Editors and Researchers Beware: Calculating Response Rates in Random Digit Dial Health Surveys

Grant R. Martsolf, Robert E. Schofield,, David R. Johnson, and Dennis P. Scanlon

Objective. To demonstrate that different approaches to handling cases of unknown eligibility in random digit dial health surveys can contribute to significant differences in response rates.

Data Source. Primary survey data of individuals with chronic disease.

Study Design. We computed response rates using various approaches, each of which make different assumptions about the disposition of cases of unknown eligibility.

Data Collection. Data were collected via telephone interviews as part of the Aligning Forces for Quality (AF4Q) consumer survey, a representative survey of adults with chronic illnesses in 17 communities and nationally.

Principal Findings. We found that various approaches to estimating eligibility rates can lead to substantially different response rates.

Conclusions. Health services researchers must consider strategies to standardize response rate reporting, enter into a dialog related to why response rate reporting is important, and begin to utilize alternate methods for demonstrating that survey data are valid and reliable.

Key Words. Survey research, random digit dial, response rate calculation, survey methodology

When evaluating survey research, editors and reviewers often look for high response rates as evidence that the results are representative and valid. For example, 89 percent of journal editors consider response rates somewhat or very important when evaluating whether to publish an article that uses survey data (Carley-Baxter et al. 2005). Response rates are particularly important for random digit dial (RDD) surveys as rates for these surveys have been falling precipitously in recent years (Bunin et al. 2007). Because many health services researchers rely heavily on RDD surveys as an important source of data, there is significant incentive to report the highest possible response rate.

Although the calculation of response rates for many types of surveys is straightforward, calculation of response rates in RDD surveys is more complicated and can be influenced by factors that are not readily apparent. The main problem is adjusting the response rate for cases in which the eligibility of the sampled unit is unknown. Eligibility might be unknown because the phone was never answered after repeated calls or the respondent hung up before any information could be obtained. It is important to estimate how many of these cases are likely to be eligible. Most methods estimate the proportion of these cases that are eligible based on information about the cases whose eligibility could be determined.

The specifics of how eligibility is considered in response rate calculation formulas can lead to markedly different results (Smith 2009). The American Association for Public Opinion Research (AAPOR) (AAPOR, 2008) and the Council of American Survey Research Organizations (CASRO) (Centers for Disease Control and Prevention 2010) have developed the two most commonly used response rate calculation methods for RDD surveys. These methods are quite similar in many respects. However, they differ on how cases of unknown eligibility are treated. The AAPOR[1] method is likely to yield substantially lower rates than those calculated using the CASRO method. These differences are particularly important for health services researchers because the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) survey uses the CASRO response rate calculation technique, and BRFSS is often used as a standard by which response rates are judged in other RDD health surveys. The AAPOR method is the approach recommended by the major survey research and polling professional association.

Features of the sampling design used to generate the list of phone numbers to call might also affect the number of cases of unknown eligibility that will ultimately be considered in the final response rate calculations. Specifically, many survey sample vendors "prescreen" the samples that they sell to survey research centers. Common practices include choosing "blocks" of numbers with more known eligibles, using an automated system to precall numbers to detect a working tone, and screening out known business

Address correspondence to Grant R. Martsolf, R.N., M.P.H., Ph.D., Pennsylvania State University, Department of Health Policy and Administration, U.S. Steel Tower, 9th Floor, 600 Grant Street, Pittsburgh, PA 15219; e-mail: martsolfgr@upmc.edu. Robert E. Schofield, is with the Survey Research Center, Pennsylvania State University, University Park, PA. David R. Johnson, Ph.D., is with the Department of Sociology, Pennsylvania State University, University Park, PA. Dennis P. Scanlon, Ph.D., is with the Department of Health Policy and Administration, Pennsylvania State University, University Park, PA.

numbers. Some survey research companies and researchers will add those prescreened numbers back when estimating the proportion of phone numbers of unknown eligibility that are assumed to be eligible. However, these prescreened numbers are likely to have very different eligibility probabilities than cases of unknown eligibility. Including the cases that were removed by prescreening in the response rate calculations would likely underestimate the proportion of unknown eligibles that are considered to be eligible, leading to inflated response rates. When calculating or reporting response rates, researchers seldom mention, and may be unaware of, prescreening and its effect on the response rates.

No published articles in the health services research literature investigate the implications of using different approaches for handling cases of unknown eligibility in RDD surveys. Using two separate samples from the Aligning Forces for Quality Consumer Survey (Alexander et al. 2011, 2012), this study demonstrates that different approaches to handling cases of unknown eligibility contribute to substantial differences in response rates. First, we calculate response rates using both the AAPOR and CASRO methods. Second, we calculate response rates for samples where prescreened numbers are included and excluded in the final response rate calculations. This study also discusses implications for editors and researchers, as well as recommendations on how response rates should be reported.

DATA

The data for this study is the Aligning Forces for Quality Consumer Survey (AF4QCS), which was collected as part of the evaluation of the Aligning Forces for Quality (AF4Q) project, a $300 million initiative funded by the Robert Wood Johnson Foundation to improve the quality of care for persons with chronic diseases in 17 communities across the United States (Painter and Lavizzo-Mourey 2008). The populations of interest for this study were adults with one or more of five chronic conditions (asthma, diabetes, hypertension, coronary artery disease, and depression) residing in the 17 AF4Q communities. A national sample was also drawn from non-AF4Q communities. The AF4QCS samples were drawn with an RDD design. Telephone interviewing in the original 14 AF4Q communities and the national sample was performed between June 2007 and August 2008 by the Penn State Survey Research Center, whereas interviews for the remaining three communities were conducted from January to May 2010 by Social Science

Research Solutions (SSRS). The sampling design oversampled for racial and ethnic minorities in 13 of the 17 communities where there was sufficient racial/ethnic diversity.

Respondents first completed a 2-minute screener interview to determine his or her eligibility for the full survey. Eligible respondents in the screener interview were then invited to take the full interview and were offered a $20 incentive for completing the full interview. The completed sample size for the entire survey is 9,975 individuals.

METHODS

We computed response rates to the screener survey (rather than the full survey) for each of the 14 original communities using formulas provided by both AAPOR and CASRO using the sample that excluded prescreened numbers. Specifically, we used AAPOR response rate 4, one of six AAPOR response rate calculation methods. AAPOR4 was chosen because it allows for an estimation of the proportion of number of respondents of unknown eligibility that should be considered eligible. Most AAPOR methodologies do not make similar estimations. Instead, they either assume that all are eligible or all ineligible. We believe that assuming some are eligible is a more plausible assumption. More information on all AAPOR response rate techniques can be found in the AAPOR Standard Definitions (2011). For the final three communities, as per the contract, the survey vendor provided the AAPOR response rates to the screener survey. However, the vendor did not provide the CASRO rates to the screener survey, and the original disposition codes necessary to calculate those rates independently were not available.

The AAPOR4 and CASRO response rates were calculated using the formulas shown in Figure 1. The terms in the formula represent the following disposition codes: I = completed interview; P = partially completed interview; R = known eligible refusal; NR = known eligible nonrespondent; UR = unknown eligibility refusal; UNR = unknown eligibility nonrespondent; NH = ineligible nonhousehold. Importantly, AAPOR is an elaboration and extension of the original work to develop the CASRO methodology. The disposition codes used in this work are based on the AAPOR definitions and provide more precise disposition codes than were available when CASRO was originally formulated.

Examining these two similar equations, the key difference between the AAPOR4 and CASRO methods is how each treats cases of unknown eligibility

Figure 1: Response Rate Calculation Formulas

$$AAPOR4 = \frac{I + P}{(I + P + R + UR + NR) + e_{AAPOR4} * UNR}$$

$$CASRO = \frac{I + P}{(I + P + R + NR) + e_{CASRO} * (UNR + UR)}$$

$$e_{AAPOR4} = \frac{(I + P + R + UR + NR)}{(I + P + R + UR + NR + NH)}$$

$$e_{CASRO} = \frac{(I + P + R + NR)}{(I + P + R + NR + NH)}$$

Note. I, completed interview; P, partially completed interview; R, known eligible refusal; NR, known eligible nonrespondent; UR, unknown eligibility refusal; UNR, unknown eligibility nonrespondent; NH, ineligible nonhousehold.

(UNR and UR). These include dispositions in which (1) the number rings with no answer (UNR) or (2) the call is broken off at introduction (UR), usually referred to as "quick hang-ups." These unknown cases could be either eligible or ineligible for the screener survey, depending on whether the phone number actually reaches a household where adults permanently reside, but the actual disposition is not known and thus eligibility must be estimated in some way. The larger the number of cases with these unknown dispositions estimated to be eligible, the lower the response rate will be.

AAPOR4 considers all unknown eligibility refusals (UR) ("quick hang-ups") to be eligible noninterviews. The AAPOR4 formula enters these cases (UR) directly into the denominator so all of these cases are considered nonresponses. However, unknown eligibility nonrespondents (UNR) (e.g., the cases in which the phone rings with no answer) are handled differently. AAPOR4 considers only a portion of these respondents as eligible. The proportion of UNR cases that are considered eligible is calculated using the "e" formula presented in Figure 1. For the AAPOR4 calculation, *e* is calculated by taking the sum of cases that are considered eligible by AAPOR4 (I, P, R, UR, NR) and dividing that sum by those cases (I, P, R, UR, NR) plus the known ineligible nonhousehold cases (NH). This proportion is multiplied by UNR to obtain the number of UNR cases that are considered "eligible non-responders."

CASRO differs from AAPOR4 and assumes that only a proportion of UR, or "quick hang-ups" cases, should be considered eligible. To determine the total number of unknown eligibility cases that are considered eligible

nonrespondents, CASRO adds the UR and UNR cases and multiplies by the proportion of these cases considered to be eligible. Unlike AAPOR4, CAS-RO calculates "e" by taking the sum of cases that are considered eligible by CASRO (I, P, R, NR) and dividing that sum by those cases (I, P, R, NR) plus the known ineligible nonhousehold cases (NH). In essence, CASRO applies the same eligibility proportion to both UR and UNR cases.

Because AAPOR4 considers all UR cases as eligible, the AAPOR4 response rate is considered more conservative than the CASRO method. The AAPOR4 logic is that the UR cases should be regarded as eligible because a "quick hang-up" is very likely to be a household (eligible) because businesses (ineligible) would be unlikely to quickly hang up on unknown callers. Applying the same eligibility proportion to the "quick hang-ups" (UR) as to numbers that ring with no answer (UNR) would not be appropriate as it underestimates eligible households.

We also examined AAPOR4 response rates when prescreened numbers were both included and excluded from the calculations using the samples from all 17 communities. The sample of phone numbers in each community had been prescreened by the sample vendors. The numbers were prescreened by auto-dialing each to determine if the number was assigned, and also eliminating many business numbers based on comparison with the Yellow Pages.

The primary difference between the two response rates is based on how the prescreened numbers are treated. When these prescreened numbers are returned to the response rate calculations, they are considered to be ineligible nonhouseholds (NH). These numbers are added back into the denominator of the *e* formula. Therefore, when these numbers are added back in, fewer of the cases of unknown eligibility will be assumed to be actually eligible. This decreases the number of cases in the denominator of the response rate formula and, therefore, increases the overall AAPOR4 response rate. However, this assumes that the prescreened numbers are similar in some way to the numbers of unknown eligibility and can, therefore, be used to estimate the proportion of those cases that should be considered eligible. This may not be a reasonable assumption.

RESULTS

The response rates for the original 14 communities using both CASRO and AAPOR4 are shown in Table 1. The two response rate methods produce large differences in results for all the communities. The AAPOR method leads to markedly lower response rates than the CASRO method. The average

Table 1:   CASRO and AAPOR4 Response Rates for Screener Survey

Community	*AAPOR Response Rate, %*	*CASRO Response Rate, %*	*Total Percentage Point Difference*	*Percent Difference, %*
Seattle, WA	29.97	44.54	14.57	48.60
Detroit, MI	27.37	44.67	17.30	63.19
Memphis, TN	24.99	47.60	22.61	90.48
Twin Cities, MN	36.78	48.32	11.54	31.36
Western New York	27.21	45.40	18.20	66.88
Western Michigan	35.99	54.30	18.32	50.90
State of Wisconsin	32.37	50.78	18.42	56.90
State of Maine	29.38	45.31	15.93	54.22
Eureka, CA	35.59	51.69	16.09	45.21
York, PA	39.17	53.23	14.06	35.89
Cincinnati, OH	29.43	57.34	27.90	94.82
Cleveland, OH	36.71	58.04	21.33	58.11
Kansas City, MO	31.54	49.23	17.70	56.12
Willamette Valley, OR	34.26	48.70	14.44	42.17
Average response rate	32.2	49.94	17.74	55.1

AAPOR response rate across all communities is 32 percent, whereas the CAS-RO rate is 50 percent. This is an 18 percentage point difference yielding a 55 percent higher rate when the CASRO method is used. Substantial variation also exists in the differences between the approaches across the AF4Q communities. This suggests that a quick and simple way to translate an AAPOR rate into a CASRO rate is not feasible. For example, in the Twin Cities, the AAPOR4 response rate is 37 percent, whereas the CASRO response rate is 48 percent; only a 31 percent difference. For Cincinnati, however, there is a 95 percent difference in the rates.

The differences between AAPOR4 response rates when the prescreened numbers are excluded and included are much smaller than the differences between AAPOR4 and CASRO (Table 2). Overall, the average AAPOR4 response rate across communities based on data that excludes the prescreened numbers is 33 percent, whereas the rate based on the data that returns the prescreened numbers is 36 percent. However, there is significant variation across the 17 communities. Furthermore, the differences between the response rates are significantly bigger in the three communities that were surveyed later (January–March 2010). The differences between the response rates across the two administrations of the surveys might be explained by changes in technology over time. Differences between these response rates are driven largely by the proportion of the original sample that can be prescreened out before the phone numbers are called. The digital technology used to prescreen the

Table 2:   AAPOR4 Response Rates for Screener Survey with Prescreened Numbers Excluded and Included

Community	Prescreened Numbers Excluded, %	Prescreened Numbers Included, %	Total Percentage Point Difference	Percent Difference, %
Seattle, WA	29.97	30.42	0.45	1.50
Detroit, MI	27.37	30.49	3.12	11.40
Memphis, TN	24.99	26.59	1.60	6.40
Twin Cities, MN	36.78	37.49	0.71	1.93
Western New York	27.21	31.98	4.77	17.53
Western Michigan	35.99	38.43	2.44	6.78
State of Wisconsin	32.37	33.21	0.84	2.59
State of Maine	29.38	32.27	2.89	9.84
Eureka, CA	35.59	37.76	2.17	6.10
York, PA	39.17	40.74	1.57	4.01
Cincinnati, OH	29.43	29.69	0.26	0.88
Cleveland, OH	36.71	37.93	1.22	3.32
Kansas City, MO	31.54	32.36	0.82	2.60
Willamette Valley, OR	34.26	34.75	0.49	1.43
Albuquerque, NM	38.24	48.74	10.50	27.46
Boston, MA	29.84	41.95	12.11	40.58
Indianapolis, IN	40.70	50.03	9.33	22.92
Average response rate	32.91	36.17	3.25	9.88

second administration of the survey was better able to assess the presence of a dial tone than that used for the first administration, allowing a larger proportion of phone numbers to be prescreened out. The technology in this area is advancing rapidly.

DISCUSSION

Our analysis demonstrates that various methods of calculating eligibility rates in RDD surveys can lead to substantially different response rates. These findings suggest a number of important implications for health services researchers. First, health services journal editors should be aware that the different methods for computing response rates for RDD surveys can lead to very different reported response rates. If response rates are used as an important publication criterion, editors and referees must fully understand the implications of using different calculation methods when judging submitted research.

Of particular importance to health services research, BRFSS uses the CASRO method to calculate response rates. Response rates reported for

BRFSS are higher than they would otherwise be if another calculation technique, such as AAPOR4, was used. Therefore, BRFSS response rates should not be used as the sole standard for other RDD health surveys. Furthermore, we believe that the key assumption of the CASRO method (and the primary difference between CASRO and AAPOR4), that prescreened numbers are similar to the numbers of unknown eligibility and should therefore be considered eligible, is not a reasonable assumption. Therefore, we recommend that more conservative AAPOR4 be adopted as the standard response rate calculation technique for health services researchers.

Although less significant than differences between AAPOR4 and CASRO rates, whether prescreened numbers are returned to the sample can also contribute to differences between response rates. This issue is likely to become more acute as the technology used to prescreen numbers becomes more advanced and better able to prescreen out ineligible numbers. However, the issue of prescreening is likely underappreciated among health services researchers and those contracting with vendors for survey research. Researchers should know whether samples are prescreened by the sample vendor before using that sample to collect data, and should understand how this might affect reported response rates. Furthermore, health services researchers should be particularly aware of both of these issues when choosing a survey research center. When using response rates as a barometer of the quality or effectiveness of a survey research center, researchers should know how the response rates have been calculated and under what conditions.

Editors should develop standards for response rate reporting and disseminate these standards to authors. Presently, researchers ignore both CASRO and AAPOR4, and report ad hoc, and often undefined, response rate formulas (Johnson and Owen, 2003; Marshall 2006). There are rarely, if ever, written or even "unwritten" criteria for response rate reporting (Carley-Baxter et al. 2005; Johnson and Owen, 2003). Editors may wish to clarify what calculation technique should be used, if response rates are used as an important barometer of the "publishability" of a study. Also, given that response rates in telephone surveys are falling precipitously with the introduction of call screening and cellular phones, researchers and editors must determine what response rate levels would be considered acceptable. Some health services journals might expect response rates that are simply unobtainable in the present telephone survey climate. This provides further incentive to use a response rate calculation technique that would produce the highest possible reported rate.

To begin to deal with this dilemma, we suggest that researchers report RDD response rates in both formats (AAPOR4 and CASRO) because of

disagreement about which response rate is most "correct." However, because the disposition code definitions are slightly different between each approach, calculating these formulas may be challenging for many health services researchers. Furthermore, many survey research companies use proprietary disposition codes that do not track back well to either AAPOR4 or CASRO codes. This requires that health services researchers understand disposition codes and requires that survey research vendors provide disposition codes that can be easily cross walked to both CASRO and AAPOR4 codes. We also suggest that these codes, in certain circumstances, be made available to editors for review or even publication in an online appendix.

The standardization and clarification of response rate reporting should also be considered within a broader discussion of why researchers should care about response rates at all and what exactly a low response rate means. A growing body of literature suggests that response rates are poor indicators of bias in an RDD survey (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Groves 2006; Keeter et al. 2006). In fact, some of the techniques used to increase response rates, including incentives and aggressive callbacks, while increasing response rates may actually increase response bias as measured by comparison with demographic characteristics to the Current Population Survey (Keeter et al. 2000). Therefore, the health service research community might consider a shift away from a reliance on high response rates as a proxy for a high-quality survey and toward the evaluation of actual response bias.

In order for researchers to better estimate response bias, we suggest that, along with standardized response rates, they also include a preplanned assessment of nonresponse bias. Nonresponse bias can be assessed in a number of ways; all of which have advantages and disadvantages when applied to random digit dial surveys. Researchers may a priori choose sampling frames for which partial data is available for both responders and nonresponders (Groves and Peytcheva 2008). However, it is very difficult to obtain any information on unknown nonresponders in the case of RDD. One approach might be to use census data to compare the likely demographic characteristics of the responders and nonresponders based on their area code. Researchers may also choose to compare earlier responders to later responders. This information might then be used to make inferences about the likely responses of the nonresponders (Groves and Peytcheva 2008). Finally, researchers might compare the responses on similar items in other nationally recognized surveys such as the National Health Interview Study (NHIS). However, differences between surveys such as data collection methods and time frame may make comparisons between surveys not definitive.

Our analysis suggests that choice of calculation method and type of sample purchased can lead to substantially different response rates. Without clear reporting standards, the various techniques can be selectively chosen for purposes of meeting peer review and publication standards or expectations. Health services researchers must consider strategies to standardize response rate reporting, enter into a broader dialog related to why response rate reporting is important, and potentially begin to utilize alternate methods for demonstrating that survey data are valid and reliable.

ACKNOWLEDGMENTS

NOTE

1. Specifically, we used the AAPOR4 method, which is discussed in more detail below.

REFERENCES

Alexander, J. A., L. R. Hearld, R. Hasnain-Wynia, J. B. Christianson, and G. R. Martsolf. 2011. "Consumer Trust in Sources of Physician Quality Information." *Medical Care Research and Review* 68 (4): 421–40.

Alexander, J. A., L. R. Hearld, J. N. Mittler, and J. Harvey. 2011. Patient-Physician Role Relationship and Patient Activation among Individuals with Chronic Illness. *Health Services Research* 47(3, Pt 1): 1201–23.

American Association for Public Opinion Research. 2008. "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys" [accessed July 12, 2010]. Available at http://www.aapor.org/Standard_Definitions1.htm

Bunin, G. R., L. G. Spector, A. F. Olshan, L. L. Robinson, R. Rosler, S. Grufferman, X. Shu, and J. Ross. 2007. "Secular Trends in Response Rates for Controls Selected by Random Digit Dialing in Childhood Cancer Studies: A Report from the Children's Oncology Group." *American Journal of Epidemiology* 166 (1): 109–16.

Carley-Baxter, L. R., C. A. Hill, D. J. Roe, S. E. Twiddy, and R. K. Baxter. 2005. "Changes in Response Rate Standards and Reports of Response Rate over the Past Decade." RTI International. TSMII Conference Paper.

Centers for Disease Control and Prevention. 2010. "*Behavioral Risk Factor Surveillance System Summary Data Quality Report*" [accessed September 22, 2011]. Available at ftp://ftp.cdc.gov/pub/Data/Brfss/2010_Summary_Data_Quality_Report.pdf

Curtin, R., S. Presser, and E. Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64: 413–28.

Groves, R. M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–75.

Groves, R. M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72 (2): 167–89.

Johnson, T., and L. Owens. 2003. "Survey Response Rate Reporting in the Professional Literature." Paper presented at the American Association for Public Opinion Research Section on Survey Research Methods.

Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser. 2000. "Consequences of Reducing Nonresponse in a Large National Telephone Survey." *Public Opinion Quarterly* 64: 125–48.

Keeter, S., C. Kennedy, M. Dimock, J. Best, and P. Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70: 759–79.

Marshall, T. R. 2006. "AAPOR Standard Definitions and Academic Journals." *AAPOR – ASA Section on Survey Research Methods* [accessed July 12, 2010]. Available at http://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000126.pdf

Painter, M. W., and R. Lavizzo-Mourey. 2008. "Aligning Forces for Quality: A Program to Improve Health and Health Care in Communities across the United States." *Health Affairs* 27 (5): 1461–3.

Smith, T. W. 2009. "A Revised Review of Methods to Estimate the Status of Cases with Unknown Eligibility" [accessed September 2, 2012]. Available at http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions&Template=/CM/ContentDisplay.cfm&ContentID=1820

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.