



ISSN 1855-931X

IMPROVING REPRESENTATIVENESS IN ONLINE SURVEYS USING A COMBINED INTERNET/POSTAL APPROACH: EVIDENCE FROM A UK-WIDE SURVEY OF CONSUMERS

Peter Atorough*

Aberdeen Business School, Robert Gordon University, Aberdeen
United Kingdom
p.t.atorough@rgu.ac.uk

Bill Donaldson

Aberdeen Business School, Robert Gordon University, Aberdeen
United Kingdom
w.g.donaldson@rgu.ac.uk

Ainslie Harris

Aberdeen Business School, Robert Gordon University, Aberdeen
United Kingdom
a.j.harris@rgu.ac.uk

Abstract

The Internet is fast gaining popularity as a medium for undertaking various forms of research, particularly research of a quantitative and survey nature. While the benefits of undertaking such research are numerous and have been recognised in many empirical projects, it has become clear that a major drawback of Internet research is the inability of researchers to achieve representative samples drawn from well-informed scientific sampling techniques. As a result, a review of studies using Internet sample recruitment techniques reveals a predominant bias towards convenience samples, with the attendant weakening of conclusions therein.

In this paper, we discuss a survey sampling approach used by the researchers to recruit samples that improved representativeness and at the same time reduced systematic non-response bias. This paper will be of interest to those considering using Internet based surveys, especially if they are concerned about the generalisability of their results to include

wider and non-Internet-based populations. While the approach used herein is not entirely new to the research methods community, our findings provide further proof that a combined online and off-line method toward Internet sampling is more robust and useful for reaching strong, reliable conclusions.

Keywords: Internet, survey, representative samples, non-response bias.

Topic Group: Research Methods

INTRODUCTION

There is ample evidence that a growing number of researchers are using the Internet as a medium via which to reach research audiences of various objectifications. Its pervasive and ubiquitous reach makes it ideal for disseminating research findings and sharing subject knowledge to the researcher's audience, be they near or far dispersed. Equally, for the researcher or student seeking supporting material and information, the Internet has become an indispensable tool by which previous research, information and data can be mined and easily accessed (Benfield and Szlemko, 2006). Indeed in some cases the Internet has completely replaced paper-based published sources, for example some journals are now published entirely online. An even more pervasive use of the Internet for research is the gathering of primary data directly from respondents by way of surveys; some of the reasons identified for this are that the cost is lower, the turnaround time faster and the presentation more intuitive (Furrer and Sudharshan, in Evans and Mathur, 2005). However, while it is relatively easy to accumulate usage and similar statistics by tracking user activity, when it is required to engage the user consumer's participation directly through eliciting of responses, Internet researchers have fared poorly thus far in achieving representativeness and reducing non-response bias as well as attrition (Mathy et al., 2003; Fricker and Schonlau, 2002, in: Evans and Mathur, 2005). Indeed these problems are common in non-Internet surveys as well, however, unlike in traditional face to face, postal and telephone surveys, the novelty of the Internet medium means that researchers have as yet not fully developed techniques to reduce poor survey results arising from low response rates, attrition, and lack of generalizability; nor have they perfected the means to improve the quality of Internet surveys by following scientifically sound sampling methods. The immediate questions that come to mind are (i) how can Internet samples recruitment be made more representative and generalisable through the application of proven scientific techniques? And (ii) how can non-response bias arising from skewed convenience coverage of potential respondents be reduced in Internet based surveys? Zhang (in Benfield and Szlemko, 2006) states that response rates to Internet surveys can be dismal enough to make the time-honoured mail-in surveys more attractive. For example, typical response rates for online surveys still stand at about 6% to 7%, a relatively poor rate compared to rates that have been achieved in telephone and postal and person administered surveys (Evans and Mathur, 2005). Therefore a secondary question arising from the literature relates to how sampling can be approached so that response rates are improved across all groups of interest to the Internet survey researcher.

There are instances where purely Internet recruited samples are appropriate, or indeed the most justifiable; however, several instances of Internet recruitment methods are related to research that focuses on topics of wider applicability. For example, in the study of online shopping adoption and behaviour, the population of interest is arguably more general (for example, the national socio-economic population at large) than that of an Internet based panel or forum (for example, population of Internet users); it is this research on a wider

population, and for more general topics, that we focus on in the present paper, and this is why this paper is of great relevance to any researcher involved in, or contemplating, collecting research data by means of Internet administered surveys to large populations of the general public. In this paper, we share our experience of undertaking a UK-wide survey of consumers using an Internet questionnaire tool. By doing so, this paper contributes to the general discussion on improving Internet survey quality and specifically, shows how, by carefully adopting elements of traditional sampling designs, the quality of Internet surveys can be improved along the important dimensions of representativeness and response rate.

We first provide a brief review of the literature on Internet surveys in the next section. Thereafter, we describe the methodology used in this research, focusing on the sample plan and survey design and actualisation (Section 3). In Section 4 we describe the results obtained and discuss these results in relation to representativeness, response error, attrition and validity. We conclude in Section 5, drawing recommendations for future practice.

LITERATURE REVIEW

The development of Internet technology has greatly influenced the way research is conducted today (Evans and Mathur, 2005) and although there are several forms of research that can be conducted on the Internet, the most prevalent appears to be the survey research (McDonald and Adam, 2003; Evans and Mathur 2005). This is reflected in the increasing number of influential publications that use Internet survey methodologies or describe and define the use of these methods for data collection (for example Dolnicar et al., 2009).

Current Internet surveys are designed around the basis of accessible but narrow sample frames – the samples usually consist of specific interest panels or are non-probabilistically drawn from social networks and online recruitment by way of techniques such as self-selection invites, snowballing, saturation sampling and sifting (Zhao and Jin, 2006). While these approaches can potentially provide the required data in specific instances, in the context of a more representative national survey, the trouble with these approaches lies in the fact that oftentimes, the demographics of these sources are dissimilar to those of the wider population that uses the Internet more generally, for example as a means of conducting commercial and purchasing activities. Hence the analysis and conclusions drawn from the data obtained are often severely limited or weakened by this constraint (Evans and Mathur, 2005). Mathy et al. (2003) argue that to the extent that a sample population is not representative of a population, the generalisability of results and findings is limited and may also be further confounded by response patterns in an Internet-based protocol versus mailing or telephone. Yet the potential advantages of undertaking research surveys via the Internet are numerous (Weible and Wallace, 1998, In McDonald and Adam, 2003), warranting that researchers strive to identify ways of eliminating or at the least minimising compromised data while fully exploiting these advantages.

It is easy to see how the Internet has become a very attractive medium for undertaking survey research. According to Schleyer and Forrest (in Hung and Law, in press) the most commonly cited benefits are low costs and quick turnaround times. For instance Adam and Deans (2000) reported a turnaround of 40% of total responses within seven days while McDonald and Adam (2003) reported a much improved 85% of total responses within the same time frame. Table 1 summarises some of the advantages that have been commonly cited in a selection of past papers. Internet survey tools do have their drawbacks, however, with invitations sent online being sometimes perceived as spam, impersonal, or raising

potential respondents concerns about protection of their anonymity should they choose to participate (Evans and Mathur 2005). Also, the proliferation of research studies using this method leads one to caution that as respondents are increasingly tasked with numerous survey requests, their willingness to partake will diminish, meaning that future Internet surveys may suffer even more from low response rates. The benefits of online surveys appear to outweigh the drawbacks, and as a result of these highlighted advantages, Internet survey research has become commonplace, especially in pure marketing insights research where issues of representation are of less concern than in academic or government oriented research. Increasingly, however, Internet research is also being used to conduct academic quality research on a wide range of policy, health, social and business issues (Couper et al., 2007; also see Berrens et al., 2003 and Hung and Law, in press). To satisfy the rigorous expectations related to such applications, more attention is necessarily focused on quality issues of representativeness, generalisability and non-bias.

Several methods have been advanced in response to the need to achieve representativeness and mitigate non-response bias in online survey methods. Mathy et al. (2002) used block sampling to obtain a small but representative sample of a lesbian population from online and then compared their profile with the larger Gallup Poll sample. Overall they found that their sample was more robust and representative than the Gallup Poll sample, thus providing evidence that adapting strategies conventionally used in offline research can yield representative samples comparable to those obtained by more traditional means of sampling. Yun and Trumbo (in Andrews et al., 2003) achieved a 72% return rate across the sample of interest within a one-month period by combining postal, email and Web-based survey forms. Finally, Atkeson and Adams (2010) used a mixed mode approach to sample and administer a Web-based survey, concluding that this approach achieved a highly representative sample.

Table 1: Advantages of Internet Surveys (adapted from Hung and Law, in press).

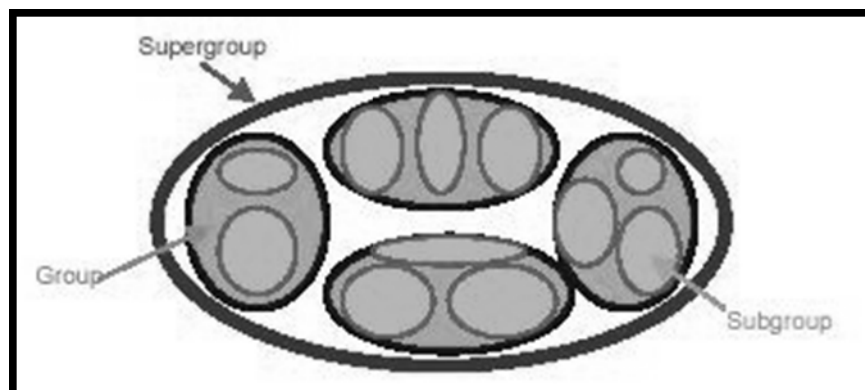
Low cost	Braunsberger et al., 2007; Duffy et al., 2005, Evans and Mathur, 2005
Fast response time	Couper et al., 2007, Evans and Mathur 2005
Easy access to wider range of geo-demographic populations	Couper et al., 2007; Fadner and Mandese, 2004
Instant data entry	Reips, 2002
Personalization	Ranchhold and Zhou, 2001
Wide geographic reach	Evans & Mathur, 2005
Recruiting difficult to reach samples	Mathy et al., 2002
Control of answer order	Evans and Mathur, 2005
Required completion of answers	
Interactive design/question	
Detail answers to open-ended questions	
Easy to obtain a large sample	
Easy follow up for non-responses	
Anonymity	Mathy et al., 2003
Avoid interviewer effects	Evans and Mathur (2005), Wiley et al. (2009)

METHODOLOGY

The purpose of our original research was to examine the relationship between adoption, usage and continuance of Internet shopping with individual personality variables, using the

Regulatory Focus (Higgins, 1997) framework, and to develop a generalised base model for understanding consumers' engagement with online shopping – the regulatory focus classification of online shoppers (REFCOS). To this end, we required a sample that reflected characteristics of current and possible Internet user population at a national level, in order to eliminate extant socio-demographic bias. Therefore, we surveyed a sample of the overall UK population using a multi-stage stratification scheme. The UK Office of National Statistics maintains a system of population groupings called "supergroups", a scheme segmenting the overall population based on shared social, economic and demographic characteristics (ONS, 2001). This segmentation uses output area classifications derived from the UK-wide census of 2001. Figure 1 below shows that there are seven supergroups overall, consisting of every postcode and address within the regions of the UK; however each supergroup is further broken down into "groups" and "subgroups", giving a final total of 52 clusters or subgroups. Various neighbourhoods from geographically dispersed areas of the UK are clustered into subgroups according shared characteristics, so that all officially identifiable population patterns are represented irrespective of geographic spread and location. The consequence of this is that capturing responses from a particular neighbourhood effectively equated to capturing responses from all neighbourhoods that belonged to that particular cluster. Therefore if at least one neighbourhood within a cluster responded adequately to the survey, then it was irrelevant if the remaining neighbourhoods demonstrated a poorer response rate, assuming no further intervening circumstances.

Figure 1: Structure of the output area classifications (source: ONS, 2001).



The number of neighbourhoods within a subgroup varies according to population size and number of postcodes but is generally between 20 and 50 neighbourhoods per cluster. It was these subgroups that we based our initial sampling on. We randomly selected an equal number of three neighbourhoods from each subgroup to generate a "neighbourhoods" sample of 156, with household addresses in these neighbourhoods constituting a final frame. From this "neighbourhoods" frame, we stepped down a level to randomly select a total of 4800 household addresses, which constituted our final sample. The decision to survey a large household sample was intended to account for invitees who might still not have participated in any form of online shopping. We did not, however, expect these particular type of invitees to be of a significant size, given that recent research suggests as much as 80% of UK adult population has used the Internet for shopping or another reason within the past year (Kalapesi et al., 2010), and that seven in ten households have Internet access (ONS, 2010).

A letter was addressed to the householder via surface mail inviting recipients to complete an online survey. To encourage participation and increase response rates, entry into a £250 prize draw was offered as a reward. The online survey was deployed using Limesurvey, a free survey tool. 655 useful responses were obtained from a total of 805 responses which represents a response rate of approximately 17%. This rate represents an improvement on other Internet surveys; but significantly, a review of the responses showed no systemic non-response bias as similar response rates were obtained from all supergroups and from subgroups within, as a result of which no targeted follow-up was deemed necessary.

RESULTS AND DISCUSSION

Response demographics are summarised in Table 2 above and show that all categories compare favourably with national characteristics when compared to the ONS statistics on national Internet user demographics (see ONS: Opinion Survey, 2010). In fact these results show that in both age and education categories, respondent demographics in this study were closer to actual national characteristics than to results reported by the ONS on Internet usage profiles in the UK. Thus whereas the ONS results show a significant skew for Internet usage towards respondents aged 18 to 24 years, our results show that Internet shopping usage is more general and reflective of national population characteristics. Similarly, using our method, we found that response rates from lower education categories were a significant improvement on reported findings. Although degree level demographics outperformed other categories, by comparison to ONS figures, respondents with lower levels of education improved considerably. Our gender demographic showed results similar to the ONS. These findings begin to illustrate the potential for improving Internet survey representativeness by using a combined recruitment and sampling. In this case, the use of a surface mail recruitment and invitation method combined with a Web-based survey instrument clearly improved response rates, reduced response bias and improved representativeness of the sample to the interesting population. This finding represents a key contribution to our knowledge about Internet research methods, especially with regards to survey sample planning. We now know that it is possible to adapt the methodology such that a scientifically rigorous approach can be adopted toward the recruitment plan, using a combination of Internet and non-Internet instruments. To this end, researchers can overcome the issues identified earlier about the generalizability, representativeness and usability of survey research that is undertaken on the Internet. Our contribution is in line with previous findings (for example Mathy et al. 2002, Atkeson and Adams, 2010) and provides further evidence that combining traditional methods of quantitative research with emerging electronic tools significantly improves the quality of the results obtained.

Table 2: Demographic Profile of Survey Respondents (source: authors)

<i>Age (n = 655)</i>		<i>Household Income (n = 655)</i>	
18–24 years	21.0%	<£25000	18.1%
25–34 years	22.0%	£25000 to 39999	25.4%
35–44 years	20.9%	£40000 to £74999	30.7%
45–54 years	17.5%	£75000 and above	25.8%
55–64 years	11.8%		
65 and over	6.8%		
Average Age	42.1		
<i>Education (n = 655)</i>		<i>Gender (n = 655)</i>	
Secondary level	26.5%		
College level	33.2	Male	47.8%
University level	40.3%	Female	52.2%

CONCLUSION

In this paper we have outlined the reasons why many researchers choose to adopt Internet surveys as their survey tool of choice, showing that many choose this method for cost considerations, and ease of administration. The Internet holds other advantages for its application as a data gathering medium. Various studies conducted via the Internet provide evidence that there are several benefits to undertaking this form of data gathering, especially when a survey approach is used. However, a number of setbacks have also been highlighted when using the Internet to conduct primary research; specifically, concerns have been raised about the representativeness of some samples that have been used in surveys, and as to whether results from such studies can be confidently generalised to the wider public. Hence it appears that for specific contexts and for clearly defined population boundaries, the issue of sampling for Internet research may not be so concerning, as sample frames can be defined from online based communities or recruited panels of Internet users. However when the population of interest is more general and the results to be obtained are expected to reflect characteristics of that general population, it is important to pay particular attention to satisfying the requirement for representativeness, both in the recruitment of the sample and the avoidance of non-response bias. Although various methods may be applied towards this end, in our study we chose to use a combined surface mail invitation and Internet-hosted survey to gather data from a national population, using predefined strata that were available at a national statistics provider. Our results vindicate our approach, as not only did we obtain a comparatively acceptable number of responses, but significantly, our sampled respondents' profiles showed a close semblance to the population of interest. As illuminated earlier, the economic case for using Internet surveys is already promising, but our research shows that to fully realise its potential, the Internet must be used appropriately for the purpose of undertaking quantitative research – in this case, it will benefit both researchers and practitioners. Researchers will be able to conduct high quality research while practitioners will be able to generalise and rely upon findings from such research in their business and economic applications.

Nevertheless we hasten to add that this approach may be used with care and caution. In the first instance, using groupings predefined by a third party could potentially be misleading. It is not explicitly clear whether initial criteria used for defining these groups was suitable for the purpose of our survey. Secondly, the population groupings by the statistics authority

were based on 2001 UK National Census figures, which at the time of our research had significantly aged. It is therefore possible that actual national demographics had changed and were no more reflective of our findings; furthermore, as this research only covered consumers within the United Kingdom, the applicability of its findings in other economies requires testing. Finally, one of the appeals of Internet research is the potential to significantly lower costs; but in contrast to this benefit, our research involved a significantly higher financial outlay than a purely Internet based study would have required. Future researchers should therefore consider innovating on ways to overcome these limitations and still achieve high quality Internet survey results.

REFERENCES

- Adam, S. & Deans, K. R. (2000). Online Business in Australia and New Zealand: Crossing a Chasm. Proceedings of AUSWEB2K, The Sixth Australian World Wide Web. Cairns: Southern Cross University, 19-34.
- Andrews, D., Nonnecke, B. & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction*, 16 (2), 185-210.
- Atkeson, L. R. & Adams, A. (2010). Mixed Mode (Internet and Mail) Probability Samples and Survey Representativeness: The Case of New Mexico 2008. Western Political Science Association 2010 Annual Meeting Paper. Accessed 15/01/2011 at <http://ssrn.com/abstract=1580588>
- Benfield, J. A. & Szlemko, W. J. (2006). Internet-based data collection: promises and realities. *Journal of Research Practice*, 2(2). Accessed 10/09/2010 at <http://jrp.icaap.org/index.php/jrp/article/view/30/51>
- Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C. & Weimer, D. L. (2003). The advent of internet surveys for political research: a comparison of telephone and Internet samples. *Political Analysis*, 11 (1), 1-22.
- Couper, M. P., Kapteyn, A., Schonlau, M. & Winter, J. (2007). Noncoverage and nonresponse in an internet survey. *Social Science Research*, 36, 131-148.
- Dolnicar, S., Laesser, C. & Matus, K. (2009). ONLINE VERSUS PAPER: Format effects in tourism surveys. *Faculty of Commerce - Papers* (2009). Accessed 02/09/2010 at <http://works.bepress.com/sdolnicar/226>
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 13, 195-219.
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52, 1280-1300.
- Hung, K. & Law, R. (in press). An overview of Internet-based surveys in hospitality and tourism journals. *Tourism Management*, article in press. Accessed 20/07/2010 at <http://www.sciencedirect.com/science/article/B6V9R-50DXCV4-2/2/ff1f288ebe77e35ee3045b93acb6d064>
- Kalapesi, C., Willersdorf, S. & Zwillenberg, P. (2010). The Connected Kingdom: how the Internet is transforming the UK economy. Report of the Boston Consulting Group, in association with Google accessed on 22/12/2010 at www.connectedkingdom.co.uk
- Mathy, R. M., Kerr, D. L & Hydin, B. M. (2003). Methodological rigor and ethical considerations in Internet-mediated research. *Psychotherapy: Theory, Research, Practice, Training*, 40 (1/2), 77-85.
- McDonald, H. & Adam, S. (2003). A comparison of online and postal data collection methods in marketing research. *Marketing Intelligence and Planning*, 21 (2), 85-95.
- Office of National Statistics (2010). ONS: Opinion Survey. Accessed 27/11/2010 at <http://www.statistics.gov.uk/cc/nugget.asp?id=8>

- Office of National Statistics (2001). ONS Neighbourhood Statistics. Accessed 12/01/2009 at <http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=userguide/detailedguidance/casestudies/censusareaclassification/case-studies-classification.htm>
- Zhao, W. & Jin, Y. (2006). A study of sampling method for Internet surveys. *International Journal of Business and Management*, April, 69-77.