

# Telephone Directories as Sampling Frames

*Don Esslemont, Susan Petersen and K S Selvakumar*

Of 1000 respondents in a face-to-face survey, 22% did not live in households listed in telephone directories, and therefore could not have been included in a sample using telephone directories as a sampling frame. Those not listed were significantly younger, more likely to be never married, separated, or divorced, and were more likely to live in the northern part of New Zealand.

Keywords: telephone survey, sampling frame, non response bias

## Introduction

There are three main methods of collecting survey data: face-to-face interviews, mail surveys and telephone surveys. Traditionally, telephone surveys were considered to be biased because of the low incidence of telephone ownership. But, due to the high level of telephone ownership in New Zealand today (95% of households), advances in telecommunications technology, and the potential for substantial cost savings, telephone surveys have become very attractive to buyers of research.

In telephone surveys, three main sampling frames are commonly used: the telephone directory, random digit dialling, and specifically compiled lists. In New Zealand, the most commonly used sampling frame for telephone surveys is the telephone directory. To date, however, no studies of the accuracy or reliability of the telephone directories as sampling frames have been conducted here.

## The Telephone Directory

The use of telephone directory as a sampling frame results in three major sources of bias: exclusion of non-telephone households, exclusion of telephone numbers that are issued after the publication of the telephone directory, and exclusion of unlisted numbers.

Earlier research, mainly in the United States, into bias resulting from using telephone directories as sampling frames, focused on the characteristics of telephone owners and non-telephone owners. Leuthold and Scheele (1971) found that people living in rural areas and those on low incomes were less likely to own telephones than those living in urban areas with higher incomes. Tyebjee (1979) found this difference in telephone ownership often resulted in over-representation of households with certain geographic and demographic characteristics when telephone directories were used as sampling frames. Hyett and Allan (1976) also noted that the telephone owners were more affluent than the general population.

Some of this bias is attributable to out-of-date listings. Cooper (1964) estimated that the proportion of subscribers not listed in the telephone directory at the end of the year amounted to 18% in United States. Blankenship (1977a) estimated that if 20% of families move annually, then the average listings of non-current names will be around 10%.

A major source of bias in recent times is the increasing use of unlisted telephone numbers. Leuthold and Scheele (1971) found that 9% of their respondents had unlisted telephone

numbers. Brunner and Brunner (1971) and Frankel and Frankel (1977) questioned whether the characteristics of the unlisted telephone owners are different from listed telephone owners.

Cooper (1964) stated that persons whose numbers are unlisted may represent, at least in part, higher income levels and greater educational accomplishment, such as lawyers, physicians, entertainers and other public personalities. In contrast, Leuthold and Scheele (1971) found that unlisted numbers average income and education levels were similar for listed and unlisted populations.

Moberg (1982) stated that if unlisted numbers are systematically excluded, the resulting sample would consist of people somewhat older, more rural, more white, more educated, more retired and more white collar than the households with telephone service.

The potential for bias has prompted several commentators to suggest that, in the United States at least, telephone directories are not adequate sampling frames (Cooper, 1964; Brunner & Brunner, 1971; Roslow & Roslow, 1972; Glasser & Metzger, 1972, 1975; Blankenship, 1977b; Tyebjee, 1979; Moberg, 1982). However, in spite of their deficiencies, the convenience of telephone directories in both hard copy and computer tape versions points towards continued reliance upon them. Furthermore, not all of the evidence weighs against their use.

For example, Leuthold and Scheele (1971) suggest that, for some specialised surveys, the telephone directory may be relatively unbiased, while Rich (1977) concludes that, for some studies, samples which include only listed subscribers may be adequate and are not different from those of the total subscriber population. Rich suggests that the choice of sampling frames should be made according to the objective of the survey.

Furthermore, Sudman (1973) concludes that directories are not only higher in quality, but less expensive than methods that use random dialling or door to door listing exclusively. He thinks that in certain areas, when the fraction of unlisted telephone is less than 5%, directories are perfectly adequate frames. In other areas, a combination of random digit dialling and the use of telephone directory is preferable.

Clearly; the possibility of bias is an important consideration when selecting a survey sampling frame, and the reliability of New Zealand telephone directories as sampling frames deserves investigation. The purpose of this study was twofold: (i) to measure the accuracy of telephone directories in obtaining a representative sample of the general population and (ii), to determine the proportion of households who had no chance of inclusion in a sample drawn from telephone directory.

## **Method**

A market research company in New Zealand supplied names, addresses and telephone numbers from an omnibus survey. The sample consisted of 1000 respondents selected nationally. Although there may have been differences between the achieved sample and the New Zealand population, the company concerned is known to sample carefully and to insist on up to three call-backs. The achieved sample was therefore taken as representative of the New Zealand population over the age of 15 years.

It was then determined if each respondent could have been contacted had the telephone directory been used as a sampling frame. Each name was checked against the telephone directory current at the time of survey. If that surname was listed, with the same telephone number, that respondent was immediately coded as being contactable using this frame. To ensure that incorrect spellings of names were not the reason why the respondents appeared not to be listed, both the Wise's Directory and the electoral rolls were referred to. If a similar name at the same address was found the telephone directory was referred to again.

If the respondent was still not verified as listed, the Universal Business Directory, which lists telephone numbers in numerical order, was used, to ensure that respondents with telephone numbers listed under other household members' names could also be included in the group that could be contacted, if their number was actually listed. This was done since the telephone directory is only used as a household sampling frame. Hence, if that household was contacted, respondent selection within household would enable the initial respondent to be included in the sample.

The Universal Business Directory enables the first four letters of the surname of the telephone subscriber to be determined by looking up the telephone number, as provided by the respondent. The telephone directory was then used to determine if that person lived at the same address and had the same telephone number as the respondent. Those that were, were coded as "listed". They represent that part of the population that could be reached by sampling from telephone directories. All remaining respondents were coded as being "unlisted", and represent those who could not have been contacted through the use of the telephone directory as a sampling frame.

Differences between listed and unlisted respondents were tested using the Kolmogorov-Smirnov Two Sample Test (Siegel, 1956, pp. 127-136) for ordinal data, or the Chi-Square Test for nominal data.

## Results

A total of 221 respondents out of a sample of 1000 respondents could not have been contacted if the telephone directory had been used as the sampling frame.

To gain some insight into the bias this may introduce into the sample, a comparison was made of the demographic characteristics of the "listed" and "unlisted" sample groups. The findings, summarised in Table 1, indicate that considerable bias is likely to result if telephone directories are used as sampling frames. In particular, those in unlisted households are more likely to be:

- i. aged between 15 and 39 years
- ii. never married, or separated or divorced
- iii. young and single, or young and married (with or without children)
- iv. from northern North Island, Auckland, Hamilton or Christchurch.

In contrast, those in unlisted households are less likely to be:

- i. aged 55 years or older
- ii. from the south of the North Island, or the South Island
- iii. widowed, or married
- iv. middle aged (whether with or without children)
- v. older (whether married or single)
- vi. from Wellington, northern North Island towns, or South Island towns.

No significant differences between the listed and unlisted groups were found for gender, size of town or city of residence, socio-economic level, size of household, educational level, employment status, respondent or household income, or respondent occupation.

## **Discussion**

These results indicate that there can be quite substantial differences between individuals in listed households and others for some demographic characteristics. Since only 80% of the population are in listed households, and bias does exist, a survey using telephone directories as a sampling frame will result in sampling distribution with mean estimates, for items affected by these characteristics, that do not correspond with the true mean of the population. Thus the onus is on the researcher to qualify the extent of bias that would be introduced by using a telephone directory sample.

Although the present research has identified differences in some demographic variables between 'listed' and 'unlisted' individuals, further studies need to be done to validate these findings, particularly since there is a limitation in the present study. This study only compared the demographic characteristics of the listed and unlisted respondents. However, in some cases, a respondent's demographic characteristics may have little bearing on the topic under investigation. For this reason, future studies should examine the bias associated with other variables, such as product usage and purchase probabilities, or other information that has a direct bearing on the decisions that will be made on the basis of the survey findings. Such information would aid market researchers in making marketing decisions.

**Table 1. Characteristics of listed and unlisted samples**

<b>Characteristic</b>	<b>Listed<sup>1</sup> Sample</b>	<b>Unlisted<sup>2</sup> Sample</b>	<b>Difference</b>	<b>D<sup>3</sup></b>	<b>p</b>
<u>Gender</u>	%	%	%		
Male	47.3	49.7	-2.4	0.347*	n.s.
Female	52.7	50.3	2.4		
<u>Age</u>	%	%	%		
15 - 24 years	15.2	31.2	-16.0	0.271	<.001
25 - 39 years	31.6	42.7	-11.1		
40 - 54 years	21.1	16.7	4.4		
55 years and above	32.1	10.0	22.1		
<u>Location</u>	%	%	%		
Northern North Island	45.1	54.2	-9.1	0.123	<.025
Southern North Island	29.0	22.2	6.8		
South Island	25.9	23.6	2.3		
<u>Population level</u>	%	%	%		
Metropolitan	28.4	34.3	-5.9	0.059	n.s.
Large cities	24.7	19.5	5.2		
Small cities	30.5	32.6	-2.1		
Towns	16.4	13.6	2.8		
<u>Marital Status</u>	%	%	%		
Never married	23.2	34.3	-11.1	29.68*	<.001
Widowed	9.2	3.2	6.0		
Separated/Divorced	6.6	13.1	-6.5		
Married/Living together	61.0	49.4	11.6		

<b>Characteristic</b>	<b>Listed<sup>1</sup> Sample</b>	<b>Unlisted<sup>2</sup> Sample</b>	<b>Difference</b>	<b>D<sup>3</sup></b>	<b>p</b>
<u>Respondent Life Cycle</u>	%	%	%		
Young single	19.6	30.7	-11.1	0.293	<.001
Young married no children	4.6	8.6	-4.0		
Young with children	11.2	25.4	-14.2		
Middle aged no children	20.4	14.5	5.9		
Middle aged with children	17.1	14.0	3.1		
Older married	16.7	3.6	13.1		
Older single	10.4	3.2	7.2		
<u>Socio-economic level</u>	%	%	%		
High I	10.1	10.4	-0.3	0.096	n.s.
II	18.2	13.5	-4.7		
III	26.7	22.7	4.0		
IV	23.8	22.6	1.2		
V	14.7	18.5	-3.8		
Low VI	6.5	12.3	-5.8		
<u>Size of Household</u>	%	%	%		
1	14.6	12.6	2.0	0.094	n.s.
2	30.9	26.7	4.2		
3	21.3	18.1	3.2		
4	20.2	20.4	-0.2		
5	8.8	12.6	-3.8		
6+	4.2	9.6	-5.4		

<b>Characteristic</b>	<b>Listed<sup>1</sup> Sample</b>	<b>Unlisted<sup>2</sup> Sample</b>	<b>Difference</b>	<b>D<sup>3</sup></b>	<b>p</b>
<u>Education Level</u>					
No Secondary School	4.8	4.5	-0.3	0.043	n.s.
School Exams	22.1	25.3	-3.2		
School Certificate	19.8	19.5	0.3		
6 <sup>th</sup> Form Certificate or UE	16.9	15.8	1.1		
Tech or Trade Cert	17.1	11.3	5.8		
Teaching Cert/Dip/Degree	18.1	22.6	-4.5		
<u>Employment Status</u>					
20 hours or more	51.4	53.8	-2.4	0.062	n.s.
Less than 20 hours	9.3	8.1	1.2		
Temporarily unemployed	1.8	6.8	-5.0		
Not income earner	37.5	31.3	6.2		
<u>Respondent Income</u>					
Up to \$5000	14.8	23.0	-8.2	0.082	n.s.
\$5001 to \$10000	20.6	10.9	9.7		
\$10001 to \$15000	10.5	11.3	-0.8		
\$15001 to \$25000	18.7	20.8	-2.1		
\$25001 to \$30000	8.9	11.3	-2.4		
\$30001 to \$35000	8.6	7.3	1.3		
\$35001 and over	14.5	12.2	2.3		
<u>Household Income</u>					
Up to \$10000	16.9	19.9	-3.0	0.071	n.s.
\$10001 to \$15000	9.8	10.4	-0.6		
\$15001 to \$25000	16.8	20.3	-3.5		
\$25001 to \$30000	10.0	8.2	1.8		
\$30001 to \$35000	8.6	7.2	1.4		
\$35001 to \$40000	10.6	15.9	-0.7		
\$50000+	17.6	13.5	4.1		

Characteristic	Listed <sup>1</sup> Sample	Unlisted <sup>2</sup> Sample	Difference	D <sup>3</sup>	p		
<u>Respondent Occupation</u>	%	%	%				
Prof/Managerial	14.3	17.6	-3.3	0.089	n.s.		
Clerical/Sales/Service	17.7	14.5	3.2				
Production/Operators	18.9	21.2	-2.3				
Other labour force	0.5	0.5	0				
Pensioner/retired	16.8	6.8	10.0				
Homemaker	21.2	19.9	1.3				
Student	8.1	11.3	-3.2				
Other non-labour fce	2.5	8.2	-5.7				
<u>Area of Residence</u>	%	%	%			0.017	<.05
Auckland	28.6	33.9	-5.3				
Hamilton	9.3	14.7	-5.4				
Northern HB towns	7.5	4.9	2.6				
Hawkes Bay	10.6	10.7	-0.1				
Wellington	14.0	6.7	7.3				
Southern Nth Is towns	4.5	4.5	0				
Christchurch	10.7	12.5	-1.8				
Dunedin	10.1	9.4	0.7				
South Is towns	4.8	2.7	2.1				

1. Listed sample N = 779

2. Unlisted sample N = 221

3. D = Kolmogorov-Smirnov statistic. \*For nominal (rather than ordinal) data, the value reported is the x2 statistic, not D.

## References

Blankenship AB (1977a). Listed versus unlisted numbers in telephone survey samples. *Journal of Advertising Research*, 17 (1), February, 39-42.

Blankenship AB (1977b). *Professional Telephone Surveys*. New York: McGraw-Hill.



- Brunner JA. & Brunner GA (1971). Are voluntarily unlisted telephone subscribers really different? *Journal of Marketing Research*, VIII (February), 121 -124.
- Cooper L (1964). Random sampling by telephone - an improved method. *Journal of Marketing Research*, November, 45-48.
- Frankel MR & Frankel R (1977). Some recent developments in sample survey design. *Journal of Marketing Research*, XIV (August), 280-293.
- Glasser GJ & Metzger GD (1972). Random-digit dialling as a method of telephone sampling. *Journal of Marketing Research*, IX (February), 59-64.
- Glasser GJ & Metzger GD (1975). National estimates of nonlisted telephone households and their characteristics. *Journal of Marketing Research*, XII (August), 359-361.
- Hyett GP & Allan GM (1976). Collection of data by telephone (CDT) - its viability in consumer research. *Market Research Society Nineteenth Annual Conference*, Metropole Hotel, Brighton, March, 29-39.
- Leuthold DA & Scheele R (1971). Patterns of bias in samples based on telephone directories. *Public Opinion Quarterly*, 35, Summer, 249-257.
- Moberg PE (1982). Biases in unlisted phone numbers. *Journal of Advertising Research*, 22 (4), August/September, 51-55.
- Rich CL (1977). Is random digit dialling really necessary? *Journal of Marketing Research*, X/V (August) 300-305.
- Roslow S & Roslow L (1972). Unlisted phone subscribers are different. *Journal of Advertising Research*, 12 (August), 35-38.
- Siegel S (1957). *Nonparametric Statistics for the Behavioural Sciences*. London: McGraw-Hill.
- Sudman S (1973). The uses of telephone directories for survey sampling. *Journal of Marketing Research*, X (May), 204-207.
- Tyebjee TT (1979). Telephone survey methods: The state of the art. *Journal of Marketing*, Summer, 68-78.
- Don Esslemont is a Senior Lecturer, and Susan Petersen and K S Selvakumar were graduate students, in the Department of Marketing, Massey University.**