Sampling and Weighting Methodology for the February 2020 Texas Statewide Study

For the survey, YouGov interviewed 1443 Texas registered voters between January 31 and February 9, 2020, who were then matched down to a sample of 1200 to produce the final dataset. The respondents were matched on gender, age, race, and education. YouGov then weighted the matched set of survey respondents to known characteristics of registered voters of Texas from the 2018 Current Population survey and 2014 Pew Religious Landscape Survey.

The respondents were matched to a sampling frame on gender, age, race, and education. The frame was constructed by stratified sampling from the full 2016 Current Population Survey (CPS) voter registration supplement with selection within strata by weighted sampling with replacements (using the person weights on the public use file). For the main sample, the matched cases were weighted to the sampling frame using propensity scores. The matched cases and the frame were combined and a logistic regression was estimated for inclusion in the frame. The propensity score function included age, gender, race/ethnicity, and years of education. The propensity scores were grouped into deciles of the estimated propensity score in the frame and post-stratified according to these deciles. These weights were then post-stratified on baseline party identification, the 2016 presidential vote, ideology, and a full stratification of four-category age, four-category race, gender, and four-category education. The weights were trimmed at 7 and normalized to sum to the sample size.

The margin of error of the weighted data for registered voters is 3.6%. The margin of error for likely Democratic voters is 5.0%.

Survey Panel Data

The YouGov panel, a proprietary opt-in survey panel, is comprised of 1.2 million U.S. residents who have agreed to participate in YouGov Web surveys. At any given time, YouGov maintains a minimum of five recruitment campaigns based on salient current events.

Panel members are recruited by a number of methods and on a variety of topics to help ensure diversity in the panel population. Recruiting methods include Web advertising campaigns (public surveys), permission-based email campaigns, partner sponsored solicitations, telephone-to-Web recruitment (RDD based sampling), and mail-to-Web recruitment (Voter Registration Based Sampling).

The primary method of recruitment for the YouGov Panel is Web advertising campaigns that appear based on keyword searches. In practice, a search in Google may prompt an active YouGov advertisement soliciting opinion on the search topic. At the conclusion of the short survey respondents are invited to join the YouGov panel in order to receive and participate in additional surveys. After a double opt-in procedure, where respondents must confirm their consent by responding to an email, the database checks to ensure the newly recruited panelist is in fact new and that the address information provided is valid.

The YouGov panel currently has over 20,000 active panelists who are residents of Texas. These panelists cover a wide range of demographic characteristics.

Sampling and Sample Matching

Sample matching is a methodology for selection of "representative" samples from non-randomly selected pools of respondents. It is ideally suited for Web access panels, but could also be used for other types of surveys, such as phone surveys. Sample matching starts with an enumeration of the target population. For general population studies, the target population is all adults, and can be enumerated through the use of the decennial Census or a high-quality survey, such as the American Community Survey. In other contexts, this is known as the sampling frame, though, unlike conventional sampling, the sample is not drawn from the frame. Traditional sampling, then, selects individuals from the sampling frame at random for participation in the study. This may not be feasible or economical as the contact information, especially email addresses, is not available for all individuals in the frame and refusals to participate increase the costs of sampling in this way.

Sample selection using the matching methodology is a two-stage process. First, a random sample is drawn from the target population. We call this sample the target sample. Details on how the target sample is drawn are provided below, but the essential idea is that this sample is a true probability sample and thus representative of the frame from which it was drawn.

Second, for each member of the target sample, we select one or more matching members from our pool of opt-in respondents. This is called the matched sample. Matching is accomplished using a large set of variables that are available in consumer and voter databases for both the target population and the opt-in panel.

The purpose of matching is to find an available respondent who is as similar as possible to the selected member of the target sample. The result is a sample of respondents who have the same measured characteristics as the target sample. Under certain conditions, described below, the matched sample will have similar properties to a true random sample. That is, the matched sample mimics the characteristics of the target sample.

When choosing the matched sample, it is necessary to find the closest matching respondent in the panel of opt-ins to each member of the target sample. YouGov employs the proximity matching method to find the closest matching respondent. For each variable used for matching, we define a distance function, d(x,y), which describes how "close" the values x and y are on a particular attribute. The overall distance between a member of the target sample and a member of the panel is a weighted sum of the individual distance functions on each attribute. The weights can be adjusted for each study based upon which variables are thought to be important for that study, though, for the most part, we have not found the matching procedure to be sensitive to small adjustments of the weights. A large weight, on the other hand, forces the algorithm toward an exact match on that dimension.