Redirected Inbound Call Sampling: A Fit for Purpose Nonprobability Sample Design

Burton Levine,* Karol Krotki • RTI International, Research Triangle Park, NC

Redirected Inbound Call Sampling enables researchers to conduct population surveillance on public health outcomes faster and cheaper than traditional data collection methodologies.

1. Introduction

Redirected Inbound Call Sampling (RICS) is a new nonprobability sampling methodology. In RICS, telephone calls to nonworking toll-free numbers are redirected to a recruitment and data collection system. Unlike outbound telephone sampling, the Telephone Consumer Protection Act rules disallowing automated telephone equipment do not apply to RICS surveys since the potential respondent initiates the call. In RICS, both recruitment and data collection can occur via an interactive voice response (IVR) system. The allure of RICS is its low cost, speed, as well as the ability to reach large samples of potential respondents. The cost of a RICS respondent to a national survey of adults using IVR for both recruitment and data collection is about $7; whereas, outbound telephone surveys cost between $70 and $100 per respondent.

The application of RICS to public health research was conceived by Scott Richards, CEO of Reconnect Research, a telecommunications firm that specializes in RICS surveys. Burton Levine has developed a collaborative relationship with Scott Richards and his team at Reconnect Research. In August 2017, Mr. Levine spent 12 days working with Reconnect Research in their offices in Los Angeles.

Burton Levine and Karol Krotki have pioneered the early development and evaluation of RICS surveys. With RTI internal research and development funds we fielded a RICS evaluation survey and the results are described in this poster.

2. Objectives

We examine bias using RICS data collection in measures that are commonly collected in public health surveillance. We accomplished this objective by fielding a RICS survey that mimicked questions from the National Health Interview Survey (NHIS).

3. Methods

We fielded an instrument that mimicked several questions from the NHIS Adult File (CDC, 2017). For the remainder of this poster, we will call this survey the RICS. We chose the 2016 NHIS Adult File as a comparison because it is a comprehensive national public health survey with a public use file, a high response rate (54.3%), and a large sample size (n = 33,028).

We used an interactive voice response system (IVR) for both recruitment and data collection. In the IVR implementation of the questionnaire, the respondent was given at most two opportunities to answer each question. Any telephone key pad press entered during the reading of the question was ignored by the IVR system. After each question is presented in its entirety, the respondent is prompted to provide a response using the telephone key pad. If a valid response option is provided any time during or after the reading of the response options, the IVR system stops reading additional response items and starts the next question. Invalid response options (e.g., pressing the 0, or 4–9 keys when only the 1, 2, or 3 keys are associated with valid options) are ignored. After the question and all response options are presented, and no valid response is forthcoming from the respondent, there is a 2-second delay before the question and response options are read a second time. If no valid response options are provided, the IVR system goes to the next question.

Reconnect Research redirected the inbound calls, recorded the IVR script, programed the IVR system and created the raw data file. We recruited 137,840 inbound calls from the 50 U.S. states and the District of Columbia to the RICS survey. Data were collected in two 1-week intervals in January and February of 2017. We screened each inbound call for adults (18 years old and older). A total of 24,735 (18%) recruited inbound calls responded to the screener. Among those who were screened, 21,998 (89%) indicated that they were an adult. Among these screened adults, 9,478 (43%) were considered respondents because they provided valid responses to 10 or more questions and were not considered straightliners. Among these valid respondents, 8,157 (86%) completed the entire questionnaire. The AAPOR4 response rate was 8.0%.

We created delete 1-group jackknife weights. For each group we initialized all respondents to an identical base weight and calibrated to the marginal distributions of sex, age category, race/ethnicity, educational attainment, and census division. We compared the unweighted demographic distributions to American Community Survey (ACS) data and used Behavioral Risk Factor Surveillance Survey (BRFSS) data as a comparison. The BRFSS data received a weight adjustment to account for the disproportionate allocation by geographical strata. We compared both categorical and continuous outcomes to the NHIS data. We evaluated the primacy effect by fielding two questions with two opposite ordering of the response categories. Inbound callers were randomized to an ordering for each question.

4. Results

We analyzed metadata on 10 million inbound calls—all redirected inbound calls to Reconnect Research between May 30, 2016, to April 12, 2017. For each state, we defined the absolute difference as the absolute value of the difference in the percentage of calls in that state and percentage of population in that state. The mean absolute difference is 0.27% (SD = 0.29). **Figure 1** displays the percent of the 10 million inbound calls by state and the percent of the population by state.

Figure 1: The Percentage of Inbound Calls and Population in Each State

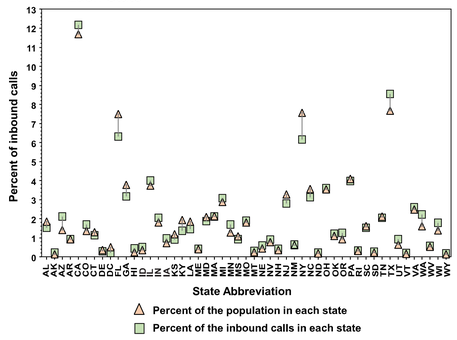

Table 1 contains a comparison of the demographic distributions among the state population, the unweighted RICS survey respondents, and the BRFSS survey respondents. Compared to the population, the demographic distributions of the RICS survey respondents are more often female, fewer are non-Hispanic white, more are African American, and they are less educated. Compared to the BRFSS, the demographic distributions of the RICS survey respondents match the 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. Overall, the RICS respondents match the state population almost as well as the BRFSS respondents. We speculate that the selection bias in the RICS study is minimal. If there were high selection bias, we would expect a much larger mismatch between the respondents and the population for the RICS study compared to the BRFSS.

Table 1: Comparison of the Demographic Distribution of State Population, the RICS Respondents, and the 2016 BRFSS Respondents

Distribution	Category	Population	RICS	BRFSS
Sex	Female	51.7	56.0	53.1
	Male	49.0	44.0	46.9
Age Category	18–24	12.4	12.8	8.4
	25–34	18.0	15.7	14.6
	35–44	16.3	15.3	14.3
	45–54	17.3	17.4	17.7
	55–64	16.7	17.5	20.0
	65 and older	19.9	21.3	25.1
Race/ Ethnicity	White	64.7	54.1	66.7
	African American	12.1	24.2	10.1
	Hispanic	15.8	14.2	13.9
	Asian	5.7	3.3	3.6
	Other race	2.4	4.2	5.7
Educational Attainment	Less than high school (HS)	13.6	17.3	9.0
	HS	28.0	38.2	25.2
	Some college	31.2	28.4	26.7
	Bachelor's or higher (BS)	27.1	16.2	39.1

Additionally, **Table 1** presents the demographic distributions of the BRFSS respondents after accounting for the study design and adjusting for the disproportionate allocation to geographical strata, as described above. Compared to the state population, the demographic distributions of the BRFSS respondents are more often female, are older, fewer are 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 night owls compared to the daytimers. It seems likely some individuals have a higher response propensity during the night than the day. Limiting data collection to 9 a.m. to 9 p.m. might be a source of bias in outbound telephone surveys that is avoided in RICS surveys.

Night Owls: Outbound telephone survey data are usually gathered between 9 a.m. and 9 p.m. in a respondent's local time zone. In contrast, with RICS one can gather data across the 24-hour daily cycle. We define night owls as survey respondents that 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." A total of 2,893 (30.6%) of the respondents are night owls and 6,572 are daytimers. **Table 2** displays the demographic distributions of night owls and daytimers. Compared to 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 differences were all statically significant. These results are suggestive that constraining data collection to 9 a.m. and 9 p.m. may result in bias.

Table 2: Comparison of Night Owls (People Who Respond between 9 p.m. and 9 a.m.) with Daytimers and BRFSS Respondents

Distribution	Category	Population	BRFSS	Night Owls	Day-timers	P-value
Sex	Female	51.7	53.1	50.2	58.5	<0.001
	Male	49.0	46.9	49.8	41.5	
Age Category	18–24	12.4	8.4	18.1	10.5	<0.001
	25–34	18.0	14.6	18.5	14.5	
	35–44	16.3	14.3	16.5	14.8	
	45–54	17.3	17.7	16.8	17.6	
	55–64	16.7	20.0	14.8	18.6	
	65 and older	19.9	25.1	15.3	24.0	
Race/ Ethnicity	White	64.7	66.7	50.0	55.9	<0.001
	African American	12.1	10.1	26.8	23.0	
	Hispanic	15.8	13.9	15.1	13.8	
	Asian	5.7	3.6	3.4	3.3	
	Other race	2.4	5.7	4.8	3.9	
Educational Attainment	Less than HS	13.6	9.0	18.3	16.7	<0.001
	HS	28.0	25.2	40.1	37.3	
	Some college	31.2	26.7	27.4	30.0	
	BS	27.1	39.1	14.3	17.0	

Table 3 displays the RICS estimates of 12 measures. All but two of these estimates are compared to their NHIS estimate; the marital status distribution is compared to estimates from the 2016 ACS 1-year summary file and the health insurance estimate is compared to an estimate from Gallup. 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 4.1. The p-value is based on testing the null hypotheses that the RICS distribution is the same as the comparison distribution. The p-value was significant at the 0.05 level for all but one distribution.

Table 3: Comparison of Estimates of Study Outcomes between RICS and the NHIS

Question	Category	Percent RICS	Percent NHIS	Absolute difference	Mean absolute difference	P-value
Ever smoked	Yes	38.8	37.0	1.8	1.8	0.004
Do you now smoke…	Everyday	34.5	30.9	3.5	4.8	<0.001
	Some days	13.7	9.9	3.7		
	Not at all	51.9	59.1	7.3		
Current smoker	Yes	18.9	15.1	3.7	3.7	<0.001
Quit attempt past 12 months	Yes	51.9	49.2	2.7	2.7	0.133
Marital status	Now married	42.7	49.5	6.8	5.9	<0.001
	Widowed	14.1	6.0	8.1		
	Divorced	16.1	11.0	5.1		
	Separated	3.9	2.0	1.9		
	Never married	23.2	30.9	7.7		
Internet use	Yes	75.0	78.2	3.2	3.2	<0.001
Health insurance	Yes	86.2	88.3	2.1	2.1	N/A
Ever used firearms	Yes	34.3	36.6	2.3	2.3	0.004
Full hours of sleep in a 24-hour periods	5 or less	18.5	9.4	9.2	4.7	<0.001
	6	28.3	22.6	5.7		
	7	24.2	29.2	5.0		
	8	19.3	30.5	11.1		
	9	4.2	4.4	0.2		
	10	2.8	2.7	0.1		
	11 or more	2.7	1.4	1.3		
Vigorous leisure-time physical activities—times per week	0	33.6	55.6	22.0	9.9	<0.001
	1 or 2	31.3	14.1	17.2		
	3, 4 or 5	24.9	20.6	4.3		
	6 or 7	5.4	8.2	2.9		
	8 or more	4.9	1.5	3.3		
Days drank alcohol per month	0	55.9	48.6	7.4	6.5	<0.001
	1–2	23.1	17.5	5.7		
	3–6	11.9	13.1	1.2		
	7 or more	9.1	20.9	11.8		
Drinks per day on days you drank	1	36.4	37.9	1.5	1.0	<0.001
	2	31.5	32.9	1.4		
	3	14.5	14.3	0.2		
	4	6.2	6.1	0.1		
	5 or 6	6.1	6.2	0.1		
	7 or more	5.4	2.7	2.7		

Primacy Effect (**Table 4**): For two questions, we randomized respondents to two different ordering of response options. In both questions, the ordering of the questions was associated with the outcome (p-value < .001). For the question about being worried about paying medical bills, when "very worried" was the first response option, 19.9% of the respondents reported that answer; whereas 13.7% of reported being "very worried" when the response options are reversed and it was presented last. We see an even larger effect, nearly 20 percentage point difference, for the question about change in health insurance.

Table 4: Evaluation of Primacy Effect: Comparison of the Effect of Reversing the Ordering of Response Categories

Question 1: How worried about paying medical bills			
Response category	Ordering 1 (% 95% CI)	Ordering 2 (% 95% CI)	Absolute difference
Very	19.9 (18.6, 21.3)	13.7 (12.4, 14.9)	6.3 (4.4, 8.1)
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			
Response category	Ordering 1 % (95% CI)	Ordering 2 % (95% CI)	Absolute difference
Better	27.9 (26.3, 29.5)	9.6 (8.4, 10.7)	18.3 (16.4, 20.3)
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)	20.0 (17.5, 22.4)

5. Conclusions

RICS sampling appears to be a promising tool for conducting population surveillance. The low cost and speed might make it a fit-for-purpose solution in many situations.

The following are the specific findings of this evaluation of RICS methodology;

1. The RICS frame members matched the state residence of the population closely.
2. The demographics of the RICS respondents match the population almost as well as the BRFSS.
3. Across 12 public health outcomes, the RICS estimates of proportions differ from the NHIS estimates by an average of 4.1%.
4. The demographics of the RICS survey respondents between 9 p.m. and 9 a.m. (night owls) look a lot like the underrepresented populations from outbound telephone surveys.
5. Large primacy effects were observed. We are not able to disentangle the sources of bias into selection bias, nonresponse bias, and measurement bias. However, we suspect that measurement bias makes a large part of the total bias because the demographic distributions of the respondents matched the population nearly as well as the BRFSS, and there is a large primary effect.

6. Potential for Impact

If we develop data collection methodology that enables us to field longer instruments, easily provide incentives, and reduces measurement error, RICS surveys have the potential to replace more expensive outbound telephone surveys and other probability surveys. Also, because of the low cost, it would expand population health surveillance to areas that were not explored.

In the evaluation we demonstrate that across 12 public health outcomes, the RICS estimates of proportions differ from the NHIS estimates by an average of 4.1%. Yeager and colleagues (2014) estimate that, on average, telephone surveys have 3% percent point absolute errors. We believe that we can reduce measurement error to 3%. And RICS surveys can have comparable quality with outbound telephone surveys.

New data collection methodology will need to be developed and tested for RICS to reach its full potential of being an alternative to outbound telephone public health surveillance surveys, such as the BRFSS. Some option for new data collection methodology is described in *Section 8: Next Steps.*

7. Limitations

For this evaluation, we used the NHIS as the comparison survey. We did not appreciate, at the time we created the instrument, the difficultly in mimicking the NHIS survey questions using an IVR instrument. We illustrate the differences between the NHIS and RICS instruments with the physical activity question: *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?* In the NHIS, the respondent can answer this 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 broke this question into categories, each category having a different time interval.

The reason why RICS is cheaper and faster than outbound telephone surveys is because RICS can recruit study participants and collect data with an automated system, whereas outbound telephone surveys require a large amount of time using telephone interviewers. However, the use of an IVR system for data collection is also a source of measurement error. For example, we observed large primacy effects. We saw 6% to 20% swings, in absolute terms, in survey estimates based on the ordering of the response categories.

8. Next Steps

We see two strategies for reducing bias in future RICS surveys. First, bias may be reduced by modifying the IVR implementation decisions. We can start by testing the effect of implementing the alternative decisions in **Table 5**. Second, we can collect data with a mode that has less measurement error than IVR, such as recruiting subjects to a web survey.

Table 5: IVR Implementation Decision and the Alternative

IVR Implementation Decision Applied	Alternative Decision
Entering data by depressing the keypad.	Having a conversational IVR in which the respondent responds with speech.
Providing a prefer-not-to-answer response option in each question.	Mentioning a prefer-not-to-answer response option once in 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 extra comforting language
Cadence of normal everyday speech	Slower or faster speech cadence than normal

Acknowledgments

RTI provided the funds for this evaluation.

Reconnect Research collaborated on the data collection methodology.

More Information

***Presenting author:** Burton Levine

RTI International
3040 E. Cornwallis Road, Hobbs 337, Research Triangle Park, NC, 27709

Phone: 919.541.1252 • **Email**: blevine@rti.org • **www.rti.org**