Self Reporting Method in **Business Surveys**

■ *T. Srivenkataramana**

Abstract

This article reviews the role of self reporting (SR), by survey respondents in business research, given the now widely spread web and online facility with readily accessible panels. In this context, it outlines (a) the effects of response format (b) uses of SR information (c) dropout pattern on the web, together with affecting factors and (d) dual frame sampling for rare groups. Relevant empirical evidences, wherever necessary, are cited.

Key words and phrases
Dual frame sampling; self reporting; Telephone cluster sampling; web based design.

* *Former Chairman & Professor, Department of Statistics, Bangalore University, Bangalore.*

Introduction

With the facility for web and online surveys, use of self-reporting (SR) is becoming increasingly common as a mode of data collection. This may be used to determine whether a person has ever engaged in a specific behaviour, owned a particular product or experienced an event of interest. Examples are, respectively, use of a scheduled drug, owning of a digital camera and experiencing harassment by tax officials. The difficulties in obtaining accurate information in SR surveys has been pointed out (e.g.., Lee, Ju and Toh, 2000). It has also been indicated that, for behaviours with objective bench marks, SR information preserves both Order and Relative frequency with a fair degree of correspondence of reality. SR serves ,many functions; e.g.; as a screening tool to determine what questions to ask the respondents, subsequently. SR product ownership serves an important role, ranging from behavioral surveillance to market sizing studies. Thus many business surveys in market research, rely to some degree on SR of behaviors or events.

2. Common formats

Two formats are common in SR surveys.

A) Yes / No grid, where a respondent is presented with a series of items for which the answer is either a Yes or a No.

B) A multiple response list from which the respondent selects all that apply to him. This often includes a 'none of these' response too. Both formats are on vogue in visually presented (paper-pen / mail or web based) SR surveys.

3. Effects of format

The context of response formulation affects responses. In fact, both the range and availability of response options have been shown to affect response choices. With dichotomous response formats in web-based surveys, Thomas et.al. (2003) reported higher endorsement rates with yes / no grids as compared to multiple response lists. Thomas and Klein (2006) analysed five experiments conducted using the Harris Interactive panel (Harris Pool Online.com) which has over 7 million members. The respondents were all aged 18 years and more and the questions concerned purchase behaviour, consumption of product

The sample size was at least 900 each of the experiment and corporate prolist actions.

Ownership of different kitchen appliances the authors conclude mainly as follows.

1. The response format has significant effects on online survey behaviour across the various topics.

2. Yes/ No grid response format leads to higher levels of endorsement than multiple response lists.

3. The differences noted are robust across multiple time frames, different types of activities and information with a wide range of salience to individual respondents, and also across question formats requiring different recall strategies.

4. The same trend was found across

different countries and languages spoken.

5. However, the endorsement frequencies were ordered similarly between the two formats, with a Correlation of about 0.98 between proportions of endorsements under the two different formats.

6. The presence or absence of the 'None of these' category is unlikely to alter the above findings.

5. Comparison of formats:

A variety of factors can affect memory and recall of respondents, which include the time frame, the salience of the topic and event priming. It is also possible the yes / no grid and multiple response lists activate different recall strategies. Also the yes / no grid has a mandatory nature which forces the respondent to one of the two dimetrically opposite responses. In contrast, with a multiple list format a response for only a minimum of one element is needed to advance to the next question on the screen. Thus the respondents tend to glance at the elements rather than reading them carefully. This process of minimally reading a list and answering items that most stand out (known as satisfying), may result in choice of the first reasonable option, rather than the most appropriate option. In this sense response endorsement demands more effort than non-endorsement with a multiple list, while these efforts are equal with a yes / no set up.

6. Other uses of SR information

a) SR data have been used as a weighting function to adjust data to be more representative of a larger group. For example, ownership of products can be used to determine size of a social class.

b) SR data can help in deciding the size of the sample to be drawn in the main study. For instance, if the number of persons who engage in a behavior is lower than anticipated, more respondents will be needed to obtain the necessary base in the main study for valid inferences about the population. Contrary will be the case with a higher than anticipated incidence rate.

7. Dropouts on web:

Dropouts pose specific challenges to web survey research. Decision field theory, a dynamic theory of decision making developed by Busemeyer and Town send (1993), helps in understanding how changes in respondents' subjective experience during a survey affect their behaviour. A basic concept here is that of the inhibitory threshold, the point which determines when the difference in the preference for one or the other action is large enough to provoke behaviour. Two aspects of respondents' subjective experiences matter: interest in the questions and burden experienced while answering them.

8. Empirical Evidence

Galesic (2006) conducted a study that sheds some light on the dynamics of respondent behaviour in a SR survey, which was part of a larger, unrestricted, banner advertised online survey on unemployment in Croatia. A total of 2339 participants started the survey; 959 stated that they were unemployed and were included in the main study. This had about 180 questions divided into 20 blocks, each of which occupied one web page and contained questions on a related topic – everyday activities, financial situation, use of internet, health, social support etc. To control the effect of question position on tendency to drop out, the order of question blocks was randomized. At the end of each block, the respondents were asked how much interesting the block was and the burden experienced while answering it, both on a seven point scale.

The announced length of the survey and the motivation of the respondents were varied within a completely randomized experiment. The 3 announced lengths were 10,20 and 30 minutes. All three groups received the same questionnaire, but, for ethical reasons, the 10 minute group received – after the 5th block or approximately one third of the survey a message that the main part was now over but those who desired could continue. The same message was shown to 20 minute group after …12th block.

9. Factors affecting dropout

Cox proportional hazards regression model was used to test whether survey design, respondents' demographics and their experiences with the survey affect dropout behaviour, with the proportion of total number of blocks in each time group as the dependant variable. Hierarchical multi level linear

Modeling was used to examine correlates of changes in momentary interest and burden, both on a 7-point scale.

The following main points emerged:

a) Besides the announced length of the questionnaire, respondents' subjective experience during the survey significantly affected chances of dropout.

b) Respondents who dropped out after a certain block often reported significantly worse subjective experience with that block than the respondents who stayed.

c) It was possible to objectively observe respondents' tendency to dropout, even before they actually gave up, which is in line with decision field theory. Likewise respondent preference to

complete a block depended on their subjective experiences with the preceding block.

d) The overall dropout rate was about 42%. Also, these respondents had a somewhat lower overall response rate and they gave shorter answers to open questions.

e) Enjoyable items tended to improve the quality of answers to the questions that followed.

10. Dual frame sampling for rate groups

Blair and Blair (2006) have considered the conditions under which dual frame web-phone designs (Kalton, 2001) may be cost effective for sampling rare groups say no more than 20% within a general population. Here he screening costs dominate the total cost. The methods used for cost efficiency of reaching rare groups include cluster, disproportionate stratified, network, and dual frame sampling, particularly in telephone data collection, because of its combination of moderate cost, excellent population coverage and reasonable response rates.

The logic of web based data collection for rare groups is that the web has lower screening costs than other methods because no incremental labour or postage is needed to contact potential respondents. Blair and Blair (2006) have compared costs of telephone interviewing using cluster sampling with analogous costs of web based data collection with two scenarios: participation solicited from members of a) opt in lists and b) online panels.

11. Telephone cluster sampling (TCS)

The TCS procedure works as follows:

a. A random number ® is dialed within a bank of telephone numbers, via list assisted random digit dialing (RDD) or any other convenient procedure.

b. If r is not a member of the target group, then no further sampling is done within the bank. Otherwise, further sampling is done until a pre-specified number (no) of group members are identified.

This procedure has the effect of rapidly dropping telephone banks with no target group members and 'fishing where the fish are' for additional members of each cluster. The usefulness of TCS for sampling a rare group depends on the extent to which the group is geographically clustered. Otherwise, sampling of opt-in web lists is likely to be cheaper

Opt-in online panels consist of persons who have consented to participate, and typically, these contain at least 200,000 members. Empirical study by Blair and Blair (2000) has shown that substantial savings may result from using web panels to study rare groups if the target group is very rare and / or a large sample is desired. Dual frame web-phone designs will typically offer savings over telephone only designs if one has access to an online list in which target group's prevalence rate is at least as high as in the general population.

12. Biases in web-based sampling

One may mention three bias related issues in this context. These are, scope for

a) Coverage bias due to that several population units may not be online,

b) Volunteerism bias due to that available sampling frames cover only a fraction of online population and

c) Non-response bias because of possibly low response rates, at least for opt-in samples.

12. Conclusion

A notable feature of dual frame web-phone designs is that they allow a comparison of web and phone results. If serious bias is suspected, it might be managed through post-stratification, possibly based on information obtained in telephone sample. Thus, there is an attractive cost-saving role for the web in studies of rare groups, in conjunction with telephone or other methods. Also there exists a good platform for empirical research studies.

REFERENCES

1. Blair E. and Blari, J (2006), Dual Frame web-telephone sampling for rare groups. Journal of official statistics, 22(2), 211-220.

2. Buse Meyer J.R and Townsend, J.T (1993). A dynamic-cognitive approach to decision making in an uncertain environment. Psychological review, 100, 432-459.

3. Galesic M (2006). Dropouts on Web: Effects of interest and burden experienced during an online survey, Journal of official statistics 22(2), 313-328

4. Kalton G (2001). Practical methods for sampling rare and mobile populations. Proceedings of the American Statistical Association, section on Survey Research Methods, August.

5. Lee E., Hu, M.Y and Toh, R.S (2000). Are consumer survey results distorted? Systematic Impact of Behavioural Frequency and Duration on Survey Response Error, Journal of Marketing Research, 37,125-133

6. Thomas R.K, Behke,S., Smith R. and Lafond, R (2003). It's only incidental: Effects of Response Format in determining incidence. Paper presented at the 58th Annual Conference of the American Association for Public Opinion Research, Nashville, TN.

7. Thomas R.K. and Klein, J.D (2006). Journal of Official Statistics22(2), 221-244