Public Opinion Quarterly

REDIRECTED INBOUND TELEPHONE CALL SAMPLING (RICS)
TELEPHONE SURVEYING VIA A NEW SURVEY SAMPLING PARADIGM

BURTON LEVINE*
KAROL KROTKI
PAUL J. LAVRAKAS

Abstract Redirected Inbound Call Sampling (RICS) is an emerging, nonprobability sampling methodology in which calls to nonworking numbers from callers throughout the United States are redirected to a telephone survey recruitment protocol and data collection system. The use of automated, interactive voice response technologies to recruit, screen, and collect data achieves significant cost savings and greatly reduces the time of fielding a survey compared with traditional dual-frame random-digit dialing (DFRDD) surveys that use interviewers for recruitment and data collection. The implementation of RICS that we fielded resulted in respondents who match some demographic of the population about as well as DFRDD telephone surveys. However, we demonstrate some non-ignorable challenges with measurement error in certain types of questions that arise from primacy effects associated with using an interactive voice response system for data collection. We present the results of a RICS study that was designed to better understand the reliability and validity of the data these surveys generate. The investigation presented in this manuscript is a first step to evaluating if RICS can be a fit-for-purpose solution for some survey needs. Our data suggest there is sufficient promise in the RICS methodology to warrant continued development and refinement.

Burton Levine is a senior research statistician at RTI International, Research Triangle Park, NC, USA. Karol Krotki is a senior research statistician at RTI International, Washington, DC, USA. Paul J. Lavrakas is a self-employed independent researcher, Evanston, IL, USA. The authors are grateful to the three anonymous reviewers for helpful comments. This work was supported by RTI International. Lavrakas's time working on the manuscript was volunteered/unpaid. *Address correspondence to: Burton Levine, RTI International, 3040 E. Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709, USA; phone: (919) 541–1252, email: blevine@rti.org.

doi:10.1093/poq/nfz024

Redirected Inbound Call Sampling (RICS) is one of several emerging nonprobability telephone sampling methodologies (Lavrakas et al. 2017). In RICS, telephone calls to nonworking numbers, placed by the potential respondents throughout the nation during their normal phone use, are redirected to a telephone survey recruitment and data collection system. In the United States, Telephone Consumer Protection Act rules state that automated telephone equipment cannot be used to place a call to a potential respondent; these rules do not apply to RICS because potential respondents initialize the call. The major advantages of RICS are its low cost, speed, and ability to gather data from large samples of respondents throughout the nation and throughout the 24-hour day. In this article, we describe RICS sampling, recruitment, and data collection methodologies, as well as implementation decisions. We present results from a RICS study that mimicked questions from the 2016 National Health Interview Survey. Using these data, we quantify bias with respect to basic demographics and study outcomes, investigate primacy effects, and characterize "night owl" respondents from whom data can be gathered outside normal telephone data collection hours.

How Does RICS Work?

Anyone who has ever placed a phone call and heard a variant of the message "The number you dialed is not in service…" has placed a call to a nonworking number. In RICS, when someone calls a nonworking number, rather than playing an "error" message and ending the call, the caller is redirected to a survey recruitment and data collection system. There are two types of phone numbers: direct inward dialing (DID), also known as toll calls, and toll free. The method of obtaining calls to redirect is different for each type of telephone number. Redirecting DID calls requires forming partnerships with local telephone companies. In contrast, calls to nonworking toll-free numbers are sold by telecommunication companies that control large banks of toll-free numbers. Redirecting toll-free calls to nonworking numbers onto a data collection system requires control of a large bank of nonworking toll-free numbers and the purchase of calls from the entity that is selling them. For the remainder of this paper, we focus on surveys that only use inbound calls to nonworking *toll-free* numbers that have been redirected to a survey recruitment and data collection system.

There is an active market for inbound redirected toll-free calls, and telemarketing to toll-free calls is commonplace. In September 2017, we conducted a study to estimate the proportion of toll-free numbers that are redirected. For this study, we called a random sample of 756 toll-free numbers and found that 9 percent of those numbers were redirected. We are not suggesting that 9 percent of all toll-free calls are redirected; rather, that approximately 9 percent of misdials to toll-free numbers are redirected. We estimate that in September 2017 inbound calls to approximately 5.4 million different toll-free numbers in the United States were redirected by some organization for marketing or data collection.

In November 2013, Scott Richards, CEO of Reconnect Research, conceived of the application of RICS methodology for survey research. Reconnect Research is a telecommunications company that specializes in redirecting inbound calls, and obtains millions of inbound calls monthly through its bank of nonworking toll-free numbers and by purchasing inbound calls from other entities. For example, Reconnect Research received 11 million inbound calls in a 30-day period ending in July 2018. At any given time, it is able to field several surveys and marketing campaigns simultaneously. For our 2017 full evaluation study of RICS, we selected inbound calls randomly from all redirected toll-free calls coming into Reconnect Research.

RICS SAMPLING

Nonprobability sampling is a term that refers to a broad range of sampling methods. Based on the classification presented in the "Report of the AAPOR Task Force on Non-Probability Sampling" (Baker et al. 2013), RICS is a form of convenience sampling, which the report defines as "a form of nonprobability sampling in which the ease with which potential participants can be located or recruited is the primary consideration" (17). Among the subcategory of convenience sampling described in the report, RICS most closely matches *river sampling*, which is characterized by a "cluster" of ways that a respondent can enter the sample. In the case of the RICS application discussed in our evaluation study, the cluster comprises the hundreds of thousands of nonworking toll-free numbers that a potential respondent might misdial. Consequently, the entire adult population of the United States that uses a telephone has a nonzero probability of misdialing a toll-free number that is among the group of numbers that are redirected. Thus, the potential coverage error associated with RICS is very small. During the second half of 2017, 3.1 percent of adults did not own a phone (Blumberg and Luke 2018), but even these individuals who do not own a phone could use a phone to dial a nonworking toll-free number and be sampled in a RICS survey.

RICS RECRUITMENT

Two basic options exist for recruiting respondents from a RICS sample. First, the redirected call can be connected to an interactive voice response (IVR) system with a recorded recruitment protocol that is played for the caller. Once cooperation is gained via this method, the data collection stage begins, most often via IVR. The IVR method for recruitment is one of the reasons that RICS radically lowers the cost of surveying compared with using interviewers (e.g., in outbound dual-frame random-digit dial [DFRDD] surveying) to do the recruitment. Second, and in contrast to the IVR approach, the redirected inbound call can be sent to a telephone recruiter-interviewer for recruitment, and once cooperation is gained, the data collection stage can proceed, which can be carried out via live interviewers or IVR.

RICS DATA COLLECTION

A RICS survey offers three options for the data collection: (1) using an IVR system, (2) using a live interviewer, or (3) sending the caller to a website to complete a questionnaire via a computer-assisted web interviewing (CAWI) system. Any of these three data collection methods can be combined with either of the two recruitment approaches. The standard approach for RICS recruitment and data collection is IVR. It is possible to use IVR to push respondents to CAWI for RICS data collection. However, from past research (cf. Tourangeau, Steiger, and Wilson 2002), we estimate that this could result in an increase in nonresponse of at least 20 percent (and possibly more than 50 percent) compared with gaining cooperation from RICS respondents while offering them the telephone as the mode for data collection.

RESEARCH ON THE USAGE OF IVR TO RECRUIT RESPONDENTS

Because IVR is the primary mode associated with RICS recruitment and data collection, we conducted a literature review of IVR surveying methods. We looked for findings in which IVR was tested experimentally *as a recruitment mode against other recruitment modes*. We were able to locate only one methodological study (Corkrey and Parkinson 2002) that involved the testing of recruitment conducted via IVR (i.e., no human interviewer was used to recruit). The authors found that the IVR-only mode suffered from a far lower response rate compared with the interviewer recruitment method and a far higher refusal rate. On other recruitment metrics in this study, the IVR-only mode performed comparably to the telephone interviewer recruitment mode (e.g., the final unweighted samples were similarly representative across age, gender, education, marital status, and employment status).

RESEARCH ON THE USAGE OF IVR TO GATHER SURVEY DATA

In contrast, there is a fairly extensive body of experimentation reports on IVR (also called telephone computer-assisted self-interviewing, or T-CASI) as a data collection mode. The major and consistent finding of these studies is that IVR generally gathers sensitive data that are less subject to social desirability bias compared with telephone interviewer-administered data collection, but that IVR suffers in terms of more missing values and break-offs (Turner et al. 1996; Cooley et al. 1998; Tourangeau et al. 2002; Currivan et al. 2004; Gribble et al. 2009; Harmon et al. 2009; Sakshaug, Yan, and Tourangeau 2010).

The decision to use IVR for recruitment *and* data collection is a complex one that needs to balance nonresponse error considerations, measurement error considerations, and cost considerations.

SAMPLING, RECRUITMENT, DATA COLLECTION, AND DATA PROCESSING IN A RICS CASE STUDY

In our 2017 RICS evaluation case study, we fielded an instrument that mimicked several questions from the 2016 National Health Interview Survey (NHIS) Adult File (Centers for Disease Control and Prevention [CDC] 2017). The instrument is presented in the appendix. We chose the 2016 NHIS Adult File as a "gold standard" comparison because it is a comprehensive national public health survey with a public use file, uses probability sampling, has a relatively high response rate (54.3 percent), and has a large sample size (n = 33,028) (CDC 2017). For this 2017 RICS survey, Reconnect Research recorded the IVR recruitment and questionnaire scripts, programmed the IVR system, redirected their inbound calls to our IVR survey recruitment script, collected the data, and created the raw data file.

SAMPLING AND RECRUITMENT

Our RICS evaluation study sampled 137,840 inbound calls from 50 states and the District of Columbia across two seven-day periods, one in January 2017 and one in February 2017. Sampling was carried out across the entire 24-hour day. In each case, an IVR recording invited the redirected caller to complete our questionnaire:

> *Please answer our brief national health survey by RTI International. This should take less than 7 minutes. Your call couldn't be completed and was sent to this survey. Let's begin …*

IVR was used to screen each inbound call for adults (i.e., callers who were 18 years old and older). There were 24,544 (18 percent) recruited inbound calls that responded to the screener. Among those that were screened, 21,847 (89 percent) qualified as eligible adults. Among the eligible screened adults, 9,465 (43 percent) were considered respondents because they provided valid responses to nine or more questions and were not considered straight-liners, or those who repeatedly answer in the same manner to a series of questions that share the same response format. Among these valid respondents, 7,496 (79 percent) completed the entire questionnaire. The American Association for Public Opinion Research (AAPOR) response rate (RR4) was 7.7 percent.[1] The median length of time that it took respondents to complete the questionnaire was 4.9 minutes; the mean was 4.8 minutes (SD = 1.1).

1. Respondents—9,465; nonrespondents—12,382; unknown response status—113,296; ineligibles—2,697;
$e = \frac{9465+12382}{9465+12382+2697} = 0.89$; AAPOR response rate $(\text{RR4}) = \frac{9465}{9465+12382+(0.89*113296)} = 0.077$

IVR QUESTIONNAIRE IMPLEMENTATION

In the IVR presentation of the RICS questionnaire, the respondent was given up to two opportunities to answer each question. Any telephone touch-tone key that the respondent pressed while the question was being read was ignored. After each question, the respondent was prompted to respond using the telephone keypad. If a valid response option was provided anytime during or after the response options were spoken, the IVR system stopped presenting additional response items and started the next question. Invalid response options (e.g., 0 or 4–9 when only 1, 2, or 3 was a valid option) were ignored by the IVR system. After the question was spoken and all response options were presented and if no valid response was forthcoming, there was a two-second delay before the question and response options were presented a second time. Then, if a valid response option still was not provided, the IVR system continued with the next question. Overall, only 1.5 percent of the valid responses were provided on the second reading of the questions.

When gathering data via an IVR system, survey researchers have several alternative ways to administer the questionnaire and collect responses. The quality of the IVR data collection depends heavily on the implementation of the IVR system. The left-hand column of table 1 indicates the RICS IVR features that we chose to implement in our study, whereas the right-hand column indicates an example of an alternative option that was considered in each case.

REMOVING RESPONDENTS WITH UNRELIABLE DATA

We implemented two data-cleaning procedures to identify and remove respondents who provided too much incomplete or unreliable data. Individuals who screened into the study but provided fewer than nine valid responses, apart from the screening question, were considered nonrespondents. We also considered the problem of straight-lining, a form of satisficing. Identifying straight-lining is somewhat difficult, because a respondent might switch at any time between providing legitimate responses and straight-lining. For survey participants with nine or more valid response options, we calculated the proportion of responses for their most-used response option. If this proportion exceeded 70 percent, we identified the participants as straight-liners and considered them as unreliable respondents. We identified only 213 survey participants as straight-liners, which is 2.2 percent of the 9,691 respondents.

WEIGHTS

We created weights by first initializing all respondents to an identical base weight, equal to the ratio of the population total to the number of respondents. Then we calibrated to the marginal population distributions by sex, age category, race/ethnicity, educational attainment, and census division (Kott 2016).

Table 1. IVR implementation decisions used and possible alternative decisions

IVR implementation used	Possible alternatives
The respondent manually enters data by using the keypad.	Using conversational IVR in which the respondent provides a spoken response. The spoken words are recorded, and voice recognition software assigns a numeric value to them.
Providing a “prefer not to respond” option in each question.	Mentioning a “prefer not to respond” option only once at the beginning of the questionnaire.
Allowing the respondent to answer before all response options are read.	Requiring the respondent to hear all response options before answering.
Adding comforting or encouraging language between some questions, such as “you’re doing great.”	Not including comforting language.
Using the cadence of normal everyday speech.	Using slower or faster speech cadence than normal.

NOTE.—IVR = interactive voice response.

The specific categories and the population totals for each category are displayed in table 2. The population totals for sex, age category, race/ethnicity, and census division come from the State Characteristics Population Estimates (U.S. Census Bureau 2017). The distribution for educational attainment came from the American Community Survey (ACS) 2016 one-year summary file (American Community Survey Office 2017). By using this weighting methodology, we implicitly assumed that the RICS selection mechanism mimics a simple random sample.

The calibration procedure requires respondents to have zero missing values for distributions used in the calibration. Thus, we applied a hot-deck imputation to the variables used in the calibration other than the census region. The following is a list of the hot-deck imputed calibration variables with the percentage of imputed values: sex (2.1 percent), age category (1.8 percent), race/ethnicity (4.8 percent), and educational attainment (2.5 percent). To assign the calibration variable for the U.S. Census division, we first assigned respondents to a state based on the respondent-provided ZIP code. There were 4,673 (49.3 percent) respondents who provided a ZIP code that matched a list of valid ZIP codes. For the 50.7 percent of respondents who did not provide a valid ZIP code, we imputed their state from the ZIP code metadata. For the respondents who provided a valid ZIP code, 93 percent of

Table 2. Comparison of demographic distributions among the population and the unweighted RICS sample and the BRFSS sample using the design weight

Distribution	Category	Total	Population	RICS	BRFSS
Sex	Female	128,015,520	51.3	56.0[a]	53.1[a]
	Male	121,469,708	48.7	44.0[a]	46.9[a]
Age category	18–24	30,843,811	12.4	12.8	8.4[a]
	25–34	44,677,243	17.9	15.7[a]	14.6[a]
	35–44	40,470,156	16.2	15.3	14.3[a]
	45–54	42,786,679	17.1	17.4	17.7[a]
	55–64	41,463,144	16.6	17.4	20.0[a]
	65 and older	49,244,195	19.7	21.3[a]	25.1[a]
Race/ ethnicity	White	160,321,206	64.3	54.1[a]	68.1[a]
	African American	30,090,673	12.1	24.2[a]	10.3[a]
	Hispanic	39,124,545	15.7	14.2[a]	14.2[a]
	Other race	19,948,804	8.0	7.5	7.4[a]
Educational attainment	Less than high school	33,955,989	13.6	17.2[a]	9.0[a]
	High school	69,889,267	28.0	38.1[a]	25.2[a]
	Some college	77,944,203	31.2	28.5[a]	26.7[a]
	Bachelor's or higher	67,695,769	27.1	16.2[a]	39.1[a]
Census division	New England	11,761,918	4.7	4.2	N/A
	Middle Atlantic	32,633,869	13.1	13.6	N/A
	East North Central	36,163,490	14.5	13.8	N/A
	West North Central	16,201,655	6.5	5.7[a]	N/A
	South Atlantic	49,949,626	20.0	24.9[a]	N/A
	East South Central	14,609,659	5.9	9.2[a]	N/A
	West South Central	29,380,854	11.8	11.8	N/A
	Mountain	18,068,655	7.2	5.0[a]	N/A
	Pacific	40,715,502	16.3	11.8[a]	N/A

NOTE.—BRFSS = Behavioral Risk Factor Surveillance System; RICS = Redirected Inbound Telephone Call Sampling.

[a]Statistically significantly different than the population, two-tailed t-test (p-value < 0.0012). We applied a Bonferroni adjustment. Table 2 has 43 hypothesis tests, alpha = 0.05/43 = 0.0012.

the time the state of the user-provided ZIP code matched the state of the ZIP code from the metadata.

We used the weighting procedure just described to produce *delete 1-group jackknife weights*, which we used to estimate variance (Wolter 1985). We randomized the respondents to 30 equally sized groups. First, we deleted the first group and created weights using the calibration methodology discussed previously. Then we deleted the second group from the full dataset and created the second set of deleted-grouped-jackknife weights. This procedure continued for all 30 groups. We also created one set of regular weights without using a deleted-grouped-jackknife approach to evaluate the demographic distribution of the respondents. Using these weights, we calculated the unequal weighting effect (UWE), a function of the variance of the weights. The larger the deviation between the demographic distribution of the respondents compared with the population, the larger the UWE.

A proposed methodology for calculating sampling weights for nonprobability samples involves two steps (Valliant and Dever 2011; Elliot and Valliant 2017). In the first step, a base weight is assigned to the inverse of estimated "pseudo-inclusion" probabilities. To estimate the pseudo-inclusion probabilities, the nonprobability sample is combined with a probability sample, in our case the NHIS. Then a weighted propensity model is fit. The dependent variable is inclusion in the probability sample, and the independent variables ideally are study outcomes that are correlated with the selection probability and the outcomes being estimated. One can create weighting classes and calculate the average pseudo-inclusion probabilities in each weighting class to smooth out the base weights. The second step involves calibrating the demographic distributions to population totals. We applied this two-step process for calculating the base weights using several different combinations of implementation decisions. The only variables available that we determined to be useful, in addition to the calibration variables, were marital status and frequency of telephone calls placed. We found very small differences in the study outcomes using the two-step approach compared with only calibrating to the demographics; therefore, we used the simpler methodology and only calibrated to the demographic variables.

ITEM MISSING DATA

Item missing data is the absence of information on individual data items for survey respondents who provided measurements on other outcomes. Item missing data occurs for three reasons: respondents choose the response option "prefer not to answer"; they did not provide any legitimate response option, which we considered to be a refusal; or they exited the study, also known as a break-off.

Online Appendix A contains a table that describes item nonresponse for the partial and completed respondents, the data used in the analysis file. The median value of the percentage of subjects who selected the prefer not to answer option over the 26 survey questions was 3.7 percent. The ZIP code question and the Hispanic question had large percentages of subjects who selected the prefer not to answer option, 22.4 percent and 9.2 percent, respectively. Refusals made up less than 1 percent of response options for all questions. Inbound callers were considered nonrespondents if they broke off on or before the educational attainment question. After the education attainment question, the percentage of respondents who were break-offs for each question increased to 20 percent by the second-to-last question of the survey.[2]

Of the 128,375 redirected callers who were not considered partial or complete respondents, 88.2 percent did not respond to the screener; they hung

2. If respondents answered the second-to-last question but not the last question, they were considered a refusal for the last question; they were not considered a break-off.

up without responding to any question. Of the 128,375 redirected callers, 2.1 percent responded to the screener but indicated they were not an adult and were therefore ineligible for the survey, so the survey terminated. Two hundred thirteen (0.2 percent) were considered straight-liners. Online Appendix B contains a table that describes item nonresponse for the 12,452 inbound callers who screened as adults and were not identified as straight-liners but were not treated as respondents because they did not both answer at least nine questions after the screener and make it through the demographic section. Of the 12,452 inbound callers who screened as adults, 51.9 percent broke off after the screener, and 93.9 percent of them broke off before the demographic questions were completed.

Noncoverage, Nonresponse, and Related Biases in RICS

We compared the unweighted demographic distributions in our survey to U.S. Census data (U.S. Census Bureau 2017) and to 2016 Behavioral Risk Factor Surveillance System (BRFSS) data, which come from DFRDD surveys in all 50 states. We compared both categorical and continuous outcomes to the NHIS data. To provide context, we also compared RICS demographic results with an outbound DFRDD telephone survey. The three primary reasons the demographic characteristics of a final unweighted RICS sample may not match the target population's demographic parameters are noncoverage bias, nonresponse bias, and the error of getting the initial probability of selection wrong. Note that we assigned equal base weights, implicitly assuming that the selection mechanism mimics a simple random sample. The following section describes how well the final sample in our 2017 RICS evaluation survey matched the target population, although we were not able to sort out the separate effects of the three reasons listed above.

DEMOGRAPHICS OF RESPONDENTS

Table 2 contains a comparison of the demographic distributions among the population, the unweighted RICS survey respondents, and the BRFSS survey respondents using the design weights. To calculate the distributions for the BRFSS respondents, we used the design weight with an adjustment to account for the disproportional allocation across geographical strata. This adjustment forces the sum of the design weights to equal the population total within the geographical strata. This approach eliminates the bias in the BRFSS from some geographical strata being over- (and under-) represented.

Because the BRFSS is a probability sample survey, the data most comparable to the unweighted RICS distributions are the BRFSS distributions, after adjusting for the study design but before calibration. We realize that the BRFSS design weights, included in the public use file, involve weight trimming and a nonresponse adjustment. Ideally, we would like to compare the distributions without these adjustments, but recalculating the BRFSS weights is not feasible. The weight trimming generally increases bias and, most likely, has a negative effect on the corroboration between the demographics of the respondents and the population, whereas the nonresponse adjustment has a positive effect.

Compared with the population, the unweighted RICS survey respondents are more likely to be female, less likely to be non-Hispanic white, more likely to be African American, and more likely to have lower educational attainment. Compared with the population, unweighted BRFSS surveying, like almost all DFRDD surveys, overrepresents females, non-Hispanic whites, and those with high levels of educational attainment. Compared with the BRFSS, the demographic distributions of the RICS survey respondents match the population's age distribution better. In contrast to the BRFSS data, the RICS respondents have lower educational attainment than the population and have fewer non-Hispanic whites than the population. Table 2 indicates that the RICS and BRFSS estimates are both statistically different than their corresponding population estimates. We used a two-tailed *t*-test with a Bonferroni correction to account for multiple comparisons. The population estimate is assumed to have no sampling variance.[3]

We measured the match between the RICS and BRFSS survey respondents to the population by the increase in UWE when the design weight is calibrated to the marginal totals of the demographic distributions in table 2. Calculating the increase in the UWE is straightforward for the RICS sample; the design weights are equal and therefore have a UWE of 1. For the RICS data, the calibrated weights, without using a delete 1-group jackknife, have a UWE of 1.28, so the increase in UWE is 28 percent (1.28/1). For the BRFSS sample, the UWE of the design weights with an adjustment to account for the disproportionate allocation by geographical strata is 3.21. The final calibrated weight is 3.94, an increase of 23 percent (3.94/3.21). Because the imputation to the BRFSS calibration procedure contains an adjustment that controls the geographic distribution, a fairer comparison is with the RICS-calibrated weights without calibrating to census division. When we calibrate RICS data to the marginal distributions in table 2 without census division, the UWE is 1.25. The increase in UWE for calibrating to the marginal distributions of sex, age category, race/ethnicity, and educational attainment is 1.25 for the RICS and 1.23 for the BRFSS. Overall, the RICS respondents match the population almost as well as the BRFSS respondents. We speculate that the bias caused by

3. The BRFSS has 477,665 respondents; consequently, there is adequate power to detect very small deviations in estimates.

assuming the selection bias is a simple random sample in the RICS study is minimal. If this assumption resulted in a large bias, a much larger mismatch should exist between the respondents and the population for the RICS study compared with the BRFSS.

DATA FOR OTHER MEASURES

Table 3 displays the RICS estimates for 12 measures. All but two of these estimates are compared with their NHIS estimates. The marital status distribution is compared with estimates from the 2016 ACS 1-year summary file, and the health insurance estimate is compared with an estimate from Gallup (Auter 2017). The proportion of the population that has health insurance has been changing rapidly because of the Affordable Care Act. We used an estimate from Gallup because it was closer in time to the RICS data collection than the estimates from the NHIS. The absolute difference is the absolute value of the difference in the estimate for every category. For binary variables, the mean absolute difference is the same as the absolute difference. For outcomes with three or more categories, the mean absolute difference is the average absolute difference over all the categories. The mean, across all the outcomes, of the mean absolute difference is 4.0. The *p*-value is based on testing the null hypotheses that the RICS distribution is the same as the comparison distribution using a chi-square test. The *p*-value was significant for all but one distribution. We used a Bonferroni correction to adjust for multiple comparisons. We were not able to conduct a hypothesis test on the health insurance question because the Gallup reference did not include a measure of variance. We included the unweighted RICS estimates as a measure of the effect of the weights. For almost all outcomes, the weights moved the estimate in the direction of the comparison estimate.

RICS MEASUREMENT ERROR: PRIMACY EFFECT

A primacy effect refers to the respondent-related measurement bias that occurs when response options given early in a list for a question are more likely to be chosen than those presented later in the list (Scanlan 2008). Primacy effects are most likely to occur with self-administered modes of data collection. To test for primacy effects, for two of the questions in our questionnaire, we randomized respondents to two different orderings of the response options. One question was *If you get sick or have an accident, how worried are you that you will be able to pay your medical bills?* One ordering of the response options was as follows: *If very worried, press 1; If somewhat worried, press 2; If not at all worried, press 3; If you prefer not to answer, press 4*. The second ordering of the response options was as follows: *If not at all worried, press 1; If somewhat worried, press 2; If very worried, press 3; If you prefer not to answer, press 4*. Table 4 displays the results using the analysis weights.

Table 3. Comparison of estimates between RICS and comparison studies

		Percentage			Comparison of weighted estimates		
		RICS					
Question (comparison source)	Category	Not weighted	Weighted	Comparison weighted	Absolute difference	Mean absolute difference	*p*-value
Smoked at least 100 cigarettes in your entire life (NHIS)	Yes	39.7	38.8	37.0	1.8	1.8	<0.004[a]
Do you now smoke … (NHIS)	Everyday	35.8	34.5	30.9	3.5	4.8	<0.001[a]
	Some days	13.9	13.7	9.9	3.7		
	Not at all	50.3	51.9	59.1	7.3		
Current smoker (NHIS)	Yes	19.9	18.9	15.1	3.7	3.7	<0.001[a]
Quit attempt past 12 months (NHIS)	Yes	52.8	51.9	49.2	2.7	2.7	0.133
Marital status (ACS)	Now married	38.9	42.7	47.7	5.0	5.6	<0.001[a]
	Widowed	16.0	14.1	6.2	7.9		
	Divorced	16.7	16.1	11.6	4.5		
	Separated	4.5	3.9	2.3	1.6		
	Never married	24.0	23.2	32.2	9.0		
Internet use (NHIS)	Yes	70.0	75.0	78.2	3.2	3.2	<0.001[a]
Health insurance (Gallup[b])	Yes	84.4	86.2	88.3	2.1	2.1	N/A
Ever used firearms (NHIS)	Yes	30.2	34.3	36.6	2.3	2.3	<0.004[a]
Full hours of sleep in a 24-hour period (NHIS)	5 or less	20.0	18.5	9.4	9.2	4.7	<0.001[a]
	6	29.0	28.3	22.6	5.7		
	7	22.4	24.2	29.2	5.0		
	8	18.6	19.3	30.5	11.1		
	9	4.3	4.2	4.4	0.2		
	10	2.9	2.8	2.7	0.1		
	11 or more	2.8	2.7	1.4	1.3		

Continued

Table 3. *Continued*

		Percentage					
		RICS			Comparison of weighted estimates		
Question (comparison source)	Category	Not weighted	Weighted	Comparison weighted	Absolute difference	Mean absolute difference	p-value
Vigorous leisure-time physical activities—times per week (NHIS)	0	35.5	33.6	55.6	22.0	9.9	<0.001[a]
	1 or 2	32.0	31.3	14.1	17.2		
	3, 4 or 5	23.1	24.9	20.6	4.3		
	6 or 7	4.8	5.4	8.2	2.9		
	8 or more	4.5	4.9	1.5	3.3		
Days drank alcohol per month (NHIS)	0	56.5	55.9	48.6	7.4	6.5	<0.001[a]
	1–2	24.4	23.1	17.5	5.7		
	3–6	11.1	11.9	13.1	1.2		
	7 or more	8.0	9.1	20.9	11.8		
Drinks per day on days you drank (NHIS)	1	35.8	36.4	37.9	1.5	1.0	<0.001[a]
	2	32.0	31.5	32.9	1.4		
	3	14.8	14.5	14.3	0.2		
	4	5.7	6.2	6.1	0.1		
	5 or 6	6.2	6.1	6.2	0.1		
	7 or more	5.4	5.4	2.7	2.7		

NOTE.—NHIS = National Health Interview Survey; RICS = Redirected Inbound Telephone Call Sampling.

[a]Statistically significantly different than the population, two-tailed t-test (p-value < 0.0045). We applied a Bonferroni adjustment. Table 3 has 11 hypothesis tests, alpha = 0.05/11 = 0.0045.

[b]The Gallup survey is a national, dual-frame, RDD survey (n = 45,087).

Table 4. Results of two different ordering of response options for two different questions

Question 1: How worried about paying medical bills				
Response category	Ordering 1% (95% CI)	Ordering 2% (95% CI)	Absolute difference	*p*-value
Very	19.9 (18.6, 21.3)	13.7 (12.4, 14.9)	6.3 (4.4, 8.1)	<0.001[a]
Somewhat	29.9 (28.3, 31.5)	28.6 (26.9, 30.4)	1.3 (0.0, 3.6)	
Not at all	50.1 (48.4, 51.9)	57.7 (55.8, 59.5)	7.5 (5.0, 10.1)	

Question 2: Change in health insurance compared to a year ago				
Response category	Ordering 1% (95% CI)	Ordering 2% (95% CI)	Absolute difference	*p*-value
Better	27.9 (26.3, 29.5)	9.6 (8.4, 10.7)	18.3 (16.4, 20.3)	<0.001[a]
Worse	17.3 (15.9, 18.6)	15.7 (14.3, 17.1)	1.6 (0.0, 3.6)	
About the same	54.8 (53.1, 56.6)	74.8 (73.1, 76.5)	19.9 (17.5, 22.4)	

[a]Statistically significantly different than the population, two-tailed *t*-test (p-value < 0.0045). We applied a Bonferroni adjustment. Table 4 has 2 hypothesis tests, alpha = 0.05/2 = 0.025.

In both questions, the ordering of the questions was significantly associated with the outcome ($p < 0.001$). For the question on worrying about paying medical bills, when "very worried" was the first response option, 19.9 percent of the respondents reported that answer, whereas 13.7 percent of respondents reported being "very worried" when the response options were reversed and it was presented last. We see an even larger effect, nearly a 20-percentage-point difference, for the question about change in health insurance coverage.[4]

NIGHT OWLS

Telephone survey data are usually gathered between 9 a.m. and 9 p.m. in a respondent's local time zone. In contrast, RICS researchers can gather data across the entire 24-hour day. We define *night owls* as inbound callers who initiated the interview after 9 p.m. and before 9 a.m. in the time zone from where the survey respondent is calling; we refer to other survey respondents as "daytimers." There were 2,893 (30.6 percent) respondents in our survey who were night owls and 6,572 who were daytimers. The AAPOR response rate (RR4) of night owls (10.7 percent) was higher than daytimers (7.1 percent). Table 5 displays the demographic distributions of night owls and daytimers. Compared with the daytimers, the night owls have a higher proportion of males, are younger, have higher proportions of African Americans and Hispanics, and are less educated. These

4. One order was "Better, Worse, About the Same." The other order was "About the Same, Worse, Better."

Table 5. Demographic distributions of night owls and daytimers

Distribution	Category	Population	BRFSS[a]	RICS night owl	RICS day-timers	p-value[b]
Sex	Female	51.3	53.1	50.2	58.5	<0.001[c]
	Male	48.7	46.9	49.8	41.5	
Age category	18–24	12.4	8.4	18.1	10.5	<0.001[c]
	25–34	17.9	14.6	18.5	14.5	
	35–44	16.2	14.3	16.5	14.8	
	45–54	17.1	17.7	16.8	17.6	
	55–64	16.6	20.0	14.8	18.6	
	65 and older	19.7	25.1	15.3	24.0	
Race/ethnicity	White	64.3	66.7	50.0	55.9	<0.001[c]
	African American	12.1	10.1	26.8	23.0	
	Hispanic	15.7	13.9	15.1	13.8	
	Asian	5.7	3.6	3.4	3.3	
	Other race	2.3	5.7	4.8	3.9	
Educational attainment	Less than high school	13.6	9.0	18.3	16.7	<0.001[c]
	High school	28.0	25.2	40.1	37.3	
	Some college	31.2	26.7	27.4	30.0	
	Bachelor's or higher	27.1	39.1	14.3	17.0	

NOTE.—BRFSS = Behavioral Risk Factor Surveillance System; RICS = Redirected Inbound Telephone Call Sampling.

[a]The BRFSS estimates were calculated with the design weights with an adjustment for the geographical strata.

[b]Chi-square test for the null hypotheses that the RICS night owls' and RICS daytimers' distributions are identical.

[c]Statistically significantly different than the population, two-tailed t-test (p-value < 0.0125). We applied a Bonferroni adjustment. Table 5 has 4 hypothesis tests, alpha = 0.05/4 = 0.0125.

differences were all statistically significant ($p < 0.001$). (We used a two-tailed t-test with a Bonferroni correction to account for multiple comparisons.)

In addition, table 5 presents the demographic distributions of the population, the BRFSS respondents, and two subsets of RICS respondents. The BRFSS results are adjusted for the study design and the disproportionate allocation to geographical strata. Compared with the population, and as previously noted, more BRFSS respondents are female and older, are less likely to be African American and Hispanic, and have higher educational attainment. The demographic characteristics that are underrepresented in the BRFSS are the characteristics that are more prevalent in the RICS night owls compared with the RICS daytimers. It seems likely that some types of individuals have a higher response propensity to be sampled and to participate in a survey during the night than the day. These are individuals that RICS gathers data from, whereas DFRDD does not.

Discussion

In a previous RICS evaluation study conducted at RTI in 2015, we fielded a RICS study with IVR data collection like the 2017 RICS evaluation presented in this paper (Levine 2016). Unlike the 2017 RICS evaluation, the instrument used in the 2015 RICS evaluation mimicked several items from the 2014 BRFSS instrument. The 2014 BRFSS is a DFRDD survey with an overlapping design. We believe that evaluating bias in a RICS study by comparing study outcomes with the BRFSS was hampered by biases in the BRFSS. Yeager and others (2011) estimated that, on average, good-quality national DFRDD telephone surveys have three-percentage-point absolute errors relative to benchmarks in the data they produce. For the evaluation described in this paper, we used the in-person NHIS as the benchmark comparison survey, based on the belief that the NHIS has less bias than the BRFSS, because it uses in-person recruitment and in-home data collection and has a much higher response rate than the BRFSS. However, we did not fully appreciate, at the time we created the instrument for our 2017 RICS evaluation study, the difficulty in mimicking the NHIS survey questions using an IVR instrument. Converting BRFSS questions to an IVR instrument was not nearly as problematic as converting NHIS questions. The BRFSS is a telephone survey that uses computer-assisted telephone interviewing; consequently, the questionnaire has been developed/tested for communicating via telephone.

We illustrate the differences between the NHIS and RICS instruments with the physical activity question. In the NHIS, this question was worded as follows: *"How often do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?"* The respondent can answer this NHIS question with any quantity and any time frame: *2 times per week*, *20 times per month,* and *6 times per year* are all valid responses. The NHIS interviewer enters the quantity of events in one data field and the time unit in another data field. For the RICS data collection, we asked, *"How many times per week do you do VIGOROUS leisure-time physical activities?"* We broke response options into categories (e.g., If 0 times, press 1; if 1 or 2 times, press 2; if 3, 4, or 5 times, press 3; if 6 or 7 times, press 4; if more than 7 times, press 5; if you prefer not to answer, press 9). These differences in question wording likely account for some of the differences observed between the RICS survey and the NHIS benchmarks.

The RICS survey obtained an AAPOR response rate (RR4) of 7.7 percent. This is comparable to many outbound DFRDD telephone surveys. In a meta-analysis of response rates across several DFRDD studies, Dutwin and Lavrakas (2016) reported that by 2015 response rates from the landline RDD frame were 9.3 percent and were 7.0 percent from the cell phone RDD frame. It may be surprising for the RICS response rate to mimic the DFRDD response rates, considering that DFRDD surveys use human interviewers to gain cooperation from respondents and make multiple outbound call attempts to each number over more than one day, whereas in RICS, there is one opportunity to recruit the respondent and only the IRV system was used to recruit in our evaluation study. If the RICS caller does not participate in the survey, s/he will never be recruited to that survey again.[5] What may be an important difference between the inbound RICS and an outbound DFRDD survey such as the BRFSS is that inbound callers recruited to the RICS survey are already on the phone expecting to talk to someone. Furthermore, because they are calling a toll-free number, the expectation to interact with an IVR data collection system is not unreasonable. In addition, in an outbound telephone data collection system, the potential survey respondent is necessarily "interrupted" from whatever else s/he was doing at the time. If s/he is on the phone already, the telephone CATI system/interviewer making the outbound call would receive a busy signal. Therefore, it is more burdensome and disruptive for people to receive an incoming call from an interviewer than to initiate an outbound call themselves when they are already expecting to talk to someone, as in a RICS survey.

RICS is cheaper and faster than DFRDD because RICS can recruit study participants and collect data with an automated system, whereas DFRDD requires a large amount of time using telephone interviewers. Using RICS with IVR data collection costs about $7 per respondent to recruit a national sample of adults using the methodology described in this paper. However, the use of an IVR system for data collection also is a source of apparent measurement error for certain types of questions. We observed large primacy effects for some RICS survey items, with a 6- and a 20-percentage point swing in survey estimates based on the ordering of the response categories.

Table 3 displays the comparison between RICS estimates and gold standards. Five questions were binary, other than "prefer not to answer"; current smoking is not considered binary for this purpose because it is derived from the current smoking question that has three outcomes. The average absolute difference between the RICS estimate and the comparison estimate for binary outcomes is 2.4 percentage points. We suspect that there is minimal (i.e., likely "ignorable" in most instances) measurement error when the questions are less complex. Six questions had three or more response options. For each question, we calculated the mean absolute percentage difference across the response options. The mean difference across the six questions is 5.5 percentage points.

5. A researcher using RICS to sample the public could request callbacks to be made to RICS respondents who initially decided not to participate. This would add to costs and lengthen data collection time but would certainly raise response rates, although its effect on reducing nonresponse bias is unknown.

This amount of bias would generally be considered non-ignorable for scientific purposes. Reducing primacy effects specifically, and measurement error in general, is an important goal of future RICS research.

Like the work on web panels (Kennedy et al. 2016), an area of future RICS research is to investigate methodology for removing potential selection bias by first identifying characteristics that are correlated with both the propensity to be included in the respondent pool and the study outcomes after accounting for the demographic variables used in the calibration. We can either include these variables in the propensity model used to calculate pseudo-inclusion probabilities or use these variables in the calibration.

A valid critique of RICS, as well as other nonprobability surveys, is that most often there is no theoretical justification underpinning the generalization of the survey data to a population. Further research in RICS includes creating and disseminating measures to evaluate the assumption that the selection mechanism approximates a simple random sample. But today, even without a theoretical justification, nonprobability samples are regularly used to make inferences to their respective target populations. Two popular nonprobability sampling methods are opt-in web panels and sampling from social media websites. Both methodologies potentially suffer from significant coverage bias because they do not cover non-internet users. One use of RICS methodology is to complement these alternative nonprobability sampling methodologies. For example, data from RICS and opt-in web panels (or social media samples) could be combined to ensure the non-internet users are covered.

Survey researchers face ethical considerations when recruiting redirected toll-free numbers. Recruitment to a RICS survey is inappropriate when the recruited individual is calling about an emergency. For example, a redirected caller might be attempting to contact someone about a health emergency, a company that provides roadside assistance, or myriad other emergency services. The authors plan to evaluate the effect of different informed consent scripts that include instructions to discontinue the call if the caller is in an emergency. We plan to investigate the relationship between the informed consent scripts, the user experience, and participation rates.

Conclusion

The main findings of this evaluation of RICS methodology are that (1) the demographics of the respondents match the population almost as well as the DFRDD BRFSS surveys; (2) across 12 public health outcomes, the RICS estimates of proportions differ from the NHIS estimates by an average of 4.0 percentage points; (3) large primacy effects were observed for two attitudinal questions; and (4) RICS provides access to the demographically different night owl cohort. We are not able to disentangle the sources of bias into selection bias, nonresponse bias, and measurement bias. However, we suspect that measurement bias comprises a large part of the total bias because the demographic distributions of the respondents matched the population nearly as well as the BRFSS, and some large primacy effects were identified.

We see several strategies for reducing bias in future RICS surveys by modifying the IVR implementation decisions. For example, we can test the effect of implementing the alternative decisions in table 1. Conversational IVR uses voice recognition software with an automated system to collect survey data from respondents' speech. Theoretically, conversational IVR can mimic a live interviewer and that may reduce measurement error. Another approach is to recruit subjects to a CATI or CAWI survey, data collection modes that generally have less measurement error than traditional IVR.

Although more research needs to be conducted, RICS sampling appears to be a promising tool for conducting population surveillance. The low cost and speed suggest that it likely is a fit-for-purpose solution in many situations.

Appendix. Questionnaire

Please answer our brief national health survey by RTI International. This should take less than 7 minutes. Your call couldn't be completed and was sent to this survey. Let's begin…

Screener: (_18_plus) If you're 18 or older press 1, if you're 17 or younger press 2.

a. **If 1**: "Thank you. This survey is voluntary and your answers are confidential."
b. **If 2**: "Thank you. This study is for adults 18 and older only." **[TERMINATE]**

1. (N_SMKEV) Have you smoked at least 100 cigarettes in your ENTIRE LIFE?
   a. If yes, press 1
   b. If no, press 2

[If _N_SMKEV = 1 then ask N_SMKNOW]

2. (N_SMKNOW) Do you NOW smoke every day, some days or not at all?
   a. If every day, press 1
   b. If some days, press 2
   c. If not at all, press 3
   d. If you prefer not to answer, press 4

[If N_SMKEV=1 or N_SMKNOW in (1, 2) then ask N_CIGQTYR]

3. (N_CIGQTYR) During the PAST 12 MONTHS, have you stopped smoking for more than one day BECAUSE YOU WERE TRYING TO QUIT SMOKING?
   a. If yes, press 1

b. If no, press 2
c. If you prefer not to answer, press 3

[Ask of all respondents]

4. (N_ACISLEEP) On average, how many FULL hours of sleep do you get in a 24-hour period? Round 30 minutes or more UP to the next whole hour.
   a. If 5 hours or less, press 1
   b. If 6 hours, press 2
   c. If 7 hours, press 3
   d. If 8 hours, press 4
   e. If 9 hours, press 5
   f. If 10 hours, press 6
   g. If 11 or more hours, press 7
   h. If you prefer not to answer, press 9

Thanks, you're doing great!

5. (A_SEX) If you are a male, press 1. If you're a female, press 2. If you prefer not to answer, press 3.
6. (A_AGE) How old are you?
   a. If you're 18–24 years old, press 1
   b. If you're 25–34 years old, press 2
   c. If you're 35–44 years old, press 3
   d. If you're 45–54 years old, press 4
   e. If you're 55–64 years old, press 5
   f. If you're 65 years old or older, press 6
   g. If you prefer not to answer press 7
7. (A_HISPANIC) Are you Hispanic or Latino?
   a. If yes, press 1
   b. If no, press 2
   c. If you prefer not to answer, press 9
8. (A_RACE) Which best represents your race?
   a. If White, press 1
   b. If Black, press 2
   c. If Asian, press 3
   d. If other race, press 4
   e. If you prefer not to answer, press 8
9. (A_ED) What is your highest level of educational attainment?
   a. If you did not graduate high school, press 1
   b. If high school graduate or G.E.D., press 2
   c. If some college, technical or vocational school, press 3
   d. If an Associate's degree, press 4
   e. If a Bachelor's degree or higher, press 5
   f. If you prefer not to answer press 7

10. (A_MARITAL) What is your marital status?
    a. If now married, press 1
    b. If Widowed, press 2
    c. If Divorced, press 3
    d. If Separated, press 4
    e. If Never married, press 5
    f. If you prefer not to answer press 6
11. (PHONE_FREQ) In the last 7 days, on average, how many phone calls have you placed a day?
    a. If less than 2 calls, press 1
    b. If 2 to 3 calls, press 2
    c. If 4 to 6 calls, press 3
    d. If 7 to 10 calls, press 4
    e. If 11 to 20 calls, press 5
    f. If more than 20 calls, press 7
    g. If you prefer not to answer press 8

Thanks we're about halfway done.

12. (CELL) Is this a Cell phone?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 3.

[If CELL=1 then ask CELL_w_LL]

13. (CELL_w_LL) Do you also receive personal calls on a landline phone?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 4

[If CELL=2 then ask LANDLINE]

14. (LANDLINE) Is this a landline phone?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 4

[If LANDLINE=1 then ask LANDLINE_w_CELL]

15. (LANDLINE_w_CELL) Do you also receive personal calls on a cell phone?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 5

[Ask N_AWEBUSE of all respondents]

16. (N_AWEBUSE) Do you use the internet?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 3

Thank you, your answers are very important.

17. (P_TWITTER) Do you ever use the internet or mobile app to use Twitter?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 3

[If P_TWITTER=1 then ask P_ TWITTER _FREQ]

18. (P_ TWITTER _FREQ) In the past week, on how many days did you use TWITTER?
    a. If 0 or 1 day, press 1
    b. If 2 or 3 days, press 2
    c. If 4 to 6 days, press 3
    d. If 7 days, press 4
    e. If you prefer not to answer, press 8
19. (N_VIGNO) How many times per week do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?
    a. If 0 times, press 1
    b. If 1 or 2 times, press 2
    c. If 3, 4 or 5 times, press 3
    d. If 6 or 7 times press 4
    e. If more than 7 times, press 5
    f. If you prefer not to answer, press 9
20. (N_ALC12MNO) In the PAST YEAR, on average, how many days per month, did you drink any type of alcoholic beverage? **[If N_ALC12MON>=1 then ask N_ALCAMT]**
    a. If 0 days, press 1
    b. If 1 day per month, press 2
    c. If 2 days per month, press 3
    d. If 3 days per month, press 4
    e. If 4 to 6 days per month, press 5
    f. If 7 or more days per month, press 6
    g. If you prefer not to answer press 8
21. (N_ALCAMT) In the PAST YEAR, on those days that you drank alcoholic beverages, on the average, how many drinks did you have?
    a. If 1 drink, press 1
    b. If 2 drinks, press 2
    c. If 3 drinks, press 3
    d. If 4 drinks, press 4
    e. If 5 or 6 drinks, press 5
    f. If 7 or more drinks, press 6
    g. If you prefer not to answer press 7

[Ask A_HEALTH_INSURANCE of all respondents]

Last few questions…

22. (A_HEALTH_INSURANCE) Do you currently have health insurance?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 3.

Instruction for programmers: Questions 24 and 25 have two orderings of the response categories. We will create four questionnaires, one with each combination of orderings of the response categories for questions 24 and 25. One-fourth of the recruited MIDI calls will get each questionnaire.

23. V1.(N_AWORPAY_V1) If you get sick or have an accident, how worried are you that you will be able to pay your medical bills? Are you very worried, somewhat worried, or not at all worried?
    a. Very worried, press 1
    b. Somewhat worried, press 2
    c. Not at all worried, press 3
    d. If you prefer not to answer, press 4
23. V2.(N_AWORPAY_V2) If you get sick or have an accident, how worried are you that you will be able to pay your medical bills? Are you very worried, somewhat worried, or not at all worried?
    a. Not at all worried, press 1
    b. Somewhat worried, press 2
    c. Very worried, press 3
    d. If you prefer not to answer, press 4
24. V1 (N_AHICOMP_V1) In regard to your health insurance or health care coverage, how does it compare to a year ago? Is it better, worse, or about the same?
    a. Better, press 1
    b. Worse, press 2
    c. About the same, press 3
    d. If you prefer not to answer, press 4
24. V2. (N_AHICOMP_V2) In regard to your health insurance or health care coverage, how does it compare to a year ago? Is it better, worse, or about the same?
    a. About the same, press 1
    b. Worse, press 2
    c. Better, press 3
    d. If you prefer not to answer, press 4
25. (N_HRFIRE) Have you EVER used guns or firearms for any reason?
    a. If yes, press 1
    b. If no, press 2
    c. If you prefer not to answer, press 4.
26. (ZIP_CODE) Using the phone keypad, please enter your 5-digit ZIP CODE ____. If you prefer not to answer press 9.

27. Thank you for taking part in this national health survey conducted by RTI International. Your answers will be confidential. If you have general questions about the survey
    a. press 1,
    b. otherwise press 2.

[if 1 then say] *You can call Jon Doe at 1-800-555–5555 Extension 55555. To hear that number again press 1, otherwise press 2.*
[if 1 then say] *1-800-555–5555 Extension 55555.* [go to 29]
[if 2 then go to 29]

28. If you have questions regarding your rights as a research subject
    a. press 1,
    b. otherwise press 2.

[if 1 then say] *you may call RTI's Office of Research Protection at (555) 555-5555. If you want to hear the number again press 1, otherwise press 2.*
[if 1 then say] *(555) 555-5555, thank you again, good-bye.* [Terminate.]
[if 2 then say] *Thank you again, good-bye.* [Terminate.]
[if 2 then say] *Thank you again, good-bye.* [Terminate.]

Supplementary Data

Supplementary data are freely available at *Public Opinion Quarterly* online.

References

American Community Survey Office. 2017. "*ACS Summary File Technical Documentation (2017): 2016 ACS 1-Year and 2012–2016 ACS 5-Year Data Releases.*" Washington, DC: U.S. Census Bureau.

Auter, Zac. 2017. "U.S. Uninsured Rate Rises to 11.7%." *Gallup News*. Available at https://news.gallup.com/poll/213665/uninsured-rate-rises.aspx, accessed February 9, 2019.

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. "*Report of the AAPOR Task Force on Non-Probability Sampling.*" Deerfield, IL: AAPOR.

Blumberg, Stephen J., and Julian V. Luke. 2018. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2017." Available at https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201806.pdf, accessed December 14, 2018.

Centers for Disease Control and Prevention (CDC). 2017. *Behavioral Risk Factor Surveillance System, Overview: BRFSS 2016*. Atlanta, GA: CDC.

Cooley, Phillip C., Heather G. Miller, James N. Gribble, and Charles F. Turner. 1998. "Automating Telephone Surveys: Using T-ACASI to Obtain Data on Sensitive Topics." *Computers in Human Behavior* 14:195–207.

Corkrey, Ross, and Lynne Parkinson. 2002. "A Comparison of Four Computer-Based Telephone Interviewing Methods: Getting Answers to Sensitive Questions." *Behavior Research Methods, Instruments, & Computers* 34:354–63.

Currivan, D., A. L. Nyman, C. F. Turner, and L. Biener. 2004. "Does Telephone Audio Computer-Assisted Self-Interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking? Evidence from the UMass Tobacco Study." *Public Opinion Quarterly* 68:542–64.

Dutwin, David, and Paul J. Lavrakas. 2016. "Trends in Telephone Outcomes, 2008–2015." *Survey Practice* 9. https://doi.org/10.29115/SP-2016-0017.

Elliot, Michael R., and Richard Valliant. 2017. "Inference for Nonprobability Samples." *Statistical Science* 32:249–64.

Gribble, James N., Heather G. Miller, Philip C. Cooley, Joseph A. Catania, Lance Polack, and Charles F. Turner. 2009. "The Impact of T-ACASI Interviewing on Reported Drug Use among Men Who Have Sex with Men." *Substance Use and Misuse* 35:869–90.

Harmon, Thomas C., Charles F. Turner, Susan M. Rogers, Elizabeth Eggleston, Anthony M. Roman, Maria A. Villarroel, James R. Chromy, Laxminarayana Ganapathi, and Sheping Li. 2009. "Impact of T-ACASI on Survey Measurements of Subjective Phenomena." *Public Opinion Quarterly* 73:255–80.

Kennedy, Courtney, Andrew Mercer, Scott Keeter, Nick Hatley, Kyley McGeeney, and Alejandra Gimenez. 2016. "Evaluating Online Nonprobability Surveys." Pew Research Center. Available at http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/, accessed February 20, 2018.

Kott, Phillip S. 2016. "Calibration Weighting in Survey Sampling." *Wiley Interdisciplinary Reviews: Computational Statistics* 8:39–53.

Lavrakas, Paul J., Grant Benson, Stephen Blumberg, Trent Buskirk, Ismael Flores Cervantes, Leah Christian, David Dutwin, Mansour Fahimi, Howard Fienberg, Tom Gutterbock, Scott Keeter, Jenny Kelly, Courtney Kennedy, Andy Peytchev, Linda Piekarski, and Chuck Shuttles. 2017. *"The Future of U.S. General Population Telephone Survey Research."* Deerfield, IL: American Association for Public Opinion Research. Available at https://www.aapor.org/Education-Resources/Reports/The-Future-Of-U-S-General-Population-Telephone-Sur.aspx.

Levine, Burton. 2016. "Introducing Inbound Calling Survey—A New Sampling Methodology." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Austin, TX, USA.

Sakshaug, Joseph W., Ting Yan, and Roger Tourangeau. 2010. "Nonresponse Error, Measurement Error, and Mode of Data Collection: Tradeoffs in a Multi-Mode Survey of Sensitive and Non-Sensitive Items." *Public Opinion Quarterly* 74:907–33.

Scanlan, Carla R. 2008. "Primary Effect." In *The Encyclopedia of Survey Research Methods*, edited by Paul J. Lavrakas, 610–11. Thousand Oaks, CA: Sage Publications.

Tourangeau, Roger, Darby Miller Steiger, and David Wilson. 2002. "Self-Administered Questions by Telephone: Evaluating Interactive Voice Response." *Public Opinion Quarterly* 66:265–78.

Turner, Charles F., Heather G. Miller, T. K. Smith, Philip C. Cooley, and S. M. Rogers. 1996. "Telephone Audio Computer-Assisted Self-Interviewing (T-ACASI) and Survey Measurements of Sensitive Behaviors: Preliminary Results—Technical Paper 21." In *Survey and Statistical Computing*, edited by Randy Banks, Joan Fairgrieve, Laurance Gerrard, Terry Orchard, Clive D. Payne, and Andrew Westlake, 121–30. Chesham, UK: Association for Survey Computing.

U.S. Census Bureau. 2017. "*7/1/2016 State Characteristics Population Estimates.*" Washington, DC: U.S. Census Bureau. Available at https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010–2016/sc-est2016-alldata6.pdf, accessed December 14, 2018.

Valliant, Richard, and Jill A. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociological Methods and Research* 40:105–37.

Wolter, Kirk M. 1985. *Introduction to Variance Estimation*. New York: Springer-Verlag.

Yeager, David S., Jon A. Krosnick, Linchiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." *Public Opinion Quarterly* 75:709–47.