**Description of the Weight Computed for ATP Wave 37 Using Social Media Variables (WEIGHT\_W37\_SOCIALMEDIA)**

A custom weight, called WEIGHT\_W37\_SOCIALMEDIA, was computed for the 4,581 panelists who responded to Wave 37 of the Pew American Trends Panel (ATP). This weight was created to provide continuity and consistency for estimates of social media news usage with prior waves that followed a similar procedure (waves 14 and 28). The computation of the weight was similar to that of routine wave weights, except that the calibration adjustment was revised to additionally adjust for the distribution of use of social media to get news or news headlines that was measured in earlier waves. The weight was computed in three main stages:

* Base weight adjusting for differential probabilities of selection;
* Propensity adjustment for attrition;
* Calibration to demographic distributions for the target population.

## *Base Weight*

A base weight was computed for all ATP members. The base weight adjusted for factors affecting the probability that the individual was selected for the panel. This probability came from the survey in which the respondent was recruited. Currently, all ATP members were recruited through three probability-based, national, overlapping, dual-frame landline and cell phone RDD surveys. In the landline sample of the RDD surveys, one adult was randomly selected from within the household. Interviewers asked to speak with either the youngest male or youngest female at home at the time of the call. In the cell sample of the RDD surveys, interviews were conducted with the person who answered the phone, provided they were age 18+ and spoke English or Spanish.

For most ATP members, their base weight was computed using single frame estimation to adjust for the probability that the respondent’s phone number was selected from the sampling frame, the overlap in the landline and cell phone frames, and the within-household selection in the landline sample. For most panel members, the base weight can be expressed as:

Where:

LL =1 if respondent has a landline phone

=0 if respondent has no landline phone

CP =1 if respondent has a cell phone

=0 if respondent has no cell phone

Sll= number of cases released in the landline sample

Scp=number of cases released in the cell phone sample

Ull=size of the landline RDD frame

Ucp=size of the cell phone RDD frame

AD=number of adults in the household (1, 2, 3 or more)

For a subset of the ATP members, an additional adjustment was included in the base weight to account for the fact that they belong to a group that was subsampled for invitation to the panel. In the Typology Survey, non-Internet users were subsampled at a rate of 25% from January 23, 2014 through February 5, 2014, but they were not subsampled (100% invited) from February 6, 2014 through the end of the field period. Internet users who agreed to join the panel in the Typology Survey but did not have an email address were taken at 100% from January 23, 2014 through February 5, 2014, but they were subsampled at a rate of 25% from February 6, 2014 through the end of the field period. The base weight of the affected cases was multiplied by the inverse of the subsampling rate (1 / .25 = 4). In the Panel Refresh Survey, non-Hispanic white Internet users with more than a high school education were subsampled at 50%. The base weight of such cases was also multiplied by the inverse of the subsampling rate (1 / .5 = 2).

## *Propensity Adjustment for Attrition*

In total, 19,719 RDD survey respondents were invited to join the ATP and 9,942 accepted, yielding a panel acceptance rate of 50.4%. A majority of those who agreed to join the panel were still active at the start of Wave 37 (5,475/9,942=55.1%). To the extent that active panel members may be different from individuals who are not active (either because they declined to join or because they dropped out), there is a risk that estimates from the panel could be subject to nonresponse bias. A propensity score adjustment was computed to adjust for this attrition.

Most of the information available for individuals who either declined the panel invitation or were dropped from the panel comes from the recruitment surveys. A logistic regression model was estimated in which being an active panel member was regressed on recruitment survey sampling frame, incentive amount ($10/20 vs $5 per survey), Internet user, race, child in the household, age, education, religious service attendance, survey recruitment (Typology vs. Governance/Panel Refresh), registered to vote, party identification (Republican vs. Democrats/Independent/Others), and statistically significant two-way interactions (*p* < .05). The model was estimated using the respondents in the recruitment surveys who were invited to join the panel. Hispanic ethnicity was excluded from the model because it was collinear with the incentive variable. Number of adults in the household, child in the household and incentive were not predictive and ultimately excluded from the model. The set of predictors considered for the model are variables that are routinely measured in surveys conducted for the Pew Research Center for the People & the Press. The significant predictors used in the final model are presented in Table 1.

The estimated propensities were used to divide cases into approximately equal-size groups using the quintiles of the estimated propensity score. Quintiles have been found to be effective in capturing most of the variation. The propensity score adjustment was computed as the inverse of the active status rate in each quintile. This approach helps to protect against model misspecification, relative to using the inverse of the propensity score.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1. Parameter Estimates from the Attrition Propensity Model^** | | | | |
| **Variable (reference group)** | **Estimate** | **s.e.** | **p-value** |  |
| Intercept | -2.299 | 0.532 | <.001 | \*\*\* |
| Frame (landline) | -0.314 | 0.126 | 0.013 | \* |
| Gender (male) | 0.742 | 0.133 | <.001 | \*\*\* |
| Internet User (non-user) | -0.143 | 0.488 | 0.770 |  |
| Race (other race) |  |  | <.001 | \*\*\* |
| White | 0.358 | 0.075 | <.001 | \*\*\* |
| African American | 0.239 | 0.088 | 0.007 | \*\* |
| Asian | 0.060 | 0.120 | 0.616 |  |
| Multi-racial | 0.359 | 0.110 | 0.001 | \*\* |
| Child in HH (no children) | -0.267 | 0.121 | 0.027 | \* |
| Age | -0.026 | 0.006 | <.001 | \*\*\* |
| Education | 0.066 | 0.054 | 0.223 |  |
| Religious Attendance | 0.263 | 0.067 | <.001 | \*\*\* |
| Recruitment Survey |  |  | <.001 | \*\*\* |
| Governance | 0.251 | 0.154 | 0.104 |  |
| Panel Refresh | 0.875 | 0.174 | <.001 | \*\*\* |
| Registered to Vote | -0.941 | 0.167 | <.001 | \*\*\* |
| Party Identification |  |  | <.001 | \*\*\* |
| Democrat | 1.387 | 0.264 | <.001 | \*\*\* |
| Independent/Others | 0.848 | 0.266 | 0.001 | \*\* |
| Frame x Age | 0.009 | 0.002 | <.001 | \*\*\* |
| Internet User x Age | 0.024 | 0.005 | <.001 | \*\*\* |
| Internet User x Education | 0.155 | 0.053 | 0.003 | \*\* |
| Internet User x Religious Attendance | -0.149 | 0.050 | 0.003 | \*\* |
| Internet User x Party Identification |  |  | 0.034 | \* |
| Internet User x Democrats | -0.620 | 0.238 | 0.009 | \*\* |
| Internet User x Independent/Others | -0.481 | 0.242 | 0.047 | \* |
| Gender x Age | -0.007 | 0.002 | 0.001 | \*\* |
| Gender x Education | -0.052 | 0.019 | 0.005 | \*\* |
| Child in HH x Age | 0.006 | 0.003 | 0.017 | \* |
| Age x Religious Attendance | -0.002 | 0.001 | 0.004 | \*\* |
| Education x Registered to Vote | 0.062 | 0.026 | 0.017 | \* |
| Religious Attendance x Registered to vote | 0.090 | 0.030 | 0.003 | \*\* |
| Religious Attendance x Party Identification |  |  | <.001 | \*\*\* |
| Religious Attendance x Democrats | -0.136 | 0.029 | <.001 | \*\*\* |
| Religious Attendance x Independent/Others | -0.079 | 0.028 | 0.005 | \*\* |
| Recruitment Survey x Age |  |  | 0.002 | \*\* |
| Governance x Age | -0.002 | 0.002 | 0.302 |  |
| Panel Refresh x Age | -0.009 | 0.003 | <.001 | \*\*\* |
| Recruitment Survey x Education |  |  | 0.013 | \* |
| Governance x Education | -0.036 | 0.021 | 0.093 |  |
| Panel Refresh x Education | -0.072 | 0.026 | 0.005 | \*\* |
| \*\*\* *p*<.001, \*\* *p*<.01, \* *p*<.05 |  |  |  |  |
| ^Variables are coded such that the model predicts active status in the panel. Positive coefficients are associated with a higher probability of being active. Negative coefficients are associated with lower probability of being active. | | | | |

## *Calibration to Target Population Controls*

In the final stage of weighting, the attrition-adjusted base weights for the panelists responding to a particular panel survey are calibrated to population benchmarks using raking, or iterative proportional fitting. This adjustment is designed to reduce the risk of nonresponse bias stemming from nonresponse at the various stages of the panel design. The raking dimensions and the source for the population parameter estimates are reported in Table 2. All raking targets are based on the non-institutionalized U.S. adult (age 18+) population.

Most of the dimensions are commonly observed in weighting protocols for general population household surveys in the US. One exception is the raking for Internet usage. This was included in the algorithm so that the panel survey estimates reflect the target population with respect to the proportion of people who use the Internet and the proportion who do not. In Wave 37, all ATP interviews were completed via self-administered Web survey. Therefore, there was a concern that Internet users could be over-represented in the survey estimates if this dimension was not controlled for in the raking. To correct for this potential over-representation, panelists who reported at the time of the recruitment survey that they did not use the Internet were used to represent non-Internet users in the raking. Currently, the estimated population parameter for the percent of U.S. adults who use the Internet is 90.2%, based on the 2017 ATP Panel Refresh Survey conducted for the Pew Research Center. (While it would have been preferable to use a large, federal, in-person survey (such as the American Community Survey or the Current Population Survey) to obtain this parameter estimate, the federal government does not routinely measure Internet access from any location.[[1]](#footnote-1),[[2]](#footnote-2))

Another dimension that is not typically used in weighting protocols for general population household surveys in the US is volunteering. This variable was included in the calibration to adjust for some potential bias due to the over-representation of more politically- and civically-engaged adults of the panel identified in some recent analysis.

Finally, variables describing the respondents’ use of nine different social media platforms for getting news were also included in the algorithm. Panelists were weighted using their previous responses to these questions from Wave 28. For panelists who responded to Wave 37 but not to Wave 28, values for these items were imputed using a chained equations approach via classification and regression trees (CART). These variables were then weighted so that respondents’ past social media news usage matched the distribution that was observed in Wave 28 using social media weights that were created using a similar procedure.

|  |  |
| --- | --- |
| **Table 2. Raking Dimensions and Source for Population Parameter Estimates** | |
| **Raking Dimension^** | **Source** |
| Gender(2) x Age(6) | 2016 American Community Survey |
| Gender(2) x Education (3) | 2016 American Community Survey |
| Age(3) x Education(3) | 2016 American Community Survey |
| Census Region(4) | 2016 American Community Survey |
| Race/Ethnicity(4) | 2016 American Community Survey |
| Population Density(5) | 2010 Decennial Census |
| Internet Usage(2) | 2017 ATP Panel Refresh Survey |
| Party Affiliation(3) | Average from the three most recent monthly surveys conducted for the Pew Research Center for the People & the Press |
| Volunteerism(2) | September 2015 Current Population Survey Volunteer Supplement |
| Use of Facebook to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of Twitter to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of LinkedIn to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of Instagram to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of Tumblr to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of YouTube to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of reddit to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of Snapchat to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| Use of WhatsApp to get news or news headlines(3) | ATP Wave 28 August 2017 (Social media weights) |
| ^ The numbers of categories (prior to any collapsing from small cell size) are shown in parentheses. | |

***Trimming***

The distribution of the raked weights was then evaluated and checked for extreme values. The weights were trimmed at the 1st and 99th percentiles.

***Design Effect and Margin of Error***

Weighting and survey design features that depart from simple random sampling tend to result in an increase in the variance of survey estimates. This increase, known as the design effect or *deff*, should be incorporated into the margin of error, standard errors, and tests of statistical significance. The overall design effect for a survey is commonly approximated as 1 plus the squared coefficient of variation of the weights. For this survey, the margin of error (half-width of the 95% confidence interval) incorporating the design effect for full sample estimates at 50% is ± 2.5 percentage points. Estimates based on subgroups will have larger margins of error. It is important to remember that random sampling error is only one possible source of error in a survey estimate. Other sources, such as question wording and reporting inaccuracy, may contribute additional error. A summary of the weights and their associated design effect is reported in Table 3 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3. Design Effect and Effective Sample Size** | | |  |  |
| **Weight Variable** | **Completed Interviews** | **Approximate Design Effect** | **Effective Sample Size** | **Margin of Error (95% confidence level)** |
|
| WEIGHT\_W37\_SOCIALMEDIA | 4,581 | 2.97 | 1,541 | ± 2.50 |

1. The July 2011 Current Population Survey estimated that 73% of US residents age 15 and older access the internet from some location. Given the increasing trends in internet access, particularly on mobile devices, this 2011 CPS estimate was deemed too out-of-date to be helpful in the ATP weighting. [↑](#footnote-ref-1)
2. Starting in 2013 the American Community Survey is measuring internet access, but it only measures access inside the sample household. Members of the ATP are permitted to complete the surveys from any location. So the more relevant parameter for the ATP is the proportion of adults who can access the internet from any location, not just at home. [↑](#footnote-ref-2)