

User's Guide for the ANES 2016 Methodology Dataset

Natalya Maisel
Matthew DeBell

American National Election Studies
The University of Michigan and Stanford University
December 10, 2018

Suggested Citation

Maisel, Natalya, and Matthew DeBell. 2018. *User's Guide for the ANES 2016 Methodology Dataset*. Ann Arbor, MI and Palo Alto, CA: the University of Michigan and Stanford University.

Acknowledgments

The study was funded by the National Science Foundation with grants to the University of Michigan (grant no. SES-1444721; Principal Investigators Vincent Hutchings and Ted Brader) and Stanford University (grant no. SES-1444910; Principal Investigator Shanto Iyengar).

Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation, Stanford University, the University of Michigan, Westat, Inc., or other individuals who worked on the study. Sections of this report may reprint parts of previous documentation of the American National Election Studies without explicit attribution.

Natalya Maisel, Jaime Ventura, Patricia Luevano and Matthew DeBell prepared these methodology data for release.

See the documentation for the ANES 2016 Time Series Study for more information on the ANES 2016 Times Series Study.

Contact

The ANES website address is <http://www.electionstudies.org>

ANES sends occasional updates on Twitter @electionstudies

Any questions not answered on the ANES website or by this report may be directed to ANES staff by email at anes@electionstudies.org

Introduction and Purpose of this Dataset

This dataset provides ANES users with supplemental methodological variables for the ANES 2016 Times Series, including further information on the recruitment efforts and data collection process. This dataset includes: interviewer characteristics, final disposition data for the full sample of addresses, sample and language information, all data for the face to face (FTF) and Web screeners, all records of contact and attempted contact with addresses, mailings sent to respondents, web login data, information on weights, and the data for the Non-Response Follow-Up (NRFU) study. The previously-released 2016 Time Series survey data focused on the households with completed surveys; when relevant, the current dataset includes data for all FTF and Web addresses sampled.

ANES 2016 Methodology Dataset at a Glance

Title:	ANES 2016 Methodology Dataset
Purpose:	To provide users with supplemental methodological variables from the ANES 2016 Times Series on all sampled addresses, including interviewer characteristics, disposition data, sample and language information, screener data, records of calls, mailings sent to respondents, web login data, weighting information, and the NRFU study data.
How to use with ANES 2016 Time Series:	The data can be merged with the Time Series data on case ID, but the case ID has been updated since the release of the 2016 Time Series. In the current Methodology Dataset, variable V160001_orig can be joined to V160001 in the previously-released 2016 Time Series datasets.
# of records in this dataset:	10680 records, which represent all sampled addresses from 2016. There were 7800 Web addresses sampled and 2880 FTF addresses sampled.
# of variables in this dataset:	3651 variables. The number of variables is large given the detailed information we provide on topics such as call records, every mailing sent to addresses, and weight replicates. This user guide describes the overall categories of variables so that users can see which types of variables may be of interest.

Methodology

This dataset was generated by ANES based on the data received from Westat, the survey firm that conducted the 2016 Time Series. ANES project staff then prepared the data for release by extracting data from call logs, renaming variables for clarity, adding variable and value labels, labeling missing data, investigating any inconsistencies, creating summary variables, and joining separate files into one dataset.

The final section of this dataset contains the data from the ANES 2016 Non-Response Follow-Up Study. This study is fully described in a separate report, *Methodology of the ANES 2016 Non-Response Follow-Up Study*.

Naming conventions

We have used previous Time Series (TS) conventions for naming the case ID variables and the interviewer characteristics variables. Historically, interviewer characteristics have been released to the public when available for completed cases, and we followed those naming conventions. However, the remaining variables in this methodology dataset use variable names that are descriptive in nature. Descriptive names were used to help users make sense of the data, given the large number of variables in this release and the fact that much of this information is not commonly released from ANES.

Where possible, we used a common prefix to make it clear that a group of variables are related. For example, all FTF screener data variables begin with the prefix “ftf_screener,” all variables providing information on the letters mailed to Web respondents begin with the prefix “web_mailing,” and so on.

Subsampled cases in the FTF sample

As noted in the Methodology Report for the ANES 2016 Time Series Study, case dispositions show that 531 FTF cases were subsampled out for adaptive design, but weighting data and other field records show 527 cases subsampled out. The discrepant cases are case IDs 300084, 300981, 301585, and 302794. For these cases, the adaptive design adjustment factor was 1, indicating the weights treat these cases as having been finalized before adaptive design was implemented, but the sample disposition data indicate the cases were dropped. These four cases amount to about one seventh of one percent of the FTF sample, so this discrepancy has no material effect on the weights or response rates, but researchers may encounter some inconsistencies between the disposition codes and the field data presented in this dataset for these cases.

Restricted-use data

As with other ANES studies, most variables from the ANES 2016 Methodology Dataset are included in the public-use data file that is available to the public for free. Access to some variables is restricted to protect respondent privacy. Such data are coded - 3, “Restricted access” on the public-use file. Restricted data include geographic details about where the respondent lives and the unedited open-ended text responses. Most of these data may be obtained by following the procedures for special access described on the ANES website.

Overview of variables

Table 1 provides information on the different types of variables in this dataset and provides the position of the variables in the file. From Table 1, users can hyperlink to the description of the variables.

Table 1. Overview of variables in the ANES 2016 Methodology Dataset, including brief description and position in the data file.

Description	Variable(s)	# of Vars	Position
1. Version date	version	1	1
2. Case ID for all addresses	V160001_orig	1	2
3. Case ID for completers, CDF version	V160001	1	3
4. Interviewer characteristics (Pre)	V168250 - V169266	17	4-20
5. Interviewer characteristics (Post)	V168300 - V168316	17	21-37
6. Mode (Web vs FTF)	mode	1	38
7. Geography	state, zip, fips, censustract, censusblockgroup	5	39-43
8. Dispositions (Web)	disp_web_pre	1	44
9. Dispositions (FTF)	disp_ftf_scr - disp_ftf_post	9	45-53
10. Sampling Information	ftf_predprob, ftf_droppoint, ftf_dropcount, web_telephone_dispcode, spanishprop, hasphone	6	54-59
11. Language of interviews	pre_language, post_language	2	60-61
12. FTF screener	ftf_screener_incentiveoffer - ftf_screener_screenerselected	110	62-171
13. FTF contact information collected after Pre	ftf_contact_accept_precheck - ftf_contact_allow_text	7	172-178
14. Web screener	web_screener_completed - web_screener_screenerselected	111	179-289
15. FTF Record of Call (ROC)	ftf_roc_pre_callcount - ftf_roc_refusaldesc_78	1564	290-1853
16. Web ROC	web_roc_contactdate_1 - web_roc_isinboundyn_14	56	1854-1909
17. Web ROC codes (reasons for contact)	web_contcode_tot_cont - web_contcode_rcont4_c10	64	1910-1973
18. FTF mailings	ftf_mailing_date_1 - ftf_mailing_code_7	14	1974-1987
19. Web mailings	web_mailing_taskcode1 - web_mailing_post_count_nopay	176	1988-2163
20. Web login	web_login_useragent_1 - web_login_time_70	210	2164-2373
21. Weights	weight_varstrat - weight_hhpostpswt	1249	2374-3622
22. NRFU	version_nrfu - nrfu_selection_weight	29	3623-3651

Description of variables

The next section provides a description of the variables in this release.

1. version

Gives the version date of the current file.

2. V160001_orig

This variable contains the 6-digit IDs we received from Westat for all addresses in the sample. It was originally named `V160001` in earlier 2016 TS releases. This variable contains values which range from 300001 – 407800. Please use this variable to join previously-released 2016 TS data to the current dataset.

3. V160001

This is a new variable with 4-digit IDs for Time Series completers that has been created in order to match the Cumulative Data File. The new values range from 1 – 5090 (IDs are only for records with a completed interview).

4. Pre interviewer characteristics: V168250 - V169266

These variables provide information on the Pre interviewer descriptive characteristics. Information collected about the interviewer who conducted the Pre interview includes: skintone, age, education, gender, Spanish spoken, experience with current survey firm, race/ethnicity.

5. Post interviewer characteristics: V168300 - V168316

These variables provide information on the Post interviewer descriptive characteristics. Information collected about the interviewer who conducted the Post interview includes: skintone, age, education, gender, Spanish spoken, experience with current survey firm, race/ethnicity.

6. mode

This variable distinguishes FTF and Web for all cases in the sample. This variable expands on `V160501` in the main TS release, which was released for completed cases.

7. Geography: state, zip, fips, censustract, censusblockgroup

The variable `state` gives the state for all cases in the sample. This variable expands on `V163001a` in the main TS release, which was released for completed cases. The remaining restricted variables provide more information on the location of each address in the sample.

8. Web disposition variable: disp_web_pre

Final disposition variable for the Web Pre-election interview. This variable constitutes the final case status, or disposition, as assigned by the survey firm. In some cases the disposition may appear inconsistent with other data, but the disposition was intended to be the final judgment of the status of the case. Inconsistencies do not necessarily constitute data errors, but can reflect the process of data collection.

In particular, there are 44 cases that appear to have screener data but were given disposition codes of “Returned mail” (i.e., 52/53/54 in `disp_web_pre`). The screener data remains in this file, but users can exclude these cases using the web disposition variable (`disp_web_pre=52` or `53` or `54`) or using the final

web screener selection variable (`web_screener_selectedpre=-6`). In these cases, for example, the household may have been vacant at the start of the study and mail was returned, but before that information could be logged, additional recruitment letters were sent and the survey was completed by new residents.

9. FTF disposition variables: `disp_ftf_scr` – `disp_ftf_post`

This set of variables represent the FTF disposition variables. These are case dispositions for the face-to-face survey. We included all information provided by the survey firm, including some interim disposition variables. The main variables of interest will likely be: `disp_ftf_scr` (screener disposition), `disp_ftf_pre` (final disposition summary for the pre-election) and `disp_ftf_post` (final disposition summary for the post-election).

The numeric codes for these different variables vary from variable to variable, depending on their provenance. For example, `disp_ftf_pre` has values ranging from 11-45, which is similar to `disp_web_pre`, although the codes used for each variable were not the same; these variables acquired their codes from ANES's request for specific information on the disposition of each case from the survey firm. The remaining FTF disposition variables tend to use disposition codes preferred by the survey firm, which included some interim and some final codes. Figures 1-4 below provide more information about the meaning of different codes. Some codes in the figures below do not appear in any of the disposition variables in the dataset.

Since the release of the ANES 2016 TS data, one FTF case (`v160001_orig=302252`) has been recoded from "complete" to "ineligible." After an investigation of screener data, it was determined that this respondent was likely not a citizen and had been interviewed by mistake.

Figure 1. Disposition codes for the FTF screener, provided by the survey firm.

Screener	Interim - Assigned by FI through EROC/mROC	Final - Assigned by FS through SMS	Final - Assigned by IMS/CAPI
			100 Not Worked - Screener
			101 Breakoff - Screener
			197 Complete, SP selected - Screener
			Complete, no adult citizen -
			198 Screener
			Complete, not perm occ HH -
			199 Screener
111	No one home - Screener	151 Max Attempts - Screener	
112	Refusal - Screener	152 Refusal, pre-selection	
113	Appointment - Screener		
114	Callback - Screener		
115	Spanish FI Needed - Screener		
116	Broken Appt/No Show - Screener		
	Disabled permanently, can't interview -		
117	Screener	157 Disabled permanently, can't interview -	
118	Can't speak English/Spanish - Screener	158 Screener	
119	Unavailable Field Period - Screener	159 Can't speak English/Spanish - Screener	
131	Vacant - Screener	171 Unavailable Field Period - Screener	
132	Vacation Home - Screener	172 Vacant - Screener	
133	Not a DU - Screener	173 Vacation Home - Screener	
134	Invalid address, Other - Screener	174 Not a DU - Screener	
135	Unable to Locate - Screener	175 Invalid address, Other - Screener	
138	Unable to Access - Screener	177 Unable to Locate - Screener	
139	Multi Unit - Screener	178 Unable to Access - Screener	
		179 Multi Unit - Screener	
		180 Closed Out Adaptive Design - Screener	

Figure 2. Disposition codes for the FTF pre-election survey, provided by the survey firm.

	Interim - Assigned by FI through EROC/mROC	Final - Assigned by FS through SMS	Final - Assigned by IMS/CAPI
PRE			400 Not Worked - PRE 401 Breakoff - PRE 499 Complete - PRE
	411 SP not home - PRE	451 Max Attempts - PRE	
	412 Refusal - PRE	452 Refusal (post-selection) - PRE	
	413 Appointment - PRE		
	414 Callback - PRE		
	415 Spanish FI Needed - PRE		
	416 Broken Appt/No Show - PRE		
	417 Disabled permanently, can't interview - PRE	457 Disabled permanently, can't interview - PRE	
	418 Can't speak English/Spanish - PRE	458 Can't speak English/Spanish - PRE	
	419 Unavailable Field Period - PRE	459 Unavailable Field Period - PRE	
	421 SP Moved - PRE	461 SP Moved - PRE	
	438 Unable to Access - PRE	478 Unable to Access - PRE	
		480 Closed Out Adaptive Design - PRE	

Figure 3. Disposition codes for the FTF post-election survey, provided by the survey firm.

	Interim - Assigned by FI through EROC/mROC	Final - Assigned by FS through SMS	Final - Assigned by IMS/CAPI
POST			700 Not Worked - POST 701 Breakoff - POST 799 Complete - POST
	711 SP not home - POST 712 Refusal - POST 713 Appointment - POST 714 Callback - POST 715 Spanish FI Needed - POST 716 Broken Appt/No Show - POST	751 Max Attempts - POST 752 Refusal (post-selection) - POST	
	717 Disabled permanently, can't interview - POST	757 Disabled permanently, can't interview - POST	
	718 Can't speak English/Spanish - POST 719 Unavailable Field Period - POST	758 Can't speak English/Spanish - POST 759 Unavailable Field Period - POST	
	721 SP Moved - POST 738 Unable to Access - POST	761 SP Moved - POST 778 Unable to Access - POST	

Figure 4. Disposition codes for the FTF interviewer observations, provided by the survey firm.

	Interim - Assigned by FI through EROC/mROC	Final - Assigned by FS through SMS	Final - Assigned by IMS/CAPI
DU OBS			200 Not Worked - DU OBS 299 Complete - DU OBS
PRE OBS			500 Not Worked - PRE OBS 599 Complete - PRE OBS
POST OBS			800 Not Worked - POST OBS 899 Complete - POST OBS

10. Sampling information: **ftf_predprob, ftf_droppoint, ftf_dropcount, web_telephone_dispcode, spanishprop, hasphone**

This set of variables provides information on the sampled addresses:

- **ftf_predprob**: The predicted probability of response for every address in the FTF sample. This variable was used to determine which addresses would receive a larger starting incentive offer. Addresses with a lower predicted probability of response were randomly assigned to receive either \$25 or \$50 as an initial incentive offer. Addresses with a higher predicted probability of response were assigned \$25 as an initial incentive offer.
- **ftf_droppoint**: Information on addresses with a drop point address flag.
- **ftf_dropcount**: Provides information on the estimated number of units for addresses with a drop point address flag. Drop point and drop count variables are not relevant for the Web sample because drop point addresses were excluded from the frame before the addresses were selected for the study.
- **web_telephone_dispcode**: Information on addresses that received additional telephone contact efforts from the survey firm for the Web sample. As the election deadline approached, ANES added telephone contact as another recruitment strategy to supplement the scheduled mailings. Telephone numbers were acquired by using a commercial service to provide telephone numbers for given addresses.
- **spanishprop**: Information for Web and FTF on the areas thought to have a high proportion of Spanish speakers, as indicated from Census Bureau data. In areas with a higher proportion of Spanish speakers, bilingual mailings in English and Spanish were sent to addresses.
- **hasphone**: A frame flag for Web and FTF indicating whether a phone number was associated with the address. This variable does not correspond to addresses that received phone calls because this variable indicates that the sample frame data associated the address with a telephone number, while telephone recruitment was performed when initially non-responding addresses were later matched to telephone numbers using a different service.

11. Language: **pre_language, post_language**

Information on whether the pre-election and post-election interviews were conducted in English or Spanish.

12. FTF Screener: **ftf_screener_incentiveoffer - ftf_screener_screenerrselected**

This set of variables contains the data from the questions on the FTF household screening questionnaire.

The first two variables provide overall information on the FTF screener. The variable `ftf_screener_incentiveoffer` indicates which households received \$25 or \$50 as an initial incentive offer. The variable `ftf_screener_completed` is an indicator for whether or not the FTF screener was completed.

Please consult the FTF screener specs to better understand the variables `ftf_screener_intro` through `ftf_screener_lang`. Please note that the data for the person in position 1 of the roster always represents the screener respondent. When `ftf_screener_s2` was greater than 1, respondents completed a household roster. In the programming for the screener interview, the data for screener respondents were repeated in the roster as “person 1” variables. In other words, for respondents who needed to complete the roster, the following variables match perfectly:

- `ftf_screener_scrn_cit_s = ftf_screener_scrn_cit_1`

- `ftf_screener_scrn_sex_s = ftf_screener_scrn_sex_1`
- `ftf_screener_scrn_age_s = ftf_screener_scrn_age_1`
- `ftf_screener_scrn_age_s_x = ftf_screener_scrn_age_1_x`
- `ftf_screener_scrn_educ_s = ftf_screener_scrn_educ_1`

All screener respondents who were sent to the roster were residents, so `ftf_screener_scrn_res_1=1` for all relevant cases. Age was collected for all screener respondents, so `ftf_screener_confirm2_1` and `ftf_screener_elig_ckage_1` were not relevant. The labels for the roster person 1 variables indicate that this is repeated screener data, so that users do not think that person 1 should be counted in addition to the screener respondent.

We have created several summary variables to try to ease in interpretation of the selection process for the pre-election interview:

- **`ftf_screener_numadults_final`**: The number of adults in the household (updated after the roster).
- **`ftf_screener_numelig_final`**: The number of eligible U.S. Citizen adults in the household (updated after the roster).
- **`ftf_screener_eligselect_rost`, `ftf_screener_selected_rostpos`, and `ftf_screener_finalrosterpos`**: We have three variables with information on which person was selected from the roster to complete the Pre interview. Please use `ftf_screener_finalrosterpos` as the final, definitive selection of the respondent from the roster. Here is a brief description of the differences between these three variables:
 - o **`ftf_screener_eligselect_rost`**: The selection for the Nth citizen listed in the roster. This variable is in the original format received from the survey firm.
 - o **`ftf_screener_selected_rostpos`**: The selection for the Nth person listed in the roster (citizen or not). This variable is in the original format received from the survey firm. This variable had 48 cases where it seemed that no roster selection happened, but indeed there was some selection that took place. In those cases, there was only 1 eligible person in the roster and that person was the screener respondent. For those cases, the selection of "Person 1 (Screener R)" was added to the variable `ftf_screener_finalrosterpos`.
 - o **`ftf_screener_finalrosterpos`**: Uses the information from `ftf_screener_selected_rostpos` but includes updates for the 48 cases described above, and includes a code for households where there was no eligible person after the roster. Please use this variable for analyses involving selection from the roster.
- **`ftf_screener_rostersize`**: The number of people listed on the roster (whether eligible or not).
- **`ftf_screener_selectedpre`**: A summary variable created to indicate the selection to the Pre for all cases. This variable combines information on the person selected from the roster with selection information for cases where no roster was needed. It also indicates whether random selection was used or if the person selected was the only eligible person.
- **`ftf_screener_screenerselected`**: This variable indicates whether the screener respondent was the person selected.

Below is a general outline for the selection process from the FTF screener to the pre-election interview:

- (1) 1487 respondents started the FTF screener
- (2) 440 cases where screener respondent was only adult in the household
(`ftf_screener_s2=1`)
 - a. 19 of these were non-citizens; marked as `ftf_screener_noeligible=1`
 - b. 4 screener respondents refused citizenship question; marked as `ftf_screener_noeligible=1`
 - c. Therefore 417 screener respondents were selected for the Pre without a roster needed
- (3) 1047 screener respondents continued to the roster
 - a. We originally had roster position data on 934 selections (`ftf_screener_selected_rostpos`).
 - b. For the remaining 113 cases, the following occurred: As noted above, for 48 of these cases, there was only 1 eligible person in the roster and that person was the screener respondent; for those cases, no value was added to the variable `ftf_screener_selected_rostpos` (this has been updated in `ftf_screener_finalrosterpos`). The remaining 65 cases were households where no one was eligible after the roster. In 63 of these cases, there were no U.S. citizens. In two cases, there was one U.S. Citizen but the person was not 18.
- (4) 1399 households had an eligible person selected to complete the pre-election interview. The screener respondent was selected for the pre-election interview in 886 cases and another person in the household was selected for the pre-election interview in 513 cases.

13. FTF contact information given: `ftf_contact_accept_precheck` - `ftf_contact_allow_text`

This set of variables presents some information on the type of data that FTF respondents allowed to be collected after the pre-election interview. These variables indicate whether or not the respondent accepted the check payment (`ftf_contact_accept_precheck`), if the respondent gave a name for the check (`ftf_contact_givename`), if the payment had to be given in cash (`ftf_contact_paymenttype`), if the respondent gave a phone number for follow-up (`ftf_contact_give_phone`) and whether this phone was a cell phone (`ftf_contact_cell_phone`) and whether the respondent would allow texts to this phone (`ftf_contact_allow_text`).

14. Web Screener: `web_screener_completed` - `web_screener_screenerselected`

This set of variables contains the data from the questions on the Web screening questionnaire.

The first two variables provide overall information on the Web screener. The variable `web_screener_completed` is an indicator for whether or not the Web screener was completed. The variable `web_screener_checkamount` indicates when checks were issued to the screener respondent who was *not* selected for the pre-election interview. As a reminder of the Web study design: When screener respondents were not selected for the pre-election interview, they still completed a “mini-survey” that consisted of approximately 20 ANES questions (e.g., party ID, education, race/ethnicity), and they were asked for information on the person who was selected for the pre-election interview. In these cases, the screener respondent still received the full promised payment, and the variable `web_screener_checkamount` indicates the amount paid in these situations (\$40 versus \$80, depending on whether the offer had been escalated).

Please consult the Web screener specs to better understand the variables `web_screener_s3` through `web_screener_s_age_10_s103_x`. Please note that the data for the person in position 1 of the roster differs somewhat from the FTF screener. For the Web screener, the person in position 1 of the roster is not always the screener respondent. There were 16 cases in the Web where the person completing the screener was under 18 years old or did not provide age data (`web_screener_s8=2` or `web_screener_s8=-7`). In those cases, the first person listed by the screener respondent was entered as person 1 on the roster. For the cases where the screener respondent was over 18, the screener respondent data for gender and citizenship was re-populated into the household roster as person 1. Therefore, the person 1 variables include a mixture of mostly screener respondents and some non-screener respondents.

In general, we have used “-5. Inapplicable question” where it seems that roster data was not needed, but please note there may be some instances where data should have been provided and it was not.

Please see the notes for the variable `disp_web_pre` for more information on when the Web screener data may sometimes seem to conflict with the final disposition codes.

We have created several summary variables to try to ease in interpretation of the selection process for the pre-election interview.

- **Number of adults in the household:** We recorded the number of adults listed before the roster (`web_screener_numadults_prerost`) and the number of adults listed after the roster (`web_screener_numadults_final`). There turned out to be no difference between these two variables.
- **Number of eligible (i.e., U.S. Citizen) adults in the household:** We recorded the number of eligible adults from the roster (`web_screener_numelig_rost`) and then combined that with information on the screener respondent's citizenship to create a final count of the number of eligible U.S. Citizen adults in the household (`web_screener_numelig_final`).
- **`web_screener_eligselect_prerost`, `web_screener_eligselect_rost`, `web_screener_selected_rostpos`, and `web_screener_finalrosterpos`:** We have four variables with information on which person was selected from the roster to complete the Pre interview. Please use `web_screener_finalrosterpos` as the final, definitive selection of the respondent from the roster. Here is a brief description of the differences between these four variables:
 - o **`web_screener_eligselect_prerost`:** Identifies the person selected for the pre-election interview before the roster. Unlike in the FTF, an initial selection happened before the roster. If the screener respondent was selected (random number `selection=1`), no household roster was completed.
 - o **`web_screener_eligselect_rost`:** The selection for the Nth citizen listed in the roster. This variable is in the original format received from the survey firm.
 - o **`web_screener_selected_rostpos`:** The selection for the Nth person listed in the roster (citizen or not). This variable is in the original format received from the survey firm.
 - o **`web_screener_finalrosterpos`:** Uses the information from `web_screener_selected_rostpos` but recodes cases that did not proceed to the Pre due to a disposition code of “Returned mail” (coded as `web_screener_finalrosterpos=-4`) as well as other reasons for not proceeding to the Pre (respondent does not live at address; screened, not a household; incomplete

screeener). Please use this variable for analyses involving selection from the roster, but keep in mind that not all households were rostered.

- **web_screener_rostersize**: The number of people listed on the roster (whether eligible or not).
- **web_screener_selectedpre**: A summary variable created to indicate the selection to the Pre for all cases. This variable combines information on the person selected from the roster with selection information for cases where no roster was needed. It also indicates whether random selection was used or if the person selected was the only eligible person. As with the variable `web_screener_finalrosterpos`, this variable recodes cases that did not proceed to the Pre due to a disposition code of "Returned mail" (coded as `web_screener_selectedpre=6`) as well as other reasons for not proceeding to the Pre (respondent does not live at address; screened, not a household; incomplete screener).
- **web_screener_screenerselected**: This variable indicates whether the screener respondent was the person selected.

Below is a general outline for the selection process from the Web screener to the pre-election interview:

- 3779 households started the screener or had a disposition code of "Incomplete screener"
 - 54 households had "30. Incomplete screener" disposition code (`disp_web_pre`)
 - 44 households had codes 52/53/54 Returned mail disposition code
 - 3681 started the screener & did not have "Incomplete" code or "Returned mail" code
 - o 35 people did not live at the address (`web_screener_s3=2`); screener stopped [Note: 9 additional households had `web_screener_s3=2` but Returned mail/Incomplete dispo code took precedence]
 - o 3646 respondents had values for `web_screener_s3` of "1=I live in this address" (n=3631) or "-7=No data" (n=15). These cases moved forward with the screener. [Note: 42 additional households had `web_screener_s3=1` or -7 but Returned mail/Incomplete dispo code took precedence]
 - 12 cases coded as "Not a household" [`web_screener_S5a==2` | `web_screener_S5a==3`]
 - 65 households with no eligible adults
 - 62 households had no eligible adults before the roster → no roster completed
 - 3 additional households had no eligible adults after the roster
 - 3569 households with at least 1 person eligible for Pre
 - 916 households → Only 1 person in household; Screener R is adult U.S. Citizen → Screener R automatically selected for Pre; no roster needed
 - 83 households → More than 1 adult in household, but only 1 person is Citizen and Screener R is adult U.S. Citizen → Screener R automatically selected for Pre; no roster needed
 - 2 cases had slightly different logic than usual (`web_screener_5>1`, `web_screener_numelig_prerost=1`, `web_screener_s6a=1`, `web_screener_s8=1`), and screener R was selected without roster
 - 1174 households → More than 1 U.S. citizen adult in household and screener R is one of them; random number generated=1 so that screener R was randomly selected → Screener R automatically selected for Pre; no roster needed

- If the case was sent to Roster, Screener R normally should not be selected – because if “1” had been the random number, no roster would have been needed. The only time Person 1 in the roster had a chance of selection would be in the 16 cases where the screener R was not 18 years old or had no age data, and therefore Person 1 represented someone else other than the Screener R.
 - 8 households - Person 1 (not screener R) selected
 - 1137 households - Person 2 selected
 - 184 households - Person 3 selected
 - 48 households - Person 4 selected
 - 14 households - Person 5 selected
 - 3 households - Person 6 selected

15. FTF Record of Call (ROC) variables: `ftf_roc_pre_callcount` - `ftf_roc_refusaldesc_78`

This set of 1,564 variables provides all the information available on the attempted contacts with addresses in the FTF. For each contact or attempt, a range of information was collected, including the date, status code, person contacted, etc. The variable suffixes range from “_1” for the first call through “_78” for the one case that had 78 records of calls. The number of calls reported here was higher than we expected, and an investigation of the calls revealed instances of duplicate or potentially-duplicate entries. This might have occurred when interviewers entered a contact attempt two or more times, possibly due to not being clear on whether the call had been recorded or maybe not remembering if they had logged it already. After an investigation by ANES project staff, we determined that as a practical matter there was not enough information to determine efficiently which records were duplicated, because records that appeared similar but might have slight difference due to duplication or due to legitimately similar call details. For example, there was a case of two sequential records, recorded 10 minutes apart (based on the automatic time stamp for entering a record), where the interviewer had recorded the exact same date and time of the contact with the respondent except that one entry had the time recorded as AM and one had it recorded as PM. From examining the comments section for that case, it seemed that both entries represented the same visit and that the initial entry was likely an error, but it would have been hard to categorize this as a duplicated entry based solely on an algorithm to compare records.

Given this issue, we wanted to find a way to estimate the number of calls without having to manually search through the records for duplicates, which would have been labor intensive and also open to subjective interpretations that would have made the process unreliable. We examined whether using the number of unique days when calls occurred would be a more reliable method. An ANES staff member examined 75 randomly selected cases. She examined calls with a valid status code (in this case, we excluded codes that indicated a letter was sent and codes for final dispositions assigned by the home office) and conducted a visual call review to determine unique calls. Compared to her visual call review, taking a sum of all call records overestimated the call count by 25.9%. Alternatively, compared to her visual call review, taking a sum of unique days when a call occurred only underestimated the call count by 0.9%.

Based on this information, we decided to use the count of unique days with a valid call record as our metric for creating a final call count. Although we lose some information this way (e.g., cases where an interviewer visited the house twice in one day to try to reach someone), it is still likely to work as an approximation of the effort expended on each case. We found a correlation of $r=0.94$ between the call

count using every record with a valid status code and the call count using unique days with a valid status code.

Below is more information on the record of call variables. ROC data are operational field records used for purposes of case management and have some gaps.

- **ftf_roc_pre_callcount**: As outlined above, this is an estimated count of the number of calls for the pre-election interview, which sums the number of unique dates with a valid call. To determine a “valid” call, we used the status code of the call record, found in variables `ftf_roc_statuscode_1` - `ftf_roc_statuscode_78`. “Valid” calls had the following range of values: 111-139, 197-198, 401-438, 499. Please see Figures 1-2 for more information on these codes. The “final” codes assigned by the home office were not counted, as these would duplicate the interim codes that were assigned in the field by interviewers. The calls counted in this variable include in-person, telephone, and text.
- **ftf_roc_post_callcount**: This replicates the procedure followed in the creation of `ftf_roc_pre_callcount` for the post-election interview. The “valid” calls for the post-election interview had the following range of values: 701-739, 799. Please see Figure 3 for more information on these codes.
- **ftf_roc_pre_callcount_inperson**: The previous variables counted all types of calls, whether they happened in-person or via telephone or text. We also wanted to examine in-person calls in particular, as these require the most effort and cost. For this pre-election variable, we restricted the calls counted in `ftf_roc_pre_callcount` to those that occurred in-person with the use of the variables `ftf_roc_contacttype_1`-`ftf_roc_contacttype_78`. There were some call records that were missing information on contact type. After some investigation, there seemed to be a pattern to the missing data. For example, codes associated with survey completion and survey breakoff were not generally coded with a contact type, but given the nature of these codes, we could assume that the contact was in-person. This variable therefore provides an estimated count of the number of days with a valid in-person visit.
- **ftf_roc_post_callcount_inperson**: The procedure for `ftf_roc_pre_callcount_inperson` was repeated for the post-election interview.
- **ftf_roc_dt_rec_1-78**: Date record was created.
- **ftf_roc_tm_rec_1-78**: Time record was created.
- **ftf_roc_taskid_1-78**: Type of task (e.g., screener, Pre, Post).
- **ftf_roc_actiontype_1-78**: Type of action (e.g., assign, transfer, add EROC).
- **ftf_roc_statuscode_1-78**: Status code variables provide a descriptive code for the call that occurred, such as if the interviewer reported that no one was home, that the house was vacant, that the respondent refused, etc. Please see Figures 1-4 for more information on these codes.
- **ftf_roc_suid_1-78**: An ID generated by the survey firm.
- **ftf_roc_idseries_1-78**: Source of record (e.g., field, home office, mobile device).
- **ftf_roc_activlogid_1-78**: Activity log values generated by the survey firm.
- **ftf_roc_completedby_1-78**: ID number of person completing the record.
- **ftf_roc_anesuserid_1-78**: ANES ID of person completing the record – very similar to “completedby” variables.
- **ftf_roc_dt_contact_1-78**: Date of attempted contact with respondent.
- **ftf_roc_tm_contact_1-78**: Time of attempted contact with respondent.

- **ftf_roc_contacttype_1-78**: Type of attempted contact (e.g., in person, telephone, text).
- **ftf_roc_statussourceid_1-78**: Source of the status code (e.g., electronic call record, survey instrument, supervisor management system).
- **ftf_roc_whocontacted_1-78**: Who was contacted (e.g., no one, selected person, household member, neighbor, community contact, other).
- **ftf_roc_appointment_dttm_1-78**: If a future appointment was scheduled, date and time of this appointment.
- **ftf_roc_restrictaccesstype_1-78**: Type of restricted access (e.g., gated community, locked apartment complex, other-inaccessible housing unit).
- **ftf_roc_duaddressstype1-78**: Type of dwelling if not a household (e.g., institutional or group quarters, business, vacant, other).
- **ftf_roc_refbreakoffdesc_1-78**: Type of refusal – hard or soft.
- **ftf_roc_refusaldesc_1-78**: More detailed information on refusal (e.g., hard-do not contact, hard-hostile/threatening, soft-not interested, soft-privacy concerns, soft-too busy).

16. Web Record of Contacts (ROC): web_roc_contactdate_1 - web_roc_isinboundyn_14

The Web ROC variables have a different meaning than the FTF ROC variables. The Web design was such that recruitment efforts occurred mainly through the mail (with some email and phone efforts as well). These efforts are documented in “web_mailing” variables below.

For the Web ROC, these variables recorded instances when respondents contacted the survey firm and when the survey firm responded to those contacts. As can be seen in the next set of variables (**web_contcode_tot_cont-web_contcode_rcont10_c4**), these contacts pertained to issues such as when the respondent was having trouble accessing a computer, when the respondent was having trouble with login/PIN/website access, when the respondent wanted to check on the status of the payment, and other reasons. The Web ROC variables give an overview of all the contact that happened between the respondent and the survey firm, and the next set of variables (**web_contcode_tot_cont-web_contcode_rcont10_c4**) provides codes for the reason(s) why the respondent contacted the company.

- **web_roc_contactdate_1-14**: Date of contact between respondent and survey firm.
- **web_roc_endtime_1-14**: Time contact was logged by the survey firm.
- **web_roc_typeofcontact_1-14**: Type of contact (email, phone, mail, fax).
- **web_roc_isinboundyn_1-14**: Whether contact was initiated by respondent or survey firm.

17. Web Coding of Contact: web_contcode_tot_cont - web_contcode_rcont4_c10

As noted above, this set of variables describes the reasons that the respondent contacted the survey firm. This set expands upon the call logs in the previous section by coding the content of the respondent’s inquiry – the survey firm recorded all emails, phone messages, summaries of phone calls, letters received, etc. Although the verbatim content is restricted to ANES staff to protect respondent confidentiality, the survey firm staff indicated the type of contact by assigning a contact code.

Because we were interested in the reasons why respondents were contacting the survey firm, the follow-up responses from the survey firm were not coded (e.g., when the survey firm would call back to provide additional instructions on how to log on to the survey). Telephone calls were the most common way for respondents to contact the survey firm, followed by email and finally mail. Multiple issues might be addressed in a single contact (e.g., respondent calls to ask if she is eligible to take the survey and she lost her PIN number and she is not sure how to access the survey online). Therefore, each contact could be coded with up to 4 codes.

- **web_contcode_tot_cont**: Total number of contacts that the respondent made to the survey firm in all modes of contact.
- **web_contcode_tot_email**: Total number of email contacts that the respondent made to the survey firm.
- **web_contcode_tot_mail**: Total number of mail contacts that the respondent made to the survey firm.
- **web_contcode_tot_phone**: Total number of phone contacts that the respondent made to the survey firm.
- **web_contcode_rcont1_typ (1-10)**: Type of contact (e.g., email, mail, phone).
- **web_contcode_rcont1_dt (1-10)**: Date of the contact.
- **web_contcode_rcont1_c1 (1-10)**: The first code given to the content of the contact.
- **web_contcode_rcont1_c2 (1-10)**: The second code given to the content of the contact.
- **web_contcode_rcont1_c3 (1-10)**: The third code given to the content of the contact.
- **web_contcode_rcont1_c4 (1-10)**: The fourth code given to the content of the contact.

18. FTF mailings: ftf_mailing_date_1 - ftf_mailing_code_7

The FTF design used a variety of letters to attempt to make contact or persuade respondents to cooperate. Please see the Methodology Report for the ANES 2016 Time Series Study for more information on this strategy, such as Table 6-2, which describes the letters mailed to respondents. The variables **ftf_mailing_date_1-ftf_mailing_code_7** provide data on the letters sent to respondents.

- **ftf_mailing_date_1-7**: Date the letter was sent.
- **ftf_mailing_code_1-7**: Letter code. The “letter type” in the label for each code corresponds to the information in Table 6-2 in the Methodology Report for the ANES 2016 Time Series Study.

19. Web mailings: web_mailing_taskcode1 - web_mailing_post_count_nopay

The Web recruitment strategy focused mainly on mailings. For more information on this strategy, please see the Methodology Report for the ANES 2016 Time Series Study, such as Table 5-1 which describes the letter protocol. We have two main sets of variables that address the same information in slightly different ways.

The variables **web_mailing_taskcode1** through **web_mailing_prepost_26** give information on the mailings that were sent to Web respondents and also include other information the survey firm recorded, such as when mail was returned and when the address status of a household was changed. These variables are not cleaned, but they allow us to present all data that we received on the mailings.

- **web_mailing_taskcode_1-26**: Information on letters sent to respondents as well as returned mail. See Table 5-1 in the Methodology Report for the ANES 2016 Time Series Study for the letter protocol.
- **web_mailing_taskstatus_1-26**: Status of the mailing (e.g., completed, no such address).
- **web_mailing_taskdt_1-26**: Date the status was assigned.
- **web_mailing_adresstatus_1-26**: Status of the address (e.g., active, ineligible).
- **web_mailing_prepost_1-26**: Mailing for recruitment to pre-election or post-election survey.

The variables `web_mailing_letterdate_pre_1` through `web_mailing_lettercode_post_9` provide a cleaner version of this information that focuses only on the mailings sent to respondents. The “lettercode” variables will make it easier for users to ascertain which mailings were sent to each address. But in general, the variables listed below contain the same underlying information as the variables listed above.

- **web_mailing_letterdate_pre_1-12**: Date letter was sent for pre-election recruitment. Again, the focus here is only on the letters sent to respondents, without any additional information on returned mail or address status changes.
- **web_mailing_lettercode_pre_1-12**: Code assigned to each letter sent for pre-election recruitment. These codes match the information in the taskcode variables above, but they are numeric variables and only focus on mailings sent to the addresses, so they are likely easier to use for analytic purposes. Some numeric codes represented more than 1 type of mailing, as noted in the value labels.
- **web_mailing_letterdate_post_1-9**: Date letter was sent for post-election recruitment.
- **web_mailing_lettercode_post_1-9**: Code assigned to each letter sent for post-election recruitment.

Finally, we calculated summary variables on the number of letters sent to addresses.

- **web_mailing_pre_count_all**: For the Pre-interview, the count of all letters and emails sent to the address, including mailings with the payment for completing the screener or survey. This focuses on the following codes: 01,12,13,14,20,21,23,25,26,30,32,33,41,43,44,50,91.
- **web_mailing_post_count_all**: For the Post-interview, the count of all letters and emails sent to the address, including mailings with the payment for completing the survey. This focuses on the following codes: 34,35,36,37,38,39,40.
- **web_mailing_pre_count_nopay**: For the Pre-interview, the count of all letters and emails sent to the address, *excluding* mailings with the payment for completing the screener or survey. This variables drops code 33 from the count in `web_mailing_pre_count_all`.
- **web_mailing_post_count_nopay**: For the Post-interview, the count of all letters and emails sent to the address, *excluding* mailings with the payment for completing the survey. This variables drops code 40 from the count in `web_mailing_post_count_all`.

20. Web login information: `web_login_useragent_1` - `web_login_time_70`

This set of variables provides a record of how many times households logged onto the Web survey. The date and time of each login is provided for public use and user agent information is available for restricted use.

- **web_login_useragent_1-70**: User agent for the login (e.g., browser type). Restricted use due to the details of the user agent information.
- **web_login_date_1-70**: Date of the login for the web survey.
- **web_login_time_1-70**: Time of the login for the web survey.

21. Weight information: weight_varstrat - weight_hhpostpswt

In this release, we are providing all the weight variables we received from the survey firm. These variables are presented as they were received. Several of the variables included in this section duplicate or are similar to other variables in the file (e.g., `weight_hasphone` matches our earlier variable `hasphone`). However, users interested in weighting procedures may want to see the kinds of information considered in the construction of the weights. Additionally, users interested in conducting analyses using the jackknife replication method may appreciate the provision of the replicate weights. Unlike other parts of this file, the weight variables do include some system-missing values.

22. Non-response follow-up (NRFU) study: version_nrfu - nrfu_selection_weight

The NRFU study was conducted to collect information about survey non-respondents. Please see the Methodology of the ANES 2016 Non-Response Follow-Up Study.

The NRFU variables first include information on the inclusion of the case in the NRFU study (`nrfu_flag`), the experimental group (`nrfu_group`), the disposition of the address (`nrfu_dispo`), whether or not we received the NRFU questionnaire after the close of the study (`nrfu_latepnd`), the person selected to complete the NRFU (`nrfu_selection`), and the number of the NRFU mailing that resulted in the completed questionnaire (`nrfu_mailnum`). Next, all of the questionnaire responses are included (`nrfu_q1 - nrfu_q15`). Finally, `nrfu_mainstatus` indicates the response status for the 2016 Times Series, and `nrfu_weight` provides the weight for the NRFU data accounting for selection and non-response (`nrfu_selection_weight` provides the weight accounting only for selection probability).