HOW LOW CAN RESPONSE RATES GO?

Stephen A. Douglas, The Douglas/Jones Group
David Napior, Roper Starch Worldwide

Introduction

Response rates for many companies who measure newspaper and local market radio on an ad hoc or syndicated basis are in the 25% to35% range using AAPOR's (American Association for Public Opinion Research publication entitled Standard Definitions) strictest definition:

- MRC accredited SMM with 15%
- Nielsen had some very low levels in the mid teens in Miami

In the United States the random-digit dial (RDD) telephone interviewing methodology is used for many important syndicated and publisher-sponsored readership studies. One the most prominent syndicated studies, Scarborough, which measures newspaper and magazine readership in 75 local media markets uses RDD. It's average response rate is 38% and does have markets where is lower.

Also many individual newspapers in smaller markets and small newspapers in major markets have also conducted, or want to conduct RDD studies to estimate a total audience of their individual newspapers for their use in ad sales. The Audit Bureau of Circulations is currently setting up standards and audit procedures to insure that a common readership question is used and a standard set of demographic questions asked. ABC will also audit the studies to insure that the minimum standards are adhered to.

This process has drawn attention to the very low response rates that have been achieved. In a separate paper ABC will report the response rates for eight test markets. The response rates ranged from 20% to 40% with only one exceeding the desired standard of 40%. In the United States this decline in response rates for the RDD study has been an eight to ten year problem. It is being exacerbated by:

- the increase in the number of phone lines;
- the increase in answering machines;
- the increase in devices that allow people to screen out phone calls;
- the increase in the number of cell-phones; (in the US many cell phones are for outgoing only others are the primary phone part of the time.)
- increased frustration with respondents because of over surveying;
- a dramatic increase in telemarketing and SUGGING (selling under the guise of research)

Unchecked or, at minimum, unparried, the alarming RDD non-response trends we are witnessing today will, in the authors' opinion, undermine the confidence in all RDD studies and audience research in particular.

In this paper we will review the merits and demerits of various strategies for increasing our confidence in audience estimates produced with RDD sample selection and recruitment methodology in this climate of diminishing availability of potential respondents to telephone contact methods:

- Side-by-side comparison of RDD to probability postal and probability door-to-door sample selection and recruitment methods;
- Adoption of more vigorous RDD practices in order to fight off the non-response infection;
- Examine the potential of relatively recent developments in statistical inference theory and practice for protecting RDD studies from the potentially misleading statistical biases of large magnitude non-response when using traditional frequentist/probability-based statistical inference method.

Response Rates: Definition and Trends

The ABC experience has clearly shown that every research company had slightly different ways of calculating response rate. It is safe to say that many of the decisions made on how to classify the various responses was driven by the principle to make the response state as high as possible.

AAPOR has at least three (3) different definitions of response rate as well as definitions of recovery rate and cooperation rate, which are sometimes confused with response rate. The one used by the ABC is the most rigid.

Response rate for the ABC studies and the most rigid AAPOR definition is defined as the percentage of pre-designated eligible respondents who provide completed interviews. For those respondents who could not be contacted, the percentage must be estimated based upon the percent of eligible respondents who were contacted. The achieved rate must be clearly stated. Where the sample employs disproportionate stratification, separate response rates must be provided by stratum, and a weighted average must be provided for the total sample. The example presented in the following table will illustrate the calculation process.

Sample Disposition and Recovery Information:		
Telephone numbers assigned		**1150**
Contacted and established as eligible*		675
Respondent interviewed		500
Respondent not interviewed		175
Respondent never home	**35**	
Respondent refused/language barrier	**58**	
Household member refused (either before or after screening)	**55**	
Miscellaneous (respondent illness, blindness (infirmity, etc.)	27	
Contacted and established as ineligible		367
Non-household number	115	
Non-working number	201	
No eligible individual	51	
Eligibility not established		108
No answer (includes some telephone answering machines*)	30	
Busy line	53	
No adult after six calls	9	
Unreported (not attempted)	16	

(This is an old example. The categories in Bold are areas growing at alarming rates)

Of 1150 telephone numbers originally assigned, 1042 (i.e., 675 + 367) were contacted; 675 of these 1042, or 65%, were established as eligible. The response rate is therefore equal to the 500 respondents who were successfully interviewed, divided by the sum of the 500, plus the 175 eligible respondents who were not interviewed, plus 65% of the 108 respondents (70) whose eligibility was not established because they were never contacted.
Response Rate = 500 ÷ (500 + 175 + 70) = 67%

Possible changes in definition of response rates (or what is allocated where in the response rate definition)

The experiences with the test markets in the US lead us to consider some of the following modifications to the operational definition of response rate.

1. Ignore the non-English speaking for audience measures of English speaking publications by projecting to the English speaking population. The interviews for the audited study are conducted only in English. The ratio of non-English speaking, to English speaking would be established by the interview. The population estimate would be for the English speaking population. For example, 6% of the people called could not speak English. Therefore, the English-speaking universe would be 94% of the total population, and the reach would be calculated off that universe. This would also mean that non-English speakers would be removed from the response rate calculation.

2. If the respondents are not in the geographic definition of the market for the study (defined by county and zip) they should be removed from the numerator and the denominator. The proportion of the total non-respondents matching the proportion of interviews conducted outside the study's geographic area should be taken out of the denominator.

3. They have some markets where there are a substantial number of no telephone households. These are generally Native American reservations. We should probably deal with this situation by obtaining telephone household penetration data from the phone companies. This means we can only project to telephone households. The question: Do we always project to only telephone households or only when non-telephone research exceeds a specified percentage of the total?

4. How to handle unresolved phone numbers is a subject we hope is addressed by AAPOR/CASRO. Numbers called numerous times spread out over months should be considered ineligible and non-working. We should develop a standard for this situation.

CASRO, APPOR and other organizations called a major conference together on the non-response problem in early October, 1999 (after this paper was prepared) in the United States. This symposium is devoting almost two days to discussion of response rate problems. Many of these conferences will have some suggestions for how to improve response rates: one here will be The Wall Street Journal paper which proposes a non-response adjustment procedure that he developed by Virginia Cable, Dan Jennings, and Val Appel.

We know RDD response rates are probably too low right now and we have to turn this problem around!!! To do this we are suggesting managing this process before it manages us. We have to develop new techniques or have a systematic process or set of processes that will prove the non-response might not affect the outcome.

Side-by-side comparison of RDD to probability postal and probability door-to-door sample selection and recruitment methods

The primary goal of the side-by-side comparison is to assess the direction and magnitude of the bias in the audience estimates produced by high non-response RDD samples relative to the estimates rates produced by alternative methodologies with lower levels of non-response. The best possible finding would be that there is no significant difference between the inferences produced from high non-response RDD samples and inferences from low non-response alternative methods of sample development. Of course, this unlikely finding does not necessarily lead to the conclusion that all three are "correct," merely that they are consistent.

There are three possible examples of side-by-side studies we could do in media research. One could be a personal interview with a high response rate, the second could be a simultaneous method approach and the third way could be a high quality direct mail approach.

Using the U.S. newspaper local market situation as an example, theoretically we could commission MRI to increase the sample in one small, one medium size and one very large market. MRI uses a face-to-face personal interview, and uses a first time read yesterday readership question very similar to the ABC question. MRI's random probability household sample is one of the best in the world. If the RDD sample with a 25% response rate produced the same demographic composition that would give some confidence that the data was good. By definition we would expect the audience levels to be different because of the methods.

The weakness of this study and analysis is the possibility of a different outcome for each market. Then we are right back where we started.

The second approach is to conduct RDD studies in the same market using two different suppliers. Assuming they got the same answer, we would at least have replication as some form of validity. As frequent symposium attendees are sure to recall, this technique is also fraught with error because of the dreaded "contractor effect."

The third approach could be use of high quality direct mail procedures to determine if the RDD and the direct mail produce the same or different numbers. If the results are different and the Direct Mail results show a more up-market demographic profile, some could argue that the direct mail procedure, which would probably get a higher response rate in the United States, could be perceived as "the better data." We would still not know if it was better data.

The side-by-side approach, particularly the personal interview, may be the last resort to prove RDD and/or direct mail are an appropriate way to measure newspapers in the United States. One way or the other, sooner or later, some form of validation experiment will have to be developed.

Go After the Non-Responder

This tactic has been tried again and again. Barry Cook of Nielsen, James Peacock of Arbitron and many others have written extensively on this subject. Dick Lysaker's Hong Kong paper on multiple methods is a landmark in this kind of work. As you recall Dick used three kinds of methods: personal, phone and mail in all kinds of combinations to achieve a response rate of over 90%. In San Francisco, Berlin and Vancouver Julian Baim, Erhard Meir and others described how to improve response rate for personal interview studies.

This will always be an ongoing process. Some new tactic could increase the response rate or more likely to reduce the velocity of the response rate decline. Those tactics will always be the subject of one or two papers at a symposium. The odds are they will not be a cure all.

The ABC has adopted as a standard a set of RDD procedures designed to minimize non-response:

- At least six attempts must be made to contact all pre-designated respondents. For telephone interviews, each time a number is dialed the phone must ring at least six times before the attempt is classified as a "no answer." To the extent feasible, callbacks to reach numbers that do not answer (including answering machines) must be made at different times of day and on different days of the week. In this regard, the research firm must keep individual records of success and explanations of failure.

- At least one call attempt may be made on Sunday, when respondents who are unavailable during the week might be present. For daily newspapers, these Sunday callbacks would use Friday as a "yesterday" surrogate.

- An eligible respondent is defined as an adult living in a residence within the survey area. When an interview attempt is frustrated by a telephone answering machine, sometimes it is obvious that the number reached is a residence and sometimes it is not. When it is obvious from the content of the machine's outgoing message that the number is a residence, and repeated attempts fail to reach a live person, the respondent who should have been interviewed must be treated as an eligible respondent who has refused. (See the American Association for Public Opinion Research publication entitled Standard Definitions ...page 26, published in 1998.)

Other telephone answering machine contacts must be classified either as "ineligible" or "eligibility unknown" depending upon the content of the outgoing message.
Although the response rate should be as high as possible, ABC requires a minimum 40%. Should a given study fail to achieve this required rate, it may be approved if ABC determines that:

1. All reasonable actions were taken to achieve the 40% minimum

2. The results of the study reflect the realities of the corresponding market. Any newspaper denied approval may appeal this decision to an ABC committee composed of research professionals.

That is ok, but with the drops in response rate we will continue to have data buyers and sellers may not have confidence in. If the average is below 30% we are probably in a lot of trouble.

Reducing vulnerability of RDD to high levels of non-response using model-assisted statistical inference

Inferences from RDD samples infected with high levels of non-response can benefit from application of non-traditional, model-assisted statistical inference methods developed by survey practitioners for situations where the probability sampling requirements underlying the traditional frequentist school of inference cannot be satisfactorily attained (Rubin, Sarndoff, etc.)

In this approach, the randomization aspect of RDD is still retained since it still provides some protection against unwitting selection bias by the surveyor. However, the randomization probabilities play no role in the inference. This is useful because the high potential for respondent self-selection inherent in a high non-response methodology seriously undermines the validity of the basic assumption underlying frequentist inferences that the probability of selection for every member of the target population is "known."

In model-assisted inference, sometimes called the extended probability sampling approach, frequentist design-based reasoning is used to estimate sampling error and model-based reasoning to handle non-sampling error. When non-sampling error, in the form of frame imperfections and non-response/self-selection bias, outweighs sampling variance in terms of contribution to total survey error, the model-based, non-probability "extension" wags the design-based, probability dog.

The paper presented by Virginia Cable, Dan Jennings and Val Appel at this same symposium is very much in the spirit of the model-assisted school of inference, even though not directly acknowledged as such. The Non Response Adjustment procedure suggested by the Cable/Appel paper holds great promise for direct mail studies that can be matched up to huge data bases and for phone studies with a large number of home phone numbers that can be identified. The "supermodel" they are anchoring their data to is the Acxiom database.

Unfortunately the NRA adjustment approach does not hold much promise for business-to-business studies attempting to identify say computer purchase influencers. However, It should be noted that IntelliQuest is working with Val to see if he can replicate the WSJ success it has had for direct mail studies for their RDD sample of in-home computer purchase influencers.

Advocates of fully model-based inference advocate more extreme modification of the RDD methodology, by appending quota and balancing mechanisms to the randomized telephone number frame. Quota and balanced sampling can, in the right hands, be quite sophisticated non-probability sampling methods when "control" variables with known population totals are available to guide the sample selection process and the development of the underlying "superpopulation" model upon which the inferences rest.

Another example of a fully model based inference approach in media research is the IntelliQuest CIMS Business Computer purchase influencer study in the United States. IntelliQuest, measures magazine readership of computer technology purchase influencers. IntelliQuest uses RDD for recruitment only; also not all recruitment is done by RDD. More than half of the business purchase influencers are recruited using non-probability sampling methods based on D&B lists of establishments, then going through switchboard and snow-balling from initial switchboard referrals**.** In this instance the RDD survey is used to build the "superpopulation" model for the Business purchaser influencer survey, which is based on a combination of RDD and quota, based sampling.

The appropriate control variables essential to a good superpopulation model for newspaper audience measurement seem to be readily identifiable and their population totals are available from MRI. The authors are going to recommend the Audit Bureau of Circulations look at this procedure for possible way to handle the declining response rates of total audience studies in the future.

Things to Think About!

- Side by side studies using different methods. May have to use as a last resort.
- Go after non-responders. May have exhausted all of our tactics. New tactics could be limited.
- Identify the non-responder correct through weighting.
- Balance to high quality control data from external sources and use RDD with quota sampling and model based statistical inference. This might be the best course of action, but it might still require a validation step.
- Re-examine the AAPOR definition. This is being done, but the response rate definitions should be changed only if it will not hurt the quality of the data.

We do not have all the answers. We would like to debate these and any other solutions that emerge at the symposium now.

Thanks for your time and attention.