

WHAT A DIFFERENCE A SAMPLE MAKES!

*Art Thomas, Eastern Institute of Technology, New Zealand
Gary Pickering, Brock University, Canada
athomas@eit.ac.nz*

Abstract

The importance of sampling to the overall outcomes of research and generalisations that may stem from any research can not be over emphasised. In particular, sampling frames would appear to be at the heart of the problem in that they project some array of population elements which may or may not be representative of some population at large. A possible end result is incorrect interpretation of actual results due to bias and the involvement and varied orientation of persons incorporated within a sample frame. This paper presents an after-event comparison of results between two sample frames used in a wine label study. The results portrayed by this comparison highlight significant and divergent outcomes that would lead to misinterpretations if only one sample frame had been used. It is argued that the use of two sampling frames acts as a levelling agent on possible biased results and that all researchers need to be constantly vigilant about sampling frames and their impact on results and interpretations.

Introduction

For those involved in research, there is countless reference in extant literature about the importance of sampling to the overall outcome of research and generalisations that stem or flow from findings and discussions. Timing, costs, objectives and quick access to information and population elements underlie some of the reasons why samples and not censuses are taken (Cooper and Schindler, 2003; Zikmund, 2003). However, considerations in these areas are weighed against the intended accuracy and reliability of possible results, how the information will be used and the very nature of the research itself; e.g., exploratory versus descriptive.

Notwithstanding these considerations, additional vexing problems are presented when attention is directed to sampling frames; that is, the register of population elements from which samples are drawn. Sampling frames can take many forms such as electoral rolls, telephone books, patient lists, magazine subscribers or a company's customer or

client list. Such listings are often thought of as convenient aggregations of potential respondents, hence we are warned about the inclusion or exclusion of certain members of the greater population; many people do not subscribe to particular magazines or deal with particular businesses. Thus, the representativeness of some sampling frames is questioned and this raises the spectre of bias. As pointed out by Aaker, Kumar and Day (1998, p.376), readers of a special interest magazine on tennis are likely to be more involved and knowledgeable about the subject than some average player.

In addition to problems presented by bias or involvement, the currency and accuracy of any list is important. Saunders, Lewis and Thornhill (2000) highlight that telephone books are printed annually, that not every one has a telephone and that some people have unlisted numbers. To combat these and other related problems, multiple list usage by researchers has been suggested (Cowan, 1991; Bolton, 1994; Cooper and Schindler, 2003).

This paper reports a comparison on a New Zealand survey that utilised two sampling frames in a wine label study. One sampling frame was a wine retailer's, national mailing list, whilst the other was drawn from a regionally-based tertiary institute. In addition to assessing information factors of wine labels, the study also examined wine purchase behaviours and information element importance.

SAMPLING FRAMES, MAILING LISTS AND RELATED MATTER

In the execution of research, sample design and sample frames are seen as important contributors to the overall quality and validity of surveys (Stephen and Soldo, 1990; Lin and Jones, 1997). The basis for this concern rests with the representativeness of a sample frame and the consequent bias that may emerge if sample frames or lists in general are not representative of some population (Bolton, 1994). On the issue of representativeness, Kaldenberg and Becker (1990) highlight the fact that inappropriate sampling frames lead to mistakes and problems. In their case, differences were found between practicing professionals who were members or non-members of a professional association. In short, practice was not dependent on being an association member.

To combat potential problems posed by some sample frames and lists, researchers have reverted to the use of dual, and even, multiple sample frames (Bolton, 1994; Cowan, 1991, Murphy, 2002). Whilst questions of representativeness is an underlying reason for these decisions, cost savings, accuracy, coverage and efficiency are further reasons for multiple frame usage. Whilst Bolton (1994) notes that two lists guards against the variability of either group's characteristics, Murphy (2002) still found bias to exist in a study involving three sample frames. This finding may be a result of the completeness of any list, a factor questioned by Iannacchione, Stabb and Redden (2003), or comprehensiveness as identified earlier by Lin and Jones (1997).

Despite these potential problems, a variety of sampling frames and mailing lists continue to be used possibly because they act to balance any potential bias inherent in the use of a single list or source. However, as Krysan, et al. (1994, p.382) state, "membership lists may have the advantage of drawing on the respondent's connection to the sponsoring organisation to motivate participation". This connection may suggest a form of involvement, but of itself would appear to contain some potential bias.

Notwithstanding a variety of reasons for which lists might be used in research, mailing lists can be viewed from at least two perspectives; the 'owner' or business and the individual. From the business side, the mailing list represents individuals who can be classed as prospects, customers and non-buyers (Sterk, 2003). For whatever reason (interest or relevance), people have volunteered personal contact details and thus have given a business 'permission' to send various amounts of information. Phelps, Nowak and Ferrell (2000) intimate that the release of this type of personal information is a trade off against perceived benefits to the individual; the individual benefits by gaining convenience, discounts, information and announcements.

Apart from the sales or direct marketing aspects of mailing lists and the potential for sharing some form of common interest with some product, Stillwell (1999) has indicated that mailing lists allow people to be kept informed (benefit) without being swamped by information. In this context, there is an element of control vested in the recipient of any information in that they decide whether or not to engage with any of the information provided. Some years earlier, Block, Sherrel and Ridgway (1986) established that people on mailing lists engage in on going search, itself connected to product

involvement. Amongst other things, product involvement is about interest and relevance to individuals. More recently, Dholakia (2001) linked enduring involvement with an individual's self-concept, values and ego.

In research, the use of mailing lists must be recognised as enacting a complex interaction or relationship amongst interest, relevance, benefits and engagement. This interaction may not guarantee respondent participation as suggested by Krysen, et al. (1994), but seemingly it will guarantee divergent and questionable results.

METHODOLOGY

A questionnaire consisting of a mock wine label, ten questions and respondent background information was developed. The mock label used *clip art* visuals (glass of wine and grapes), and was accompanied by vintage year, the name of a fictitious winery and the type of wine. A variety of information such as the wine's attributes, how it could be used, how the wine was made, its attributes and that it had won an award was included. In all, seven information elements were used and these, derived from extant literature, were manipulated in a research design of eight versions of the final questionnaire; one version contained all seven information elements, whilst the remaining versions eliminated a single and different piece of information.

Based on the information provided, respondents were asked to indicate their likelihood of purchasing the wine described. An 11-point, labelled probability scale was used to measure purchase likelihood. The scale ranged from a high of '10' (Certain, Practically Certain) to a low of '0' (No Chance, Almost No Chance). Each point was accompanied by expressions of chance. For example, the uppermost scale point was represented by a *99 in 100 chance*, whilst the lowest scale point showed only a *1 in 100 chance*. In addition, respondents were also asked to provide a reason for their response likelihood and indicate what they would be prepared to pay for the wine described. A number of wine purchasing behaviour questions such as types of wine bought, place of purchase, average number of bottles purchased per month and usual price paid were also asked, together with other demographic information. No attempt was made to determine wine type preferences.

The questionnaire was distributed to a total of 1,144 potential respondents. This sample size was based largely on the treatment of information elements in the survey (eight versions) and two types of wines, and the possibility of gaining about 40 responses for each version. Some respondents were drawn from a wine retailer's nation-wide mailing list (n=640) and for these reply-paid mail was used for returning questionnaires. The remainder (n=504) were staff and students from a tertiary institution and these respondents were provided two drop-off zones on campus. No indication of wine consumption activity amongst this latter group was known. The total sample represents a dual sample frame and this was used in an attempt to combat potential bias of a mailing list only sample.

The 640 mailing list respondents were randomly selected from a base of 12,500. The list's owners undertake monthly promotions. Their mailings are organised by postal codes in 25 boxes, with each box containing 500 envelopes. A random number between 1 and 25 was selected as a start point, together with a random start point within the first box selected. From here, every 19th envelope ($12,500 \div 640$) was chosen and a questionnaire was inserted.

For the staff sample, questionnaires were distributed through an internal mailing system. Each staff mail slot received a questionnaire. Where mail slots were shared, a different questionnaire version was distributed. Questionnaires for students were distributed via lecture venues across all faculties. Regardless of the sample base, a rotation between issuing a Cabernet Sauvignon and Chardonnay version of the questionnaire was followed not only to address a possible 'wine type bias', but also to assess whether 'wine type' impacted on information assessment of the mock wine label (see Thomas and Pickering, 2003). Data was analysed using SPSS Version 10.0.

In total, 320 returns were received, representing a 28% response rate. A different response rate exists for each sample component; for mailing list respondents this was 23%, whilst that for staff and students was 38% and 28%, respectively. Given that no follow-up activities were undertaken, and that the research was exploratory, these response rates are reasonable and yield an overall $\pm 4.9\%$ error margin.

RESULTS AND DISCUSSION

Sample Characteristics

Of the 320 returns, 54% was contributed by tertiary institute respondents (n=173), whilst the remainder (n=147) came from the mailing list sample. Close examination of these respondents across demographic categories reveals significant differences. The mailing list sample is characterised as married males between 41 and 60 years of age who have an education level of diploma or less, live in higher income households where a majority would classify themselves as *Connoisseur* wine consumers. The tertiary sample is predominantly female, aged 31 to 50 years in households of diffuse income, and who are most likely to classify themselves as *Beverage Wine Drinkers*. Full details of the contrasts are shown in Table 1.

Table 1. Sample contrasts by demographics

Demographics	Tertiary		Mailing List		Chi Sq.	df	p=
	n	%	n	%			
<u>Gender:</u> Male	66	38	100	70	29.943	1	.000
Female	106	62	43	30			
<u>Marital Status:</u>							
Single	49	29	17	12	12.517	2	.002
Married	116	69	117	85			
Other	4	2	4	3			
<u>Age:*</u>							
30 or less	42	24	7	5	44.166	5	.000
31 – 50	96	55	79	54			
51 or more	34	20	60	41			
<u>Education: *</u>							
No formal	3	2	13	9	16.744	5	.005
Diploma or less	90	52	58	40			
Degree	52	30	36	25			
Post-graduate	28	16	37	26			
<u>Household Income:*</u>							
\$35,000 <	52	32	8	6	52.980	4	.000
\$35,001-\$45,000	17	10	15	11			
\$45,001-\$55,000	24	14	7	5			
\$55,001 >	72	48	113	79			
<u>Wine Consumer:</u>							
Connoisseur	36	21	86	59	57.507	4	.000
Aspirational	16	9	13	9			
Beverage Wine	84	49	30	20			
New Wine	12	7	-	-			
Combination	23	14	18	12			

* Number of items reduced to fit within Table

Whilst it was anticipated that some differences would appear between the two samples, it was not expected that these differences would be so acute. As a result, close attention was devoted to the main parameters of the study.

Purchase Likelihood

Respondents were asked to read the information provided by the mock label and indicate their likelihood of buying the wine described, provide an explanation and indicate how much they would be prepared to pay for the wine described. Despite the demographic differences highlighted, no significant difference in purchase likelihood is revealed. The tertiary sample showed a 54% likelihood of purchase compared to the mailing list of 52% ($t=.643$, $p=.521$).

In examining respondent reactions and explanations to the described wine, general categories of *positive*, *negative* and *neutral* were devised. No significant difference between the two sample components existed (Chi Sq. 1.810, $df=2$, $p=.405$). Similarly, no significant difference existed amongst respondents as to the price they would be prepared to pay for the wine described (Chi Sq. 4.084, $df=5$, $p=.537$). On average, the tertiary group were willing to pay NZ\$ 18.86, whilst the mailing list group were marginally higher at NZ\$ 19.71.

Whilst no significant difference between the two samples exists in terms of *prepared to pay* for an unknown wine, the same can not be said about the price *usually paid* for wine. Here, a significant difference exists; tertiary respondents on average usually pay NZ\$ 14.88, whilst the mailing list group usually pay NZ\$ 17.02 (Chi Sq. 22.018, $df=5$, $p=.001$). Table 2 highlights average price differences between the two groups for the two conditions.

Table 2. Prepared and usual price comparisons

Price	Tertiary	Mailing List	diff. NZ\$
Prepared to pay	18.86	19.71	+0.85
Usually pay	14.88	17.02	+2.14

diff. NZ\$	-3.98	-2.69	
------------	-------	-------	--

The information presented in Table 2 suggests a comparable scrutiny of information between the two groups when it comes to being prepared to pay for a wine (a guess really as no price information was given). However, the usual behaviours of the two groups are markedly different with mailing list respondents paying a great price, on average, than the tertiary group. If one assumes the mailing list respondents have a greater level of knowledge about wines than the tertiary group, then it may be possible for new wine releases to extract a higher premium from lower knowledge groups based on constant information to both groups.

Wine Purchase Behaviour

Apart from the price usually paid for wine, respondents were asked to indicate from where they bought their wine, from which outlet they bought mostly, bottles purchased per month and types of wine bought. In terms of the various outlets from which wines can be obtained (treated as a multiple response variable), respondents were free to mention any or all that applied. From the data (see Table 3), it would appear that the tertiary group concentrate their wine purchases through *supermarkets* (90%) followed by *bottle stores* (79%) and *wineries* (60%). On the other hand, mailing list respondents have a firmer leaning firstly to *bottle stores* (88%) followed by *supermarkets* (79%), *wineries* (73%) and *direct mail* (70%). The mentioning/not mentioning relationship for each outlet is significant between the two groups.

Table 3. Purchase outlet mentions

	Tertiary	Mailing List	
--	----------	--------------	--

Outlets	n	%	n	%	Chi Sq.	df	p=
Winery	104	60	107	73	5.561		
Bottle store	137	79	128	88	1	.018	
Supermarket	156	90	115	79	3.469		
Direct mail	27	16	102	70	1	.063	
Other	9	5	23	16	7.191		
					1	.007	
					94.531		
					1	.000	
					8.633		
					1	.003	

Following immediately on from determining the various outlets used for purchasing wine, respondents were asked to indicate the one outlet from which they bought most of their wine. As Table 4 points out, the tertiary group are most likely to concentrate most of their wine purchases in *supermarkets* (53%), whilst the highest concentration for mailing list subjects is *bottle stores* (32%). However, this group of respondents is by no means as uniform as their tertiary counterparts; it would appear they are more likely to spread their purchases around several outlets. These elements contribute to the significant relationship shown amongst the data.

Table 4. Main purchase outlets

Outlets	Tertiary		Mailing List		Chi Sq.	df	p=
	n	%	n	%			
Winery	17	10	23	16			
Bottle store	51	29	47	32			
Supermarket	92	53	36	24	38.633		
Direct mail	10	6	31	21	1	.003	
Other	3	2	10	7			

Bottled Wine Purchases

Respondents were asked to indicate the number of bottles of wine they purchase in an average month. The options available were 1-3, 4-7, 8-10, 11-20, 21-30 and 31 or more. Whilst it is possible for people to understate their actual level of purchase activity, it is noted that a significant difference between the two samples exists for the full data set (Chi Sq. 118.584, df=5, p=.000). Based on the options and data available, it is estimated that the tertiary and mailing list groups purchase six and 13 bottles of wine respectively per month.

The response option were recoded to reflect *light*, *medium* and *heavy* purchase activity, the divisions being 7 or less, 8 - 20 and 21 or more respectively. Based on this categorisation, the tertiary group are overwhelmingly *light* purchasers of wine, whilst the mailing list group are mainly *medium* in orientation. Details of this are shown in Table 5.

Table 5. Monthly wine bottle purchases

Bottle per month	Tertiary		Mailing List		Chi Sq.	df	p=
	n	%	n	%			
7 or less	139	81	40	27	91.647	2	.000
11 – 20	25	14	83	57			
21 or more	8	5	23	16			

Types of Wine Purchased

Respondents were free to indicate all or any types of wine purchased. The wine options available were still red, still white, rose, sparkling and fortified. Although both groups tended to mention still red and white wine purchases, the nomination levels for these were far greater amongst the mailing list sample. This nomination level generally was greater for the mailing list group across all wine categories except sparkling wine resulting in significant differences between the two groups. Details are shown in Table 6.

Table 6. Types of wine purchased

Types of Wine	Tertiary		Mailing List		Chi Sq.	df	p=
	n	%	n	%			
Still red	126	73	133	91	14.912	1	.000
Still white	133	77	136	93	13.363	1	.000
Rose	4	2	11	7	3.669	1	.055
Sparkling	58	33	44	30	-	-	-
Fortified	16	9	27	18	4.924	1	.026

Importance of Information Elements

Various amounts and types of information appear on front and back labels of wine. From a list of 14 items, respondents were asked to indicate how important each item was to them when deciding which wine to buy. Also, respondents were asked to indicate how important front and back labels were to them in their wine buying decisions. A seven-point scale was used for both areas where '1' indicated *very little importance* and '7' was *very great importance*.

In terms of front and back label importance for wine decisions, front labels were seen to be slightly more important for the tertiary sample (4.86) than they were for mailing list respondents (4.73) though this difference was not great enough to exhibit a significant difference ($t=.741$, $p=.459$). Whilst this was the case for the front label, the same can not be said for its counterpart; the back label was significantly more important to mailing list respondents (4.84) than the tertiary sample (4.40; $t=-2.425$, $p=.016$).

Of the 14 information elements examined, 11 were determined from extant literature and included attributes, manufacture, wine company, wine brand and how the wine should be used. Three additional elements normally associated with labels were added; these included alcohol level, image, logo or coat of arms and label colours. Of the 14 information elements, *t-tests* showed nine significant differences between the mean scores of the two sample groups. Four of these items were more important to the

tertiary group and included *type of situation, how the wine should be used, image/logo and colours used on labels*. By contrast, the mailing list group stressed greater importance on more core-related wine aspects inherent in *the wine region, the wine maker, the wine brand, wine company and how the wine was made*. Information items, mean values, *t*-scores and significance values are shown in detail in Table 7.

Table 7. Information item importance (Means)

Information Item	Tertiary	Mailing List	<i>t</i> -Test	p=
Wine company	4.85	5.45	-4.052	.000
Wine brand	4.77	5.16	-2.377	.018
Expert opinion	4.75	4.99	-1.541	.124
Attributes	4.78	4.52	1.436	.152
Wine maker	3.70	4.37	-3.877	.000
How wine made	3.82	4.18	-1.932	.054
Wine region	3.80	4.18	-2.044	.042
How wine used	3.98	3.33	3.412	.001
Image/Logo	3.89	3.14	4.122	.000
Label colours	3.85	3.03	4.271	.000
Alcohol level	3.27	3.02	1.244	.214
Unique/unrivalled	3.09	2.93	.892	.373
Type of situation	2.96	2.33	3.329	.001
Type of person	2.61	2.35	1.462	.145

CONCLUSIONS

The act of volunteering contact information for a retailer's mailing list for the purpose of receiving product information on a regular basis is strongly suggestive of both interest in and relevance of a product to an individual. Interest, relevance and increased levels of information or knowledge are central elements of involvement. It is postulated that this situation would likely skew research results. Evidence of this can be seen, for example, in a higher concentration of *Connoisseurs* in the mailing list group, and this group's emphasis on the importance of core-related wine information elements.

There is ample evidence in terms of sample characteristics and wine purchase behaviour throughout this paper that points to significant differences and skewed results between the two samples used. Had only one sample frame been used, then any number of mistakes or problems could have eventuated regardless of either sample frame. For example, Thomas and Pickering (2003) highlight a number of differences in relative importance of various label elements as a function of demographics and type of wine consumer. Hence, the decision to include what was thought to be a 'general public' component in the original research design and thus utilise a dual sample frame is believed to have been justified. In the case presented, the use of the two sample frames would appear to act as a balancing or levelling factor. It is interesting to note here that the study was not designed with a sampling frame comparison as an objective. The approach taken is one of a reanalysis of secondary data, and thus the experimental design may not be as robust as it would otherwise have been.

As a final comment and regardless of research experience, researchers need to be constantly vigilant about the impact that sample frames have on the results and interpretations of research undertaken. Without this vigilance, incorrect interpretations can easily follow.

REFERENCES

- Aaker, D.A., Kumar, V., & Day, G.S. (1998). *Marketing Research*. (6th ed.). New York: John Wiley & Sons, Inc.
- Anonymous. (n.d.). Make money with your mailing list. [http://www.hrdirect.net/free_report_copy\(1\).htm](http://www.hrdirect.net/free_report_copy(1).htm), retrieved from the Internet, August 27, 2004.
- Block, P.H., Sherrell, D.L., & Ridgway, N.M. (1986). Consumer search: An extended framework. *Journal of Consumer Research*, 13, 119-126.
- Bolton, R.N. (1994). Covering the market. *Marketing Research*, 6, pp.30-35.
- Clark, M. A., & Wood, R.C. (1999). Consumer loyalty in the restaurant industry: A preliminary exploration of the issues. *British Food Journal*, 101, p. 317+, retrieved through Proquest, August 17, 2004.
- Cooper, D.R., & Schindler, P.S. (2003). *Business Research Methods*. (8th ed.). Boston:

- McGraw-Hill Irwin.
- Cowan, C.D. (1991). Using multiple sample frames to improve survey coverage, quality, and costs. *Marketing Research*, 3, pp.66-69.
- Dholakia, U. M. (2001). A motivational process model of product involvement and consumer risk perception. *European Journal of Marketing*, 35, 1340-1361.
- Exter, T. (March, 1997). What's behind the numbers. *American Demographics*, 52-59.
- Innacchione, V.G., Staab, J.M., & Redden, D.T. (2003). Evaluating the use of residential mailing addresses in a metropolitan household survey. *Public Opinion Quarterly*, 67, p. 202+, retrieved through Proquest, October 8, 2004.
- Kaldenberg, D.O., & Becker, B.W. (1990). Research on attitudes of professionals towards advertising: A methodological caveat. *Journal of Advertising Research*, 30, pp. 17-23.
- Krysan, M., Schuman, H., Scott, L.J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly*, 58, pp.381-399.
- Lin, B., & Jones, C. (1997). Some issues in conducting satisfaction surveys. *Journal of Marketing Practice*, 3, p. 4+, retrieved through Proquest, August 17, 2004.
- Murphy, G.B. (2002). The effects of organisational sampling frame selection. *Journal of Business Venturing*, 17, p.237+, retrieved through Proquest, July 26, 2004.
- Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing*, 19, 27-41.
- Saunders, M., Lewis, P., & Thornhill, A. (2000). *Research Methods for Business Students*. (2nd ed.). London: Prentice Hall.
- Sterk, L. (2003). Editorial: A grand analogy. *Journal of Targeting, Measurement and Analysis*, 11, 200-202.
- Stillwell, D. (September 29, 1999). Mailing lists can deliver crucial data. *Journal of Commerce*, p. 13.
- Thomas, A., & Pickering, G. (2003). The importance of wine label information. *International Journal of Wine Marketing*, 15, 58-74.
- Zikmund, W.G. (2003). *Business Research Methods*. (7th ed.), Mason: South-Western.