

# **SURVEY METHODS**

## **About the survey**

In the spring of 2015, the Center for Survey Research at the University of Virginia entered into an agreement with Foothills Forum, a nonprofit, nonpartisan group of citizens in Rappahannock County, Virginia, to design and then conduct a mail-out survey of households in Rappahannock County. Foothills Forum is currently organized as a 501(c)3 non-profit. Foothills Forum was represented by its chairman, Larry “Bud” Meyer, and a survey committee charged with assisting in the development of the questionnaire. The goal of the survey was to determine citizen opinion on issues important to them regarding life in the County.

## **Questionnaire development**

Beginning in January, 2015, the staff at the Center for Survey Research and the survey committee for Foothills Forum discussed the aims of the survey, based on an initial conceptual outline formulated by Foothills Forum. Foothills Forum also conducted a series of focus groups, not designed or assisted by the Center for Survey Research, in order to help them clarify issues to be included in the questionnaire.

A preliminary questionnaire was developed by August, 2015 and was pretested at a focus group in Washington, Virginia, held on September 15, 2015. Foothills Forum was responsible for the recruitment of volunteers for the focus group and for arrangements and set-up of the meeting. The group was facilitated by Kathryn Wood, assisted by Matthew Braswell, who served as recorder. As a result of the focus group, significant modifications were made to the questionnaire.

The final questionnaire was approved by the Foothills Forum survey committee on October 9, 2015 and was submitted for review and approval by the University of Virginia’s Institutional Review Board for the Social and Behavioral Sciences. The project, IRB #2015-0397-00, was approved on October 13, 2015. The questionnaire was only available in English language; translation to Spanish was not justified due to the small size of the Hispanic and immigrant population in Rappahannock and the even smaller number of non-English speakers as reported in the American Community Survey’s five-year estimates for the county.

## **Sample design**

After consideration of several possible random sample designs in consultation with CSR staff, Foothills Forum decided it would be best to attempt to contact all households in Rappahannock County, instead of drawing a random sample. Since some households in Rappahannock do not receive their mail at a home mailbox, the sampling frame was defined as all residential household addresses in the County and all active, residential post office boxes in the County. Accordingly, CSR obtained a full list of residential addresses and active residential P.O. boxes from Marketing Systems Group, an industry leader in providing address-based samples for scientific surveys. The list provided by M-S-G is updated quarterly by the US Postal Service, and represents the complete list of postal delivery points in the County. We excluded from the list addresses marked

as vacant in the post office files, but seasonal homes were included in the listing. At CSR's request, M-S-G appended the names of residents with the address, to the extent that the names are known to them. (The Postal Service does not release names with its address listings; these are added in by the sampling company based on white pages phone directory listings as well as lists from proprietary databases.)

The P.O. Box listings for Rappahannock do not include information on whether or not the post office box is the only way for the household to receive mail. Since some households have a home mailbox as well as a P.O. Box, this design meant that some unknown number of households might receive two survey requests, one sent to their residential address and one to their P.O. Box. We addressed this issue by including a statement in the survey cover letter that each household was requested to respond only once. (As will be seen below, we also adjusted our estimate of the response rate by estimating the number of P.O. Box addresses that were redundant with residential addresses.)

Our initial mailing list included 3,874 addresses, of which only 492 lacked a resident's name. Of the entire list, 1126 (29%) were post office boxes. Technically, this complete list or sampling frame should not be considered a sample but a list of the entire population of eligible households in Rappahannock County.

## **Survey launch and production**

The project design called for an advance postcard, an initial questionnaire mailing, a thank-you/reminder postcard and a second mailing of the questionnaire. All mailings used USPS First Class mail, with address correction and forwarding requested. The survey was publicized in the *Rappahannock News* around the time of the launch. The initial postcard was mailed during the week of October 19, 2015, with the questionnaire following one week later, during the week of October 26. These mailings were addressed to the named household member (except for those lacking a name listing, which were addressed to "current resident"). The cover letters for the questionnaire mailings were personalized with the recipient's name (if available) and address. The reminder postcard was mailed on November 5, 2015. The purpose of the postcard was to thank those respondents who returned a completed questionnaire and to encourage potential respondents who had not yet done so to participate in the survey. The second questionnaire to non-respondents was mailed on November 23. Cases which had resulted in a returned, undeliverable postcard or questionnaire mailing were excluded from the second mailing. The second questionnaire mailing was addressed to "[name] or current resident." We announced that formal data collection would close on December 11, but were able to incorporate all questionnaires received by January 8, 2016.

The survey was designed to use a confidential protocol, meaning that the identity of the respondent would be known to CSR, but not disclosed to Foothills Forum or anyone else. An ID number was affixed to the front page of each outgoing questionnaire and to each address label so that CSR could track who had responded and who had not responded. A few respondents (44 cases) returned their questionnaires with the label removed; these were accepted and entered into the results file along with all the others. Of course, in these cases the respondent was sent the follow-up mailing aimed at non-respondents, since there was no way for CSR to identify who had sent back the questionnaire.

A number of packets were returned undeliverable due to incorrect addresses or the addressed person not being at the address. The mailings for the first and second survey packets resulted in 439 packets returned as undeliverable.

After launch, CSR received inquiries from a small number of Rappahannock residents who heard of the survey effort but never received the questionnaire. CSR sent questionnaires to these households and some of them responded to the survey, resulting in an addition of 8 households to the final sample list.

## Survey response

Of the 3882 households in our final sampling frame, 1117 questionnaires were completed from the first mailing. There were 439 returned because of bad mail addresses, while 13 were deemed ineligible because they were outside the Rappahannock County borders. The second questionnaire mailing yielded an additional 245 completions, for a total of 1,362 returned questionnaires received by the closing date for data entry, January 8<sup>th</sup>. (An additional 26 questionnaires trickled in after that date, too late for inclusion in the survey tabulation.) Only eight explicit refusals were received, and three cases in the sample list were determined to be duplicate listings.

### FOOTHILLS FORUM FINAL DISPOSITION REPORT

FINAL DISPOSITION	COUNT	OVERALL	QUALIFIED/ REACHABLE	Adjusted count	Adjusted %	% of returns
Complete	1362	35.1%	41.5%	1362	41.8%	98.5%
Refusal	8	0.2%	<0.1%	8	<0.1%	0.6%
Not eligible	13	0.3%				0.9%
Bad Mail	439	11.3%				
Duplicate household address	3	0.1%				0.2%
Estimated redundant PO boxes	148	3.8%				
Open Status	1909	49.2%	58.2%	1888	57.9%	
<b>Total</b>	<b>3882</b>	<b>100.0%</b>	<b>100.0%</b>	<b>3258</b>	<b>100.0%</b>	<b>1383</b>

The table above shows the final disposition for all cases in the sample list. The 1,362 completed questionnaires constitute 35.1 percent of the sample list, or 41.5 percent of the qualified and reachable households. Looking at the returned mail only, 1.1% of the cases were not eligible or represented duplicates. Accordingly, we can estimate that the same percentage of out-of-sample cases exists among the 2,057 'open' cases that we never heard from, dropping that number to 2033. Additionally, because 7.3% of respondents indicated that they receive mail at both their residence and at a post office box, we can estimate that the same is true for 7.3% of the 2,033 open cases presumed eligible. These cases can be dropped from the adjusted count, because the potential respondents associated with them are already included in the sample list on the basis of their residential address.

Based upon these estimates, the number of unique, reachable, qualified household that we did not hear back from is 1,888, yielding an adjusted total frame size of 3,258. (This figure is nearly identical to the US Census Bureau's estimate that there were 3,281 households in Rappahannock, as reported in the American Community Survey five-year estimates for 2014.) Our 1,362

completes thus represent a response rate of 41.8 percent. (This calculation is equivalent to the American Association for Public Opinion Research standard response rate definition RR3, a response rate in which the households of unknown eligibility are adjusted using an estimate of how many of those households are eligible.)<sup>1</sup> In other words, this survey is based on responses from about 42 percent of the occupied households in Rappahannock County.

## Data entry and data cleaning

The paper survey responses were entered into a computer data base by trained CSR staff members, using a nearly identical instrument set up in Qualtrics, a web-based survey tool licensed to CSR. Questionnaires were sorted by respondent number to identify duplicate responses (which usually occurred when a household returned their questionnaire from both the first and second questionnaire packet.) In these cases, data were entered from the earliest of the two returns. In the rare case where a respondent entered more than one answer on a forced-choice item, a coin flip was used to randomly select which of the chosen categories would be entered. A coin flip was also used to select the respondent when a questionnaire had been filled out by household members to represent responses or demographics for two members of the household on the same questionnaire. Data entry included entry of all the numeric responses (and check-boxes represented as numbers for analysis) as well as verbatim entry of text responses to open-ended survey questions. Once all questionnaires were entered into Qualtrics, the data were moved to the IBM SPSS statistical analysis program. CSR analysts edited the variable labels and value labels in the data set, designated missing value indicators, and reversed the coding of some of the response scales so that higher numbers would represent more favorable orientations, which greatly facilitates interpretation of results. Due to sample weighting and the fact that a small number of returned cases were excluded from analysis because they had too many missing questions, the final, weighted sample size for analysis was 1,346 cases.

Thirteen percent of the entered questionnaires were selected for validation: supervisors compared the original paper questionnaire with the data entered into Qualtrics. Any errors found were corrected in the SPSS dataset, however no systematic errors were evident. A preliminary set of weighted frequencies and means on scaled variables was prepared for Foothills Forum in mid-January, 2016.

Text responses to open-ended questions on the survey were moved from Qualtrics into an Excel workbook. CSR lab supervisor Beverley Kerr read through every response and edited out any personal references or references to specific business locations; she also cleaned up some spelling errors in these entries. The responses were then alphabetized (separately for each question) and listed verbatim in a separately bound appendix to the survey report (Appendix E).

## Weighting and analysis

The data were weighted to ensure accurate representation of the County population. This process, known as post-stratification weighting, adjusts the sample proportions to match known characteristics of the population under study. We drew our control totals (target percentages) from the 2013 and 2014 five-year estimates for Rappahannock County derived from the results of

---

<sup>1</sup> American Association for Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Eighth edition, 2015. Available at [www.aapor.org](http://www.aapor.org).

the American Community Survey and published by the US Bureau of the Census. The weight variables included homeownership, single person households, zip-code, gender, married persons, race (African-American/all other), and respondent age (four categories). CSR used iterative raking weighting (also known as ‘raking’), to estimate the weights. Raking was implemented by CSR using a proprietary program that links Python macros with SPSS. The weighted data match US Census Bureau data for Rappahannock County on these characteristics. For example, in the unweighted survey results women constitute 51.4 percent of the respondents, while after weighting women are 49.6 percent of the respondents, matching the 2014 ACS five-year estimate of the female percentage among Rappahannock adults (18 and older). Unweighted frequencies for the demographic variables are included at the end of the weighted frequencies in Appendix B.

Standard statistical tests for survey data often assume that sampling was done by simple random sampling. In this study, a fairly large proportion (42%) of the study population ended up in the final sample, a fact which lowers the standard error of the survey estimates (the ‘margin of error’) based on a statistical factor known as the ‘finite population correction<sup>2</sup>.’ In addition, the use of post-stratification weights, which act to reduce possible non-response bias in the results, brings with it a ‘design effect’ that increases the standard error of the estimates. Therefore, CSR used the Complex Sampling module of procedures in SPSS for all tests of statistical significance in this report. The tests used were independent sample T-tests available through the complex sampling GLM procedure in SPSS, and the confidence level for all tests was set at the 95% level. The T-test results reported in the mean-comparison and percentage-comparison tables in Appendix C are based on this procedure, which calculates the applicable sampling error for each question while taking into account the finite population size and the design effect from weighting.

## Margin of Error

The margin of error due to sampling for the survey is approximately  $\pm 3.3$  percent at the 95 percent level of confidence. This means that if the survey were to be repeated 100 times with a random 41.8% of the households responding each time, the results of this survey would be within 3.3 percentage points of the population percentage 95 out of those 100 iterations of the survey. Note that there are other sources of error in surveys besides sampling error that can be difficult or impossible to measure, including non-response error and measurement error. However, due to the comprehensive nature of the household listing used in this study, coverage error (meaning failure to include eligible units) is negligible in this study.

The margin of error is affected by the large number of respondents relative to the Rappahannock population (42 percent of which responded to the survey) and the weighting of the dataset. The overall estimate of  $\pm 3.3$  percent takes those factors into account. The estimated overall design effect from complex sampling is 1.51.

The margins of error are larger for questions answered by smaller numbers of respondents, and for subgroups in the data.

---

<sup>2</sup> The finite population correction is calculated as  $\sqrt{1 - f}$ , where  $f$  is the sampling ratio  $n/N$  (sample size divided by population size). With  $f = 0.42$ , the finite population correction is .76, which means that the confidence range or margin of sampling error is about three-quarters of the size it would be if a sample of the same size had been drawn from a very large population.