WEB SURVEYS: CAN THE WEIGHTING SOLVE THE PROBLEM?

Vasja Vehovar, Katja Lozar Manfreda, University of Ljubljana; Zenel Batagelj, CATI Center
Katja Lozar Manfreda, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploscad 5, 1000 Ljubljana, Slovenia

Key Words: Web surveys, sampling, post-survey weighting

1. Introduction

Web surveys are promising another break-throu improvement to the survey industry. However, as usual, initial suspicions towards new methods are generally very strong, which could also be observed previously for other modes of computer assisted interviewing.

During the few years of their existence Web surveys provided us with a wide range of empirical evidence. Many advantages of this survey mode are reported (Balden 1999; Batagelj and Vehovar 1998a; Bates and Nichols 1998; Black 1998; Chisholm 1998; Clayton and Werking 1998; Comley 1997, 1998; Coomber 1997; Davis 1998; Dillman and Tortora 1998; Eaton 1997; Farmer 1998; Gonier 1999; Hollis 1999; Iyer 1996; Kehoe and Pitkow 1996; Kottler 1997a; Lanier et al. 1998; McCullough, 1998a; Nichols and Sedivi, 1998a; Onyshkevych and McIndoe 1999; Parackal and Brennan, 1998; Pitkow and Kehoe, 1995; Pitkow and Recker 1995; Schillewaert et al. 1998; Smith 1997; Spaeth 1999; Stanton 1998; Watt 1997; Weissbach 1997; Wydra 1999), however, certain obstacles do exist, particularly when the general population is targeted. Leaving some (temporary) technical problems aside, the key problems remain the Internet penetration within the general population and, even more seriously, the participation in these surveys. In this paper we concentrate on the specific aspect of Web surveys, i.e. using the post-survey weighting corrections as a remedy for these problems. These post-survey adjustments are not specific only to the Web surveys but are frequently discussed also in the context of the general non-response problem in other survey modes. Here, we discuss the problem within a framework of an empirical project based on comparisons of the results from a Web and a telephone survey.

2. The probability sampling

The Web surveys are used in many different contexts. They can be applied as a solicited survey for certain response-motivated sub-population of Internet users. They can be used also as an option in mixed-mode surveys. However, most frequently, they are used in market and opinion research for targeting the general (Internet) population. In this paper we concentrate on the latter usage of the Web surveys. The examples can range from general self-selected surveys (e.g. GVU's surveys (GVU 1994-1999)), through Web survey panels (e.g. Harris Pool Online, http://www.harrispollonline.com) to the examples of online votings (short question(naire) on some political or public issue, e.g. on the CNN site).

In this context we cannot, in general, speak about probability samples as there is either no probability mechanism of the sample selection or – if such a mechanism exists – the response rates are to low. An obvious question arises: Can such surveys be used in serious research? We often face the attempts of inferring valid conclusions from data collected with hazardous methods not only in the context of Web surveys, but also in general (e.g. surveys based on self-selected forms in newspapers and magazines, or 1-800 telephone surveys). There are countless examples where findings from such surveys have been severely wrong. From the Literary Digest example in 1936 (Albig, 1956: 180-182, 215) to other election disasters - the message has been always the same: these are simply not reliable methods.

On the other hand, one can not deny that a certain portion of useful information exists even in such self-selected surveys. At least, such findings throw light on the characteristics of the self-selection process, which is often of particular interest. In addition, the most intriguing property of these surveys is the fact that they are robust. We can observe this in household panels where the ad-hoc recruitment into large databases/frames, the quota selection and the post-survey adjustments completely replace the probability sampling procedures. Similar is true for many telephone surveys based on an ambiguous sampling frame with a low response rate and vague within-household selection procedure. And examples of successful Web surveys (more exactly, Web survey panels with proper quota selection and weighting procedures) giving same results as parallel telephone surveys also exists (Terhanian and Black 1999a).

We inevitably face a practical and common sense argument: If the non-probability sampling methods provide good enough data or data with predictable error, why pay more for a sophisticated, time-consuming and expensive data collection? This simple argument gives enough justification for this paper to observe the differences between the Web and

telephone data collection and evaluate the extent of improvements from post-survey weighting corrections.

3. The data

In June and July 1998, the third RIS Slovene national Web survey was launched (http://www.ris.org). It was widely advertised in traditional media and on the Web. Additional to this, persons included in the public email directory (19,000) were solicited by an email message with two follow-ups. However, this frame which contains only about one fifth of all email addresses in Slovenia is not representative for the target population of all Internet users. Despite the majority of the respondents in the Web survey actually came from the explicit email solicitation, we treat the entire sample as being self-selected, since there is ambiguity regarding the eligibility of the units in the frame, and also due to the low response rate. This is additionally justified since responses from the email solicited respondents do not differ from the respondents who had self-selected themselves from other sources.

After the Web survey was completed a large telephone screening (n=10,000) of the general population was launched (in September 1998) in order to locate Internet users who were involved in (or aware of) the Web survey. Only the users that had been using the Internet at least monthly were surveyed. The telephone numbers were selected from the public telephone directory (the telephone coverage is around 90% and there are only a few percent of unlisted numbers), and a random procedure was applied to select a person within the household. The same block of target questions was used as in the Web survey. The users in the telephone survey who were aware of - but did not cooperate in - the Web survey were asked also questions concerning their reasons for (non)participation in this survey.

In this paper we are going to compare the data from both surveys and evaluate the post-survey weighting corrections for estimates from the Web survey.

4. Target population

The Slovenian active population of one and a half million (aged between 15-60) can be divided into non-Internet users (87%) and Internet users (13%). Internet users can be further-on divided to those who were aware of the RIS project (32%) and within them those who were aware of the RIS 98 Web survey (33%). Aware users were email solicited (47%) and not solicited (53%). We define the non-respondents as the users who knew about the Web survey yet did not answer the Web questionnaire (54%). Response rate among those aware of the survey was 53% for email solicited and 23% for self-selected users. The most frequently mentioned reasons for not participating were lack of time (27%) or interest (11%).

The above percentages show that the Web respondents were thus basically selected only from the 11% of the total population of Internet users in Slovenia (only these were aware of the survey). Any inference about all Internet users implicitly assumes a random selection that determines the selection of persons that are "aware of the survey". The same assumption applies to the non-response process.

5. The two populations

There are, of course, large differences between the Web respondents and the regular Internet users measured by the probability telephone survey with a representative frame. Obviously, intensive users dominate in the Web survey as showed by the following analysis including monthly users from both surveys (95% of the Web survey respondents and 50% of persons who have already used the Internet in the telephone survey).

The cluster analysis (K-means) of the Web respondents regarding their use of Web, email, telnet, IRC, USENET, FTP and MUDs shows six homogeneous clusters: four large and two small ones (where MUD users were concentrated). The largest group (one third) – *experienced users* - regularly uses Web and email, and often uses the FTP service. The next two groups represent each around one fifth of the total Web respondents: *intensive users*, using all services *except MUDs*, and the *IRC users*. One quarter of Web respondents comprises less intensive users who use the Web often and email occasionally - *non-intensive users*. The smallest two groups are, as mentioned, *intensive users* who use all services, *including MUD*, and *only MUD users*.

These clusters were compared to the representative telephone survey data. In contrast to the Web respondents, the largest part of Internet users from the telephone surveys is included to the group of non-intensive users (62% in comparison to 25%). Nobody was included in the groups of users using MUD and only 4% to the group of intensive users without MUD. Additional 10% were IRC users and 24% experienced users.

6. The differences

Despite the differences in the structure of respondents, the key variables were surprisingly close. For example, the percentage of monthly users who had performed some shopping in the previous 12 months is the same in both surveys (19%) and the same is true also for the frequency distribution of the amount of money spent on on-line purchase.

However, on the attitudinal questions there are mixed results. Answers to four out of seven items related to the electronic commerce were significantly different (t-test, sign. < 0.05) in the two samples. It seems that the items where extensive experience with the Internet is needed lead to the creation of significant differences. For example, web respondents were more likely to agree with the statement 'I'm am very interested in electronic banking' (3.8 in comparison to 3.4 on a five-point Likert scale) but there were no significant difference in agreement with the statement 'I'm very worried about the misuse of credit cards' (4.05 in comparison to 3.95). Similarly, two out of seven items regarding the role of different institutions in the development of the Internet in Slovenia, resulted in statistically significant difference. Their opinions are close as regards the role of government institutions but different as concerns the role of non-government institutions. We can conclude that the differences are not only small but also predictable. The Web respondents are more intensive users and they prefer public access provider vs. the national monopolist (Telekom Slovenia) which is the largest commercial provider.

The largest differences were observed with the measurement of the Web page visits. In Figure 1 the horizontal axis presents the percentage of respondents who have never visited a particular site, and the vertical axis presents the proportion of respondents who visit this page weekly or even more often. Clearly, the figures from the telephone survey are generally closer to the right/downwards direction, i.e. less weekly visitors and more non-visitors.

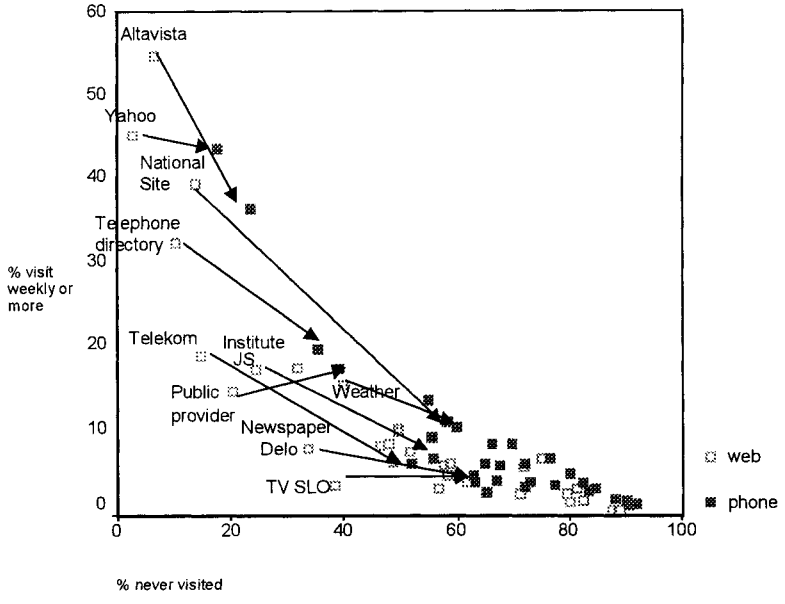

Figure 1. Differences in WWW page visits: Web vs. telephone survey

7. The weighting

The weighting procedures are a well-established practice in the survey industry and their theoretical background is well known (Kish 1992; Little and Rubin 1989; Meadow et. al. 1983; Sardnal 1990). The issue of the non-response and/or non-coverage problem is often an important reason for weighting. However, the effect of such weights is generally relatively small (Groves et al. 1988). Some empirical research in social sciences has shown that weighting could remove half of the non-response (or non-coverage) bias if we have all the relevant co-variables (Vehovar 1992). But since in the survey practice we often miss the necessary co-variables the improvements are usually much smaller or even negligible (Groves and Couper 1998).

In our example, the weighting procedures were implemented in order to correct the estimates from the Web survey. The following variables were used in a variety of combinations: gender, age, frequency of Internet use, the first year of Internet use, email usage, access from home, computer orientation and knowledge of English. The margins from the telephone survey were used as a basis for the raking (RIM weighting) procedure. Due to the large differences in the structure of respondents in the two surveys, the maximum weight obtained was generally above 100. The most over-represented users were email users, regular (weekly) users and English-speaking users. Therefore in the next step the respondents using the Internet less than weekly were excluded together with the Internet users without

email; however, the maximum weight dropped only to 25, which is still relatively high.

Due to this, another weighting procedure was implemented. In order to detect the variables that explain the differences in both surveys, the CHAID method was used. Data collection mode was the dependent variable, and the independent variables were as follows: use of email, frequency of Internet use, access from home, knowledge of English, computer orientation, the year of first use of the Internet, gender and age of respondent. The Web respondents were weighted (W1) by the margins from the telephone survey according to the above segments obtained by CHADE. Additional weighting procedure (W2) was performed using raking with the above segmentation variable and also with gender and age.

Figure 2 illustrates the example of the frequency of Alta Vista page visits which shows that weighting "works" only in the sense of offering the right direction. On the other hand, the above mentioned attitudes towards the electronic commerce successfully improved when the second weighting (W2) was used.

As showed, the results are not very optimistic, at least with the page visit items. Further research is needed since different data collection modes – in addition to under-coverage of certain segments - can also introduce discrepancies because of mode effect. The above results may also suggest that certain variables related to the Web-page-visit behavior must be additionally included in the weighting procedures as the independent variables.

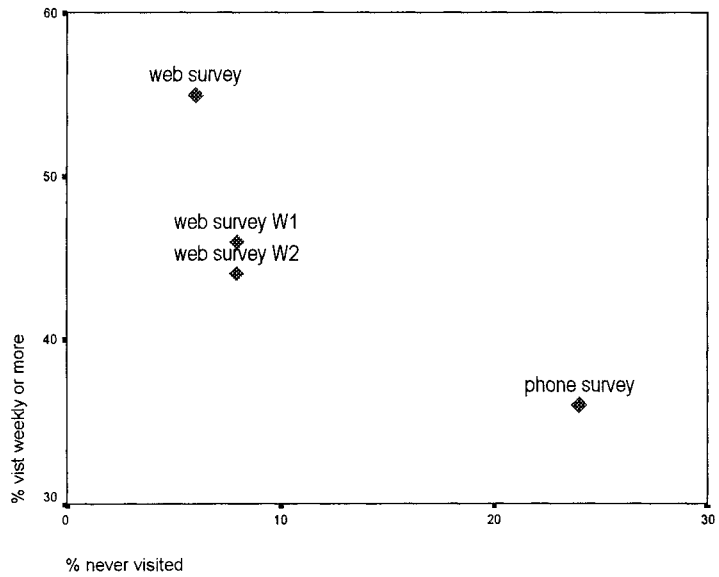

Figure 2. The effects of the weights on measuring Alta Vista visits

8. Discussion

The Web surveys targeting the general Internet population suffer from the inferential problem that cannot be solved. Such problems are common to all surveys with large non-coverage, large non-response or poor sample selection mechanism (Vehovar 1992). However, when inferring on the total Internet active population, the comparison of the telephone and the web survey results shows that errors are relatively small. The differences narrow further if we compare the Web respondents only with the intensive (e.g. daily) users from the telephone survey. This is, at least, the empirical evidence for a particular (Slovenian) social environment. Some research from the US (Flemming and Sonner 1999) shows larger discrepancies, at least with respect to political attitudes. However, in this example the general population was compared to the Web survey respondents, therefore discrepancies would be smaller if the population of intensive Internet users was compared.

When the non-response and the non-coverage problem in the self-selected Web surveys are extremely large the potential for severe discrepancies in survey results is large. The non-coverage – people not using the Internet and users not aware of the survey - is generally larger than 95% of the active population and the non-response is generally larger than 60%. These numbers are, in fact, the lower limits obtained in a very particular setting of a national survey in a small country. In practice the above figures are often much larger. The potential for errors is thus very large.

With the increased Internet usage a part of such discrepancies can disappear, however, this is not the case with the cooperation rates, which may decline even further.

Additional problem regarding cooperation is the fact that in the Web surveys the intensive Internet users and technically oriented users respond significantly more often (Kehoe and Pitkow 1996; Pitkow and Kehoe 1995; Pitkow and Recker 1994). Similar also holds true for younger, male, educated and wealthy respondents. For the non-English speaking nations the fluency in the English language is another factor. As a consequence, the Web respondents are recruited from an extremely narrow segment of the Internet users. Any weighing procedure thus inevitably produces a large range of weights. The standard social-demographic variables have a relatively small effect in the weighting procedures. The results are slightly better when Internet-specific variables are applied, such as computer orientation, measured by the readership of computer magazines.

We can conclude that when comparing the intensive Internet users from a telephone survey and the respondents from a Web survey the differences are generally small which offers a good ground for targeting this population with the Web surveys. However, the difference increases when comparing the Web results with a more general population. The weighting has a relatively limited impact. It improves the estimates, so it should be performed in this type of surveys, but it cannot solve the problem.

9. References

Albig, W. (1956): Modern Public Opinion. McGraw Hill, New York..

Balden, W. A. (1999): "Project Landmark. A Comprehensive Study to Determine if The Internet Presents a Valid Data Collection Alternative for Mainstream Consumer Goods and Services". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 14-27.

Batagelj, Z., and Vehovar, V. (1998b): "Technical and Methodological Issues in WWW Surveys". Paper presented at AAPOR '98, St. Louis. http://www.ris.org/ris98/stlouis/index.html

Bates, N., and Nichols, E. (1998): "The Census Bureau WWW Hiring Questionnaire: A Case Study for Usability Testing". Paper presented at JMS 98, Dallas, Texas.

Black, G. (1998): "Internet Surveys – A Replacement Technology". Paper presented at AAPOR '98, St. Louis.

Chisholm, J. (1998): "Using the Internet to Measure Customer Satisfaction and Loyalty". The Book of Papers from the Worldwide Internet Seminar 1998 in Paris, France.

Clayton, R. L., and Werking, G. S. (1998): "Business Surveys of the Future: The World Wide Web as a Data Collection Methodology". In: Couper, M. et al. (Eds.), Computer Assisted Survey Information Collection. John Wiley & Sons, New York, pp. 543-562.

Comley, P. (1997): "The Use of the Internet for Opinion Polls". Learning from the Future: Creative Solutions for Marketing. The Book of Papers from the 50th ESOMAR Marketing Research Congress in Edinburgh, Scotland. http://www.virtualsurveys.com/papers/Poll.htm

Comley, P. (1998): "On-Line Research: Some Methods, Some Problems, Some Case Studies". In Westlake, A. et al. (Eds.), New Methods in Survey Research 1998. Proceedings of the ASC International Conference, a satellite meeting for Compstat 98. http://www.virtualsurveys.com/papers/ASC.htm

Coomber, R. (1997): "Using the Internet for Survey Research". Sociological Research Online, 2, 2. http://www.socresonline.org.uk/socresonline/2/2/coomber.html

Davis, G. (1998): "Using the Internet for Employee Surveys". Market Research Library. http://www.researchinfo.com/library/internet_employee/index.shtml

Dillman, D. A., and Tortora, R. D. (1998): "Principles for Constructing Respondent-Friendly WEB Surveys and Their Influence on Response". Paper presented at JSM 98, Dallas, Texas.

Eaton, B. (1997): "Internet Surveys: Does WWW Stand for 'Why Waste the Work?'". Quirk's Marketing Research Review, June/July. http://www.quirks.com/CGI-BIN/SM40i.exe?docid=3000:58911&%70assArticleId=244

Farmer, T. (1998): "Using the Internet for Primary Research Data Collection". Market Research Library. http://www.researchinfo.com/library/infotek/index.shtml

Flemming, G., and Sonner, M. (1999): "Can Internet Polling Work? Strategies for Conducting Public Opinion Surveys Online". Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida.

Gonier, D. E. (1999): "The Emperor Gets New Clothes". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 8-13.

Groves, R. M., and Couper, M. P. (1998): Nonresponse in Household Interview Surveys. John Wiley & Sons, New York.

Groves, R. M. et al. (Eds.) (1988): Telephone Survey Methodology. John Wiley & Sons, New York.

GVU (1994-1999): GVU's User Survey. http://www.gvu.gatech.edu/user_surveys

Hollis, N. S. (1999): "Can a Picture Save 1,000 Words? Integrating Phone and Online Methodologies". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 41-49.

Iyer, R. (1996): "The Internet: A New Opportunity for Marketing Research Firms". Quirk's Marketing Research Review, May. http://www.quirks.com/CGI-BIN/SM40i.exe?docid=3000:58911&%70assArticleId=46

Kehoe, C. M., and Pitkow, J. E. (1996): "Surveying the Territory. GVU's Five WWW User Surveys". The World Wide Web Journal, 1, 3. http://www.cc.gatech.edu/gvu/user_surveys/papers/w3j.html

Kish, L. (1992): "Weighting for uniqual Pi". Journal of Official Statistics, 2.

Kottler, R. E. (1997): "Exploiting the Research Potential of the World Wide Web". Paper presented at Research '97, London. http://www.quantime.co.uk/web_bureau/knowledge/r97a.htm

Lanier, A. et al. (1998): "Web-Based Data Collection in National Science Foundation Surveys". Paper presented at JSM 98, Dallas, Texas.

Little, R. J. A., and Rubin, D. B. (1987): Statistical Analysis with Missing Data. John Wiley & Sons, New York.

McCullough, D. (1998): "The Internet: A New Horizon for Conducting Marketing Research". In: Herrington, D. J., and Taylor, R. D. (Eds.), Proceedings of the Annual Meeting of the Society for Marketing Advances. Society for Marketing Advances, VA, pp. 256-260.

Meadow et al. (1983): Incomplete Data in Sample Surveys. Academic Press, New York.

Nichols, E., and Sedivi, B. (1998a): "Economic Data Collection Via the Web: A Census Bureau Case Study". Paper presented at JSM 98, Dallas, Texas.

Onyshekvych, V., and McIndoe, D. (1999): "Internet Technology: Gaining Commercial Advantage". Paper presented at the 1999 AAPOR Conference, St. Petersburg, Florida. http://surveys.over.net/method/Onyshkevych.ZIP

Paracka, M., and Brennan, M. (1998): "Obtaining Purchase Probabilities via a Web Based Survey: The Juster Scale and the Verbal Probability Scale". Marketing Bulletin, 9, pp. 67-76.

Pitkow, J. E., and Kehoe, C. M. (1995): "Results from the Third WWW User Survey". World Wide Web Journal, 1, 1.
http://www.gvu.gatech/user_surveys/papers/survey_3_paper.html

Pitkow, J. E., and Recker, M. M. (1994): "Results from the First World-Wide Web User Survey". Journal of Computer Networks and ISDN Systems, 27, 2.
http://www.gvu.gatech/user_surveys/papers/survey_1_paper.html

Pitkow, J. E., and Recker, M. M. (1995): "Using the Web as a Survey Tool: Results from the Second WWW User Survey". Journal of Computer Networks and ISDN Systems, 27, 6.
http://www.gvu.gatech/user_surveys/papers/survey_2_paper.html

Sardnal C. E. et al. (1997): Model Assisted Survey Sampling. Springer-Verlag, NY.

Schillewaert, N. et al. (1998): "Non Probability Sampling for WWW Surveys: A Comparison of Methods". Journal of the Market Research Society, 4, 40, pp. 307-313.

Smith, C. B. (1997): "Casting the Net: Surveying an Internet Population". Journal of Computer Mediated Communication, 3, 1.
http://www.ascusc.org/jcmc/vol3/issue1/smith.html

Spaeth, J. (1999): "At The Dawn of a New Day". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 5-7.

Stanton, J. M. (1998): "An Empirical Assessment of Data Collection Using the Internet". Personnel Psychology, 51, 3, pp. 709-725.

Terhanian, G., and Black, G. S. (1999): "Understanding the Online Population: Lessons from the Harris Poll and the Harris Poll Online". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 28-33.

Vehovar, V. (1992): Imprecision of the Estimates From Telephone Samples. MA thesis, University of Essex.

Watt, J. (1997): "Using the Internet for Quantitative Survey Research". Quirk's Marketing Research Review, June/July. http://www.quirks.com/CGI-BIN/SM40i.exe?docid=3000:58911&%70assArticleId=248

Weissbach, S. (1997): "Internet Research: Still a Few Hurdles to Clear". Quirk's Marketing Research Review, June/July, 1997. http://www.quirks.com/CGI-BIN/SM40i.exe?docid=3000:58911&%70assArticleId=249

Wydra, D. (1999): "Online Tracking: A New Frontier". ARF's Online Research Day - Towards Validation. Advertising Research Foundation, New York, pp. 34-36.