THE SOCIAL
RESEARCH CENTRE

Title: A comparative analysis of EWP and RDD sample frames
Stream: Survey Research
Pennay, D and Challice, G.
Organisation: The Social Research Centre Pty Ltd
Contact Details: Email darren@srcentre.com.au Tel: 03 9382 0689

Overview

The context in which this paper is written is one whereby one of the most widely used sampling frames for telephone surveys (the Desktop Marketing Systems [DtMS] July 2004 release of the white pages on disc[1]) is rapidly ageing. As such the representativeness of samples obtained from this source is becoming an increasingly important consideration in telephone survey research.

The main focus of this paper is a comparative analysis of respondent characteristics obtained from surveys using listed telephone numbers as a sampling frame (i.e. those found in the July 2004 Desktop Marketing Systems product) compared with those using unlisted numbers (i.e. those not found in the DtMS product).

The observations in this paper are based on our experiences in providing survey research services to a range of government departments, government agencies and academic researchers.

The surveys used in this analysis include:

- The 2005 Victorian Population Health Surveys (VPHS) – Victorian Department of Department of Human Services. A high quality annual survey of 7,500 Victorians aged 18 years and over. The purpose of the survey is to inform and support planning, implementation and evaluation of adult health services and programs throughout Victoria.

- The 2006 Community Attitudes to Violence Against Women Survey (VAWS) – VicHealth. Community attitudes research involving scoping, design, development and reporting of multi-faceted survey relating to violence against women, including full qualitative development phase, 2,000 general community surveys and 800 surveys with persons of diverse linguistic and cultural background, and

- The 2004 International Crime Victimisation Survey (ICVS) – Australian Institute of Criminology United Nations Organisation. 6,000 general community surveys on crime victimization issues. Australian component of a large-scale international survey.

All were conducted to a high specification with strict controls on survey procedures.

---

[1] Which contained listing already between 6 and 18 months old at the time of release

Comparative profile of respondents with listed and unlisted numbers

The data in Table 1 shows the age profile of respondents with listed and unlisted telephone numbers for the Victorian Population Health Survey (VPHS) and the Community Attitudes to Violence Against Women Survey (VAWS)

As can be seen, the proportion of 18 to 40 year old persons is much higher (and closer to that of the general community) amongst samples obtained from unlisted numbers.

Table 1: Age distribution of persons with DtMS listed and unlisted phone numbers.

Project	ABS[a]	VPHS 2005		VAWS 2006	
		Listed (n=5,449)	Unlisted (n=2,111)	Listed (n=1,325)	Unlisted (n=675)
18-40 years ·············	42.7	23.7	43.3[#]	33.0	47.4[#]
41 years plus ···········	57.3	76.3	56.7[#]	67.0	52.6[#]

a) Estimated Residential Population,30 June 2005.
#Denotes significant difference between estimates for listed and unlisted samples at the 95% confidence interval.

Given the differences in the age composition of samples obtained from listed and unlisted telephone numbers, it is reasonable to expect some variance across a range of other demographic variables strongly correlated with age. Some examples are presented in Table 2.

As can be seen, the improved representation of younger persons in unlisted samples impacts the proportion of achieved interviews with employed persons (by reducing the skew towards retirees that is a consistent feature of listed samples), family households, divorcees and more "transient" household types, that is, those in rented accommodation or short term residents of an area.

Whilst these statistics are taken from the VPHS, the pattern is highly consistent across a range of similar surveys undertaken by the Social Research Centre. All of the results are statistically significant at the 95% confidence interval.

Table 2: Selected demographic characteristics of persons with DtMS listed and unlisted phone numbers.

Characteristic	Listed	Unlisted
VPHS 2005 (n=7,560)	*5,449*	*2,111*
Employed ·································	50.7	60.3
Retired ····························	33.2	15.4
University educated ··························	22.9	28.3
Household income <$40k pa ······················	49.7	46.6
Language spoken at home - English only ·············	89.8	84.5
Dependent children in household ························	32.0	42.3
Divorced / separated ····························	11.0	16.4
Resident in area less than one year ····················	1.7	18.9
Renting ·································	10.6	30.9

Whilst the demographic profiles of samples obtained from persons with listed and unlisted telephone differ quite significantly, it is important to understand whether these differences impact on key attitudinal or behavioural measures. This is briefly examined at Table 3.

As can be seen, key measures such as smoking prevalence and crime victimization rates differ by sample type. Whilst these measures may, at least in part, be related to factors like age (which is commonly corrected for by the application of weights), they may also be influenced by other factors such as the socio-economic status of respondents (an issue picked up on in our second paper).

Table 3: Selected survey measures of persons with DtMS listed and unlisted phone numbers.

Characteristic	Listed	Unlisted
VPHS 2005 (n=7,560)	*5,449*	*2,111*
Holds private health insurance	52.5	43.6
Daily smoker	14.6	22.0
Has alcoholic drink every day	18.9	13.6
Sought professional help for mental health related illness	8.6	12.2
Could raise $2,000 within 2 days in an emergency	85.0	79.4
ICVS 2004 (n=6,000)	*4,730*	*1,270*
Victim of violent crime	51.3	58.0
Victim of at least one crime	17.5	24.4
Feels safe waiting for public transport after dark	60.7	55.9
VAWS 2006 (n=2,000)	*1,325*	*675*
High gender equality rating	34.2	43.9
Interviewed late in call cycle (7 or more attempts)	23.5	31.1

All the results are statistically significant at the 95% confidence interval.

While the data presented to date shows broad attitudinal, behavioural and demographic differences between samples obtained from listed and unlisted telephone numbers, it is of interest to establish whether or not there are differences within selected population subgroups.

An example of the sorts of differences that can be found is provided in Table 4. This shows a comparative profile of persons aged 18 to 40 years reached via listed and unlisted telephone numbers.

It is evident from this data that the profile of respondents from listed numbers is significantly different from that of those with unlisted numbers. This has important consequences for the overall representativeness of samples based solely on EWP listed numbers.

Table 4: Profile of persons aged 18 to 40 years with DtMS listed and unlisted phone numbers.

18-40 year olds	Listed	Unlisted
VPHS 2005, 2205 18-40 year olds	*1,292*	*913*
Renting	17.6	45.7
Group household	6.7	10.5
Couple only	10.7	16.2
One parent family	8.6	15.6
One person household	5.1	8.2
Household income less than $40k per annum	23.9	35.6
General health rating of excellent or very good	48.6	42.7
Daily smokers	20.7	25.6

All the results are statistically significant at the 95% confidence interval.

In light of the above, survey researchers are increasingly turning to random digit dialing (RDD) sampling techniques in an attempt to include persons with unlisted numbers in their sample universe and thereby achieve a more representative sample of the general population.

RDD sample generation

There are a number of methods commonly used to generate RDD sample. The Social Research Centre typically uses the "known blocks" method, since it offers the best combination of efficiency (relative to "true" RDD) and rigour (relative to the "EWP plus one" method).

The procedure typically adopted is to:

- Undertake a random selection of records from the DtMS product within the agreed geographic strata, to be used as "seed" numbers for random number generation

- Retain the eight digit exchange prefix of the listed number (for example 02628946) and randomly generate the last two digits, to create a new randomly generated 10 digit telephone number, and

- "Wash" the resultant numbers against the latest electronic business listings to remove known business numbers.

Where the survey methodology involves sending approach letters to sample members to encourage response, an additional stage is to wash RDD selections against the DtMS product to identify which randomly generated telephone numbers can be matched to a listing in DtMS with a surname and an address (the "matched" sample) and which randomly generated telephone numbers cannot be matched to a White Pages listing (the "unmatched" sample).

This process typically yields around 30% to 40% matched numbers, depending on the location, with the matched numbers typically accounting for about 70% of the interviews achieved.

A method commonly used by researchers to improve the efficiency of approach letter mailing is to wash "matched" sample against current listings using Sensis' MacroMatch service.

This service confirms which DtMS selections remain current (with reference to the on-line version of the Electronic White Pages, which is updated daily). It also provides the new phone number for those DtMS selections where the surname and address remain the same, but the *phone number* has changed, or the new address where the surname and the phone number remain the same, but the *address* has changed.

With the ageing of the DtMS listings, the rate of successfully MacroMatched selections is gradually declining, and currently averages around 55% to 65% of DtMS selections, depending on the location.

Current limitations of RDD sample frames

Whilst it is generally accepted that this method of RDD offers improved coverage over electronic white pages sample frames, it remains an imperfect frame in that there is, for example:

- Potential under-coverage of areas that are under-represented in the source white pages listings, such as growth areas

- Lack of access to the full listing of eligible prefixes for "seed" number generation (the Integrated Public Network Database), and

- Non-coverage of households with no landline.

Historically, non-coverage has been higher for rented houses, households where the head is unemployed, young or with low income, and for single person households[2], however, it could reasonably be expected that there is also increasing under-coverage of the "3G" generation, which may represent a different demographic in terms of employment status or disposable income, for example.

When it might be appropriate to continue to use EWP sample frames

Depending on the accuracy of the estimates required, there may still be occasions when EWP sample frames are viable. These may include:

- When attempting to target specific groups, such as retirees, or older persons generally (the Social Research Centre and the Australian Institute of Health and Welfare used a sample frame that over represented households with listed numbers when undertaking a recent survey on vaccination rates amongst persons aged 65 years and over)

- When targeting long term / home owner residents

- When it is important to ensure approach letters are sent to all households in advance of interviewing

- Where there is a long, established time series, or

- When budget is limited – as using an RDD methodology adds approximately 5-10% to the costs of mounting a survey[3].

On balance, the findings in this paper suggest that telephone surveys using a sample frame that contains unlisted numbers will produce superior population estimates.

---

[2] Groves et al, Telephone Survey Methodology
[3] RDD samples are generally associated with higher refusal rates than EWP samples, due to factors such as the lack of a mechanism for matching unlisted numbers to an address (so that advanced letters can be sent to sample members), and lower overall sample yields, given that unlisted numbers feature a higher proportion of unresolved call outcomes (where no contact is established for a higher proportion of connected numbers)