IN TRYING TO GET MORE ARE WE GETTING LESS ?

Ramesh Thadani & Katy Merchant, Indian Market Research Bureau (IMRB)

Introduction

Several experiments have been conducted to study the effects of non-response on readership estimates. Most of these experiments have focussed on the relatively small group of 'Hard-To-Get' (HTG) respondents in the main NRS, who perforce act as the surrogates for the non-responders.

So far, there is no consensus on the impact of non-response on readership estimates.

Statistical techniques such as the models developed by Hans Vorster (Audit & Surveys, Inc. Toronto), presented at the Readership Symposium, San Francisco, determined the significance and the direction of changes in reading probability, as a result of including or imputing the demographic characteristics of non-responders which most significantly differentiate the reading probability of a specific magazine.

This study concluded that non-response does reduce the probability of reading and hence the AIR of the publications.

On the other hand, work done by Messrs Vorster and Frankel reported at Salsburg, that since HTG respondents constituted a small percentage of the total sample, their contribution to overall reading scores was small and did not affect universe profiles.

Upto the last NRS conducted in India, the sampling method did not provide for substitutes as surrogates for the originally selected sample. In fact, to be more specific, the sampling method followed for NRS IV (1990) was as follows:

a)   A set of cluster heads or 'starting addresses' was selected from the electoral rolls using a systematic sampling method.

b)   From each starting address, seven more addresses were selected using a smaller sampling interval.

c)   Each starting address was visited and one adult respondent was interviewed for the main readership interview. This respondent was selected using the Kish method.

The above method ensured that every single household contacted was selected from the electoral rolls. Also, using a smaller sampling interval for selection of addresses in step (b) mentioned above, gave an element of fieldwork efficiency by reducing the interviewer's commuting time between successive calls.

Starting addresses which for whatever reason, could not be traced or located, were not 'substituted'. Also, if after three visits, the selected respondent for the main readership interview was not available or refused an interview, he/she was not 'substituted'. Both these factors resulted in a non-response rate which for NRS IV (1990) was 30%, on a national basis.

In the design for NRS 1995, this issue was closely examined by the technical group from the research agencies conducting NRS 1995, and a design change was recommended. The key factors prompting the design change were:

(a)   A considerable amount of fieldwork time and cost is incurred in locating addresses and physically visiting them. Therefore, incremental costs for substituting households or individuals is relatively small.

(b)   A 30% non-response rate always leaves a question mark in the user's mind about the profile of these respondents. Notwithstanding the fact that substitution is no definite solution to this problem.

(c)   High non-response rates particularly affect smaller publications - in other words the unweighted sample available for cross-tabulations or other analyses such as duplication, is small.

The design changes recommended were as follows :

a)   While cluster heads were selected as in NRS IV from the electoral rolls, the remaining addresses in the cluster were not selected from the electoral rolls. Instead, a specific rule was followed - namely the selected household and four consecutive households were to be interviewed.

        In moving from the starting household to the other consecutive household the movement method was the `right-hand-rule'.

b)   If a household was locked (after 3 visits), or refused to participate, then the sixth consecutive household was contacted.

c)   If in a household, the selected respondent was not available for an interview (after 3 attempts), then an interview was conducted with an individual with a matching profile in the sixth or seventh consecutive household. The profile was matched in terms of socio-economic class, age and sex.

In this paper, we are focussing on how the substitute sample varies from the original sample and making an assessment whether, in the process of "trying to get more information by having a substitute sample", are we losing out or making some unacceptable trade-offs ?

The NRS 1995 fieldwork is completed and data processing is currently in progress. This paper is based on an analysis of the readership survey conducted in one major metropolitan city in India - namely Madras, with a total sample of 3,492 individuals.

The findings of our study are discussed in detail in the sections that follow.

1. Are there differences in response rate by respondent type ?

        All clusters of households were classified into three broad groups :

        Predominantly

-       socio-economic class A1, A2, B1

-       socio-economic class B2, C

-       socio-economic class D, E1, E2

As expected, based on past surveys in India, the upper social classes are more difficult to interview. The non-response rate and, therefore, the percentage of substitutes in the first group is considerably higher than the average.

Non response rate by Socio-Economic class

TABLE 1 (a)

	A1/A2/B1	B2/C	D/E1/E2	Total
(Unweighted sample)				
Original sample	431	1370	892	2693
Substitute sample	194	406	199	799
Total sample	625	1776	1091	3492
Substitute sample as % of total sample	31.04%	22.86%	18.24%	22.88%

At a later point in the paper, we will discuss the full impact of the differential response rate in the various social classes.

Tables 1 (b) and 1 (c) give the sex and age distributions for the original sample and the substitute sample. While there is no substantial difference in the age distribution of the two samples, the substitute sample does have a slight skew towards women. This is attributed to a higher non-response rate among women in the original sample, perhaps, a reluctance to talk to strangers among conservative women.

Non response rate by Sex

TABLE 1 (b)

	Original Sample %	Substitute Sample %
Males	49.0	45.9
Females	51.0	54.1
Total	100.0	100.0

Non response rate by Age

TABLE 1 (c)

Age (years)	Original Sample %	Substitute Sample %
15 - 19	10.4	9.3
20 - 24	13.2	13.9
25 - 34	29.6	32.2
35 - 44	20.5	20.3
45+	26.4	24.4
Total	100.0	100.0

2. How different are the originally selected respondents from the substitutes in terms of readership and key media related parameters ?

To answer this question, let us first look at what our estimates of selected readership related parameters would have been in a `no substitution' scenario vis-a-vis a `with substitution' scenario. For both the situations, we have followed a comparable weighting procedure, i.e. a post-facto stratification by age (5 cells) x sex (2 cells) - a 10 cell matrix.

Further, for the no substitution scenario, we have also used a cluster response level weight similar to the procedure followed in NRS IV. In the `with substitution' scenario, this step would be redundant.

A comparison of the `original' sample vis-a-vis the total sample

TABLE 2

	Original Sample Only %	Original Sample + Substitute Sample %	Index : Original Sample = 100
Key Readership parameters :			
Read any :			
English daily	16.4	17.1	104
Tamil daily	33.6	32.3	96
English magazine	19.0	19.1	101
Tamil magazine	39.3	38.1	96
Any Englishpublication	22.0	22.2	101
Any Tamil publication	48.7	48.0	99
Average Issue Readers of :			
One publication	13.7	13.8	101
Two publications	10.7	9.8	92
Three publications	6.2	6.6	106
Four publications	5.3	5.3	100
Five or more publications	19.5	18.7	96

TABLE 2 (cont'd)

Average Issue Readers of :	Original Sample Only %	Original Sample + Substitute Sample %	Index : Original Sample = 100 %
Top Five Tamil Publications			
* Daily Thanthi	27.7	26.6	96
* Kumudam	15.6	15.1	97
* Ananda Vikatan	13.1	12.7	97
* Junior Vikatan	8.5	8.3	98
* Malai Malar	8.4	7.8	93
Average Issue Readers of :			
Bottom Five Tamil Publications			
* Muththaram	2.8	2.8	100
* Idhayam Pesukirathu	2.5	2.4	96
* Kalaimagal	2.4	2.4	100
* Rani Comics	2.0	2.2	111
* Malai Mathai	1.7	1.9	112
Top Five English Publications			
* Hindu	14.8	15.4	104
* Hindu (Weekly)	13.1	13.6	104
* India Today	6.4	6.2	97
* Readers Digest	5.0	5.1	102
* Indian Express	4.4	4.9	117
Bottom Five English Publications			
* Chandamama	1.8	1.7	94
* General Knowledge Today	1.7	1.9	112
* The Week	1.7	1.6	94
* Aside	0.7	0.7	100
* BBC Worldwide	0.4	0.5	125

It is quite evident from the above table that differences in readership related parameters are extremely small, almost non existent in several cases. We, however, do notice a slight skew towards English publications in the 'Original + Substitute' sample, which we hypothesise is due to the substitutes having a slightly higher SEC profile (i.e. higher social class groups).

Table 3 shows the differences in the two samples on Television ownership and viewership related parameters.

A comparison of the 'original' sample vis-a-vis the total sample

TABLE 3

	Original Sample Only %	Original + Substitute Sample %	Index : Original Sample = 100
Key TV related parameters			
Own a TV set	71.7	71.4	100
Own a Colour TV set	25.8	25.5	99
Access to Cable & Satellite channels	16.3	16.7	102
TV viewing frequency			
Every day	74.7	73.4	98
4 to 6 days a week	2.9	3.0	103
1 to 3 days a week	13.2	14.2	108
less often	9.2	9.4	102

Even on TV ownership and viewing related parameters the differences are very small.

Since the two sample sets share a 77% common sample, another way of looking for differences, with a sharper focus, would be to only compare the original sample with the substitute sample. For this comparison, we have used a common weighting procedure i.e. the 10 cell (age x sex) matrix. Since this implies a difference in weighting procedure as compared to the earlier analysis for the original sample (which has a cluster response level weight), the results for the original sample shown in this analysis set will not be identical with the analysis shown earlier. The comparison should only be made for the original sample vis-a-vis the substitute sample. These analyses for readership related parameters and TV ownership/viewership parameters are shown in Tables 4 and 5.

A comparison of the 'original' sample vis-a-vis the substitute sample

TABLE 4

Key Readership parameters	Original Sample Only %	Substitute Sample Only %	Index : Original Sample = 100
Read any :			
English daily	15.6	22.1	142 *
Tamil daily	33.0	29.7	90
English magazine	18.0	22.7	126 *
Tamil magazine	38.3	37.6	98
Any English publication	20.8	26.7	128 *
Any Tamil publication	48.1	47.4	99
Average Issue Readers of :			
One publication	14.0	13.3	95
Two publications	10.1	8.8	87
Three publications	6.2	7.7	124 *
Four publications	5.0	6.5	130 *
Five or more publications	18.6	20.5	110
Average Issue Readers of :			
Top Five Tamil Publications			
* Daily Thanthi	27.5	23.6	86 *
* Kumudam	14.8	16.0	108
* Ananda Vikatan	12.4	13.8	111
* Junior Vikatan	8.5	7.5	88
* Malai Malar	8.1	7.0	86
Bottom Five Tamil Publications			
* Muththaram	2.7	2.9	107
* Idhayam Pesukirathu	2.4	2.4	100
* Kalaimagal	2.4	2.4	100
* Rani Comics	1.8	3.3	183 *
* Malai Mathai	1.7	2.6	153
Average Issue Readers of :			
Top Five English Publications			
* Hindu	14.2	19.3	136 *
* Hindu (Weekly)	12.6	17.1	136 *
* India Today	5.9	7.3	124
* Readers Digest	4.7	6.8	145 *
* Indian Express	4.1	7.7	188 *
Bottom Five English Publications			
* Chandamama	1.6	2.1	131
* General Knowledge Today	1.7	2.7	159
* The Week	1.6	1.7	106
* Aside	0.6	1.1	100
* BBC Worldwide	0.4	0.7	175

(* = differences are significant at 95% confidence level)

A comparison of the `original' sample vis-a-vis the substitute sample

TABLE 5

	Original Sample Only %	Substitute Sample Only %	Index : Original Sample = 100
Key TV related parameters			
Own a TV set	70.3	75.0	107
Own a Colour TV set	24.1	30.0	124 *
Access to Cable & Satellite channels	15.4	21.2	138 *
TV viewing frequency			
Every day	73.1	74.4	102
4 to 6 days a week	3.1	3.6	116
1 to 3 days a week	14.0	14.9	106
less often	9.8	7.1	72

(* = differences are significant at 95% confidence level)

The analyses in Tables 4 and 5 do bring into focus the differences between the substitute sample and the original samples and some of the differences are statistically significant. We hypothesise that these differences are primarily driven by the Social Class of the `substitute' sample as compared to the `original' sample.

3. What happens if we adjust for differences in Social Class for the substitute sample ?

To confirm this hypothesis, we conducted another set of analysis. This analysis involved the inclusion of social class into the weighting procedure. For this, we used the observed SEC distribution for the total sample.

The effect of this analysis is shown in Tables 6 and 7 below in an indexed form for the substitute sample, both with and without the SEC weight.

Impact of Weighting by SEC (Index : Original Sample = 100)

TABLE 6

	Without SEC Weights	With SEC Weights
Key Readership parameters		
Read any:		
English daily	142 *	118 *
Tamil daily	90	89
English magazine	126 *	107
Tamil magazine	98	97
English publication	128 *	109
Tamil publication	99	95
Average Issue Readers of:		
One publication	95	96
Two publications	87	85
Three publications	124 *	112
Four publications	130 *	118
Five or more publications	110	98
Average Issue Readers of:		
Top Five Tamil Publications		
* Daily Thanthi	86 *	86 *
* Kumudam	108	99
* Ananda Vikatan	111	99
* Junior Vikatan	88	79
* Malai Malar	86	88

(*= differences between the original sample and the substitute sample are significant at 95% confidence level)

TABLE 6 (cont'd)

Key Readership parameters	Without SEC Weights	With SEC Weights
Average Issue Readers of:		
<u>Bottom Five Tamil Publications</u>		
* Muththaram	107	107
* Idhayam Pesukirathu	100	84
* Kalaimagal	100	84
* Rani Comics	183 *	174 *
* Malai Mathai	153	147
Average Issue Readers of:		
<u>Top Five English Publications</u>		
* Hindu	136 *	112
* Hindu Weekly	136 *	111
* India Today	124	105
* Reader's Digest	145 *	116
* Indian Express	188 *	157 *
<u>Bottom Five English Publications</u>		
* Chandamama	131	113
* General Knowledge Today	159	147
* The Week	106	128
* Aside	183	128
* BBC Worldwide	175	140

(* = differences between the original sample and the substitute sample are significant at 95% confidence level)

Impact of Weighting by SEC (Index :Original Sample = 100)

TABLE 7

Key TV related parameters:	Without SEC Weights	With SEC Weights
Own a TV set	107	103
Own a Colour TV set	124 *	107
Access to Cable & Satellite TV	138 *	122 *
TV viewing frequency:		
Everyday	102	99
4 to 6 days a week	116	87
1 to 3 days a week	106	109
less often	72 *	91

(* = differences between the original sample and the substitute sample are significant at 95% confidence level)

The above analyses do indeed show that by using SEC in the weighting procedure, the differences do narrow considerably, though in isolated cases there is no change.

A key dilemma however, is that in India, because of its huge sample size (1,10,000), rigorous sampling method, and meticulous fieldwork procedures, the NRS is the source for SEC distributions. Perhaps, a way out could be to have a survey which gives us a very `good' SEC distribution to use as a weighting procedure. This could either be an independent survey, or for those non responding households in the original NRS sample, we make additional calls to at least establish the Social Class. The questions to determine Social Class are only two:

- Occupation of the Chief Wage Earner
- Education of the Chief Wage Earner

Households unwilling to participate in the detailed interview may not be unwilling to provide this information.

Playing devil's advocate, if the survey results as shown in Table 2 and 3 for the original sample and the total sample are not too different, then why have substitutes at all ?

Let us, therefore, examine what we gain by having substitutes. What we essentially gain is a larger sample for smaller publications.

Unweighted Average Issue Readers Sample

TABLE 8

No. of publications with:	Original Sample	Total Sample (with substitutes)
Sample size 200 and more	9	12
150 or more	13	23
100 or more	25	29
50 or more	36	41

The above analyses clearly justifies and builds a case for substitutes; if 50 is a minimum threshold sample for some analyses, then five more publications get included. Also, as many as 10 more publications have a sample size of 150 or more, thereby enabling analyses with two or three variables at the target group definition level.

Conclusions

It is evident from the analysis presented in this paper that there is a marked difference in response rates by social class. 31% of the upper social class respondent were substitutes as compared to 18% for the lower social classes.

However, if care is taken to match the profile by sex and age of the substitute sample, with the profile of the originally selected individual, the differences in estimates of readership of publications is small.

Our analysis appears to indicate that were we to include social class as one of the parameters for weighting the substitute sample data, then the differences between the original and the substitute sample would narrow down even further.

Allowing substitution increases the overall sample size in a very cost-efficient manner, yielding larger sample sizes for smaller publications. This helps in providing more accurate estimates of readership even for finer definitions of target groups.

Handled with care, substitution can provide good data in an economical way.

References :

Hans Vorster, Totum Audits and Survey Inc and John Smart, Applied Marketing Statistics. Readership and Survey Non Response. Paper presented at Readership Research Symposium, San Fransisco 1993.

Praveen Tripathi, MediaSearch. Are Hard To Get Respondents Any Different In Their Reading Behaviour From Others? Paper presented at Readership Research Symposium, San Fransisco 1993.

Stephen A Douglas, Dr Julian Baim. Response Rates : A Multinational Perspective. Paper presented at Readership Research Symposium, Hong Kong 1991.

Paul Donato. Mechanism For Improving Response Rates In Complex Media Survey. Paper presented at Readership Research Symposium, Hong Kong 1991.

Erhard Meier. Response Rate Trends in Britain. Paper presented at Readership Research Symposium, Hong Kong 1991.

Hans Vorster, Lester Frankel. Are All Those Call Backs Really Necessary. Paper presented at Readership Research Symposium, Salzburg 1985.