Are Callbacks a Waste of Time?

Philip Gendall and Peter Davis

This research investigated whether the same inferences would have been made from a survey without callbacks as with them. Comparison of the characteristics of respondents contacted on the first call with those contacted on subsequent calls over a range of behavioural and demographic variables confirmed the respondents who are not at home at an interviewer's first attempt are more likely to be in regular, paid employment, to be younger, male, better educated and to earn higher incomes. They also tend to watch less television, listen to the radio less and attend the cinema less often. However, for the behavioural and demographic variables in the study, none of the differences between estimates based on the first call sample and the total sample are significant and, for many practical purposes, the differences which do exist could be regarded as irrelevant. But although these differences may be small, they may be important, depending on the decisions which are based on them (for example, small differences are important when survey proportions are applied to population figures to determine the cost-per-thousand estimates which influence the purchase of media time and space). Consequently, the answer to the question, "Are callbacks a waste of time?", depends ultimately on how accurate the results of a survey need to be. This in turn depends on the nature and purpose of the study and on how the results will be used.

Keywords: callbacks, interviewing, survey research, non response bias

Introduction

Non-response bias is a potential problem in any sample survey, and numerous studies have documented differences between respondents contacted on the first and subsequent attempts (Stroeven 1981; Ward, Russick & Rudelius 1985; Vorster & Frankel 1987; Robinson & Lifton 1991). In surveys of the general public, respondents contacted on the first call tend to be older, less educated, female, have lower incomes and are more likely to live in rural areas.

To reduce these biases, researchers typically make several follow-up calls, or callbacks, on sample members who cannot be interviewed at the first attempt. The number of callbacks made before abandoning a prospective respondent is often two, but may be as many as five or six.

However, callbacks are expensive and time-consuming, particularly for face-to-face surveys, and this has led researchers to examine other methods of reducing non-response bias. These methods involve weighting the responses of those interviewed at the first attempt to compensate for respondents who cannot be interviewed. The weighting factors used include the reciprocal of respondents' reported probability of being at home, and demographic characteristics such as age, sex, and household size.

The problem with weighting is that it assumes attitudes, behaviour and other variables of interest are correlated with the variables used for weighting. If they are not, weighting can increase rather than reduce non-response bias. But the real issue is not the validity of weighting; it is whether the **same** inferences would have been made from a survey **without** callbacks as with them.

This research note reports a study which investigated this question using data from a face-to-face survey involving up to two callbacks on "not-at-homes".

Method

The data for this study came from the Palmerston North Household Omnibus, which is conducted annually by students from the Marketing Department of Massey University. The survey sample is based on clusters of four interviews (two with males, two with females, 15 years of age or older) around randomly selected starting points. Two callbacks are made before a selected household is abandoned and replaced. The 1989 Omnibus survey on which this study was based involved 667 completed interviews and had a response rate of 57%. Two thirds of the respondents interviewed were contacted on the first attempt.

The analysis involved three phases. First we compared demographic and behavioural characteristics of respondents contacted on the first and on subsequent calls to establish whether there were any significant differences. Then we weighted the first-call sample and the total sample so that their age-sex distributions matched that of the Palmerston North population. Finally, we compared the weighted estimates obtained from those contacted on the first call with weighted estimates obtained from the whole sample.

Results

Comparison of the characteristics of respondents contacted on the first call with those contacted on subsequent calls over a range of behavioural and demographic variables confirmed what other researchers have already shown. Respondents who are not at home at an interviewer's first attempt are more likely to be in regular, paid employment, to be younger, male, better educated and to earn higher incomes. They also tend to watch less television, listen to the radio less and attend the cinema less often (see Table 1).

The use of a sex quota in the survey resulted in 7% of potential 'first call' respondents not being interviewed because an interviewer's male or female quota was filled. It is probable that a high proportion of these people were women. Consequently the figures shown in Table 1 almost certainly underestimate the differences that would have been observed without a sex quota.

However, the practical question is not whether there are differences between first call and subsequent call respondents, we know there are, but whether this known bias can be corrected so that callbacks can be avoided. To examine this, the first call and total samples were both weighted so that their age-sex distributions matched the 1986 census figures for Palmerston North. Then the behavioural and demographic estimates derived from these weighted samples were compared. The results of this exercise are shown in Table 2.

None of the differences between estimates based on the first call sample and the total sample are significant and, for many practical purposes, the differences which do exist could be regarded as irrelevant.

Table 1. Unweighted First Call and Subsequent Call Results

Characteristic	Respondent Contacted On		Difference
	1st Call (n = 439)	2+ Call (n = 228)	
	%	%	%
Main grocery shopper	67	71	+4
Attended cinema in last 6 months	44	42	-2
Watch TV2 most	21	23	+2
Labour Party supporter	21	27	+6
CD player ownership	19	19	0
Female	53	47	-6
Regular, paid employment	55	61	+6*
Aged 60 or older	25	17	-8*
Some tertiary education	38	43	+5
Income $50,000+	33	28	+5
	Mean	**Mean**	
TV viewing (hours/week)	20.0	18.2	-1.8**
Radio listening (hours/week)	21.3	18.6	-2.7
Cinema attendance last 6 months	1.6	1.3	-0.3
Household size	3.1	3.1	0

* Difference significant at p<.10.
** Difference significant at p<.05.

Table 2. Weighted First Call and Total Sample Results

Characteristic	First Call Sample (n=439)	Total Sample (n=667)	Difference
	%	%	%
Main grocery shopper	64	66	+2
Attended cinema in last 6 months	49	47	-2
Watch TV2 most	26	26	0
Labour Party supporter	20	22	+2
CD player ownership	21	21	0
Regular paid employment	58	58	0
Some tertiary education	40	43	+3
Income $50,000+	27	27	0
	Mean	**Mean**	
TV viewing (hours/week)	19.8	19.3	-0.5
Radio listening (hours/week)	21.3	20.7	-0.6
Cinema attendance last 6 months	1.9	1.7	-0.2
Household size	3.2	3.2	0

Discussion

Before leaping to the conclusion that call backs are a waste of time, we need to examine these results and their implications a little more closely. In our survey, 66% of respondents were interviewed on the first call; this is consistent with other New Zealand omnibus surveys using similar sampling methods (see Stroeven 1981). Under these circumstances it is probably not surprising that estimates based on the first call sample are very similar to those based on the whole sample.

However, in a recent American Advertising Research Foundation (ARF) study which analysed the effect of call backs in six large surveys, the proportion of the total sample represented by first call respondents ranged from 37% to 54% (Opatow 1991). For these surveys the differences between results based on all calls and those based only on respondents reached at the first call were mostly of the order of 1% or 2%. This suggests that, even when the call back sample is relatively large in relation to the total sample, accurate estimates can be made from information collected only from respondents interviewed at the first attempt.

Nevertheless, it is also apparent from this study and the ARF study that for some variables related to the distinguishing characteristics of "not-at-homes", estimates based on first call respondents only will be different to those which would have been obtained if callbacks had been done to include "not-at-homes" in the sample, even after weighting for known biases. Furthermore, although these differences may be small, they may be important, depending on the decisions which are based on them (for example, small differences are important when survey proportions are applied to population figures to determine the cost-per-thousand estimates which influence the purchase of media time and space).

Consequently, the answer to the question, "Are callbacks a waste of time?", depends ultimately on how accurate the results of a survey need to be. This in turn depends on the nature and purpose of the study and on how the results will be used.

For many business surveys and public opinion surveys, the non-response bias introduced by not doing callbacks is likely to be relatively unimportant, particularly if the sample has been weighted to correct for known deficiencies in its characteristics. This is because the increase in accuracy as a result of callbacks is too small to make a meaningful difference to the inferences drawn. In such cases weighting of first call data is likely to be an acceptable substitute for callbacks. In fact, because weighting significantly increases the variance of estimates, it can be argued (see Ward et al 1985) that the small gains in accuracy it achieves are not worth the effort involved and that unweighted first-call estimates might just as well be used.

However, there **are** surveys for which the level of accuracy required is such that the difference made to estimates by callbacks can be meaningful. Media analysis is one such situation which has already been mentioned, but the same argument applies to any survey in which small percentage differences are important or where important subgroups are likely to be under-represented among first-call respondents. Thus callbacks are almost certainly justified in omnibus surveys because of the varying requirements of their multiple clients.

Conclusion

For many ad hoc surveys, callbacks on "not-at-homes" may not provide more useful information than estimates based on "at home" respondents. This is particularly so if the sampling method used involves quotas and procedures to maximise the possibility of reaching "mobile" respondents on the first call. However, this does not mean that non-response bias can be ignored in ad hoc surveys. At the very least, researchers should consider the likely impact of known response patterns on their survey results. Furthermore, if small differences in survey estimates are important, or if important subgroups are likely to be under-represented among first-call respondents, then there is no substitute for callbacks to reduce the potential for non-response bias.

References

Opatow L (1991). Some thoughts about how interview attempts affect survey results. *Journal of Advertising Research*, February-March, RC6-RC9.

Robinson L & Lifton D (1991). Reducing market research costs: deciding when to eliminate expensive survey follow-up. *Journal of the Market Research Society*, *33* (4), 301-308.

Stroeven R (1981). The effect of call-backs. *Australian Marketing Researcher, 5* (2), 37-39.

Vorster H & Frankel L (1987). Are all those call-backs really necessary? In H Henry (Ed.), Readership research: Theory and practice (pp. 227-231). Amsterdam: Elsevier Science Publications BV.

Ward JC; Russick B & Rudelius W (1985). A test of reducing callbacks and not-at-home bias in personal interviews by weighting at-home respondents. *Journal of Marketing Research, 12,* 66-73.

Philip Gendall is Professor of Marketing and Head of Department, and Peter Davis was a third year student, in the Department of Marketing, Massey University.