP data, they will eventually become freely available to registered ALP data users.<sup>6</sup>

---

<sup>4</sup> Election predictions will be estimated using individual surveys and the cross-sectional weights.

<sup>5</sup> Respondents are not limited to choosing a single candidate; instead, they are forced to report on the likelihood (0–100 percent) of voting for each candidate.

<sup>6</sup> Researchers may access ALP data for free by registering as an ALP Data Page user. See RAND American Life Panel, undated (b).

## References

---

- Carman, Katherine Grace, and Michael Pollard, *Methodology of the RAND Midterm 2014 Election Panel*, Santa Monica, Calif.: RAND Corporation, RR-854-RC, 2014. As of January 25, 2016:  
[http://www.rand.org/pubs/research\\_reports/RR854.html](http://www.rand.org/pubs/research_reports/RR854.html)
- Couper, Mick P., Arie Kapteyn, Matthias Schonlau, and Joachim Winter, “Noncoverage and Nonresponse in an Internet Survey,” *Social Science Research*, Vol. 36, No. 1, 2007. pp. 131–148.
- DeBell, Matthew, and Jon A. Krosnick, *Computing Weights for American National Election Study Survey Data*, Ann Arbor, Mich., and Palo Alto, Calif.: American National Election Studies, ANES Technical Report series, No. nes012427, September 1, 2009. As of January 25, 2016:  
<http://www.electionstudies.org/resources/papers/nes012427.pdf>
- Delavande, Adeline, and Charles F. Manski, “Probabilistic Polling and Voting in the 2008 Presidential Election: Evidence from the American Life Panel,” *Public Opinion Quarterly*, Vol. 74, 2010, pp. 433–459.
- Duncan, Greg J., and Daniel H. Hill, “Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics,” *Journal of Business and Economic Statistics*, Vol. 7, No. 4, October 1989, pp. 441–452.
- Gutsche, Tania L., Arie Kapteyn, Erik Meijer, and Bas Weerman, “The RAND Continuous 2012 Presidential Election Poll,” *Public Opinion Quarterly*, Vol. 78, No. S1, 2014, pp. 233–254.
- Kapteyn, Arie, Erik Meijer, and Bas Weerman, *Methodology of the RAND Continuous 2012 Presidential Election Poll*, Santa Monica, Calif.: RAND Corporation, RR-858-RC, 2014. As of January 25, 2016:  
[http://www.rand.org/pubs/research\\_reports/RR858.html](http://www.rand.org/pubs/research_reports/RR858.html)
- Manski, Charles F., “Measuring Expectations,” *Econometrica*, Vol. 72, No. 5, September 2004, pp. 1329–1376.
- RAND American Life Panel, “Welcome to the ALP Data Pages,” website, undated (a). As of January 26, 2016:  
<https://alpdata.rand.org>
- RAND American Life Panel, “New User Registration,” website, undated (b). As of February 10, 2016:  
<https://alpdata.rand.org/index.php?page=data&p=shownewuser>

Schonlau, Matthias, Arthur van Soest, Arie Kapteyn, and Mick Couper, "Selection Bias in Web Surveys and the Use of Propensity Scores," *Sociological Methods and Research*, Vol. 37, No. 3, February 2009, pp. 291–318.

Silver, Nate, "Which Polls Fared Best (and Worst) in the 2012 Presidential Race," *FiveThirtyEight* blog, November 10, 2012. As of January 25, 2016:  
[http://fivethirtyeight.blogs.nytimes.com/2012/11/10/  
which-polls-fared-best-and-worst-in-the-2012-presidential-race/?\\_r=0](http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?_r=0)