

WageIndicator



in Argentina, Belgium, Brazil, Denmark, Finland, Germany, Hungary, Italy, India, Mexico, Netherlands, Poland, Spain, South Africa, South Korea, United Kingdom, United States

www.wageindicator.org

THE *WAGEINDICATOR* WEB-SURVEY: INTRODUCING WEB-SURVEYS

Kea Tijdens, AIAS/University of Amsterdam, Netherlands

15/01/2007

Understanding the Internet and web traffic

Since the start of Internet, in the early 1990s, the world has faced a continuous boost in the number of websites, web-visitors and applications. Today, the Internet has become worldwide a major medium both for providing and finding information and for communication, gradually replacing television, radio, and printed media, respectively telephone, fax and post. With many millions of visitors every day, firms have shifted parts of their advertising budgets from other media to the Internet, leading to advertisement revenues for frequently visited websites. Consequently, web marketing is increasingly deployed in attracting large numbers of visitors. In contrast to a few years ago, the major players on the Internet are aware that they aim to attract large masses of web-visitors, and increasingly they know how to do so. Moreover, insight in web traffic has increased quickly.

Providing free information for the public at large is a major backbone of the Internet, in recent years encouraged by adds revenues. This has widened the variety of free information provided, a tendency recently in particular strengthened by Google. The free content encyclopedia project Wikipedia contains more than 3 million articles in over 200 languages. It is written collaboratively by volunteers under the slogan: "Imagine a world in which every single person is given free access to the sum of all human knowledge."

In addition, the Internet is used for eliciting reactions from the public at large. Large audiences are using eBay for trading items, special websites for congratulations or mourning messages, websites for poll voting for singers or others, and many other topics. Every time, the numbers of visitors are astonishingly high within and across national borders.

Eliciting the public is increasingly used for scientific data collection too. For example, with only developing a game with a prize incentive, tremendously large numbers of web-visitors are willing to help science by naming photographs, leading to the labeling of nearly 20 million images, so that search machines can find them. This was supervised by professor Luis von Ahn, Carnegie Mellon University, Pittsburgh, USA, who is currently developing an exiting plan for the translation of all languages into all languages, with the help of large masses of web-visitors worldwide.

So, if cases prove that providing free information presents an added value to the audience all over the world and that web-visitors are voluntarily willing to provide information, with or without a prize incentive, could this be applied to information

gathering and providing concerning wages? Yes, it can. The *WageIndicator* websites in 17 countries have shown to be able to provide information on wages and in return ask the visitors to complete a questionnaire on work and wages. The public at large in these countries has shown a desire for information about wages, and it has shown to be willing to complete voluntarily a web survey in return for this free information.

The digital divide

The low Internet access rates in some countries are a major argument against worldwide web surveys. In 2004, Internet access rates ranged from 788 per 1,000 inhabitants in New Zealand to 2 per 1,000 inhabitants in Egypt. Particularly in Africa, access rates are low. For two reasons Internet access rates can be expected to grow. First, urbanization is a continuous process and Internet is better accessible in urban areas. In Sub Saharan African countries for example average annual growth of internet dial-up traffic is currently mainly driven by increases in the number of customers through both residential and public *access in cyber cafes, and not by increasing* mainline telephone services.¹ Therefore, Internet access rates will increasingly underestimate Internet use.

Second, the so-called digital divide across countries reflects a persistent gap in the availability of mainline telephone services. Yet, cell phones may be a promising new platform for *Internet access*. Currently, major computer and telecom firms are developing technologies to make Internet accessible through cell phones, aiming at markets in developing countries. Compared to telephone line, mainline telephone services the diffusion rates of cell phones are in many developing countries much higher.² Particularly in Asian and Latin American countries, Internet access rates may increase fast in the coming years. In African countries, prospects may be slightly different, but here also growth can be expected. Here, we may expect an increase of wireless Internet.

With an increasing share of the worldwide population using the Internet, it has become an interesting medium to collect volunteer survey data worldwide, while at the same time providing free information.

Volunteer web surveys: recruitment

Sampling methods are classified as either probability or non-probability. In probability samples, each member of the target population has a known non-zero probability of being selected. This method requires sampling frames. A range of sampling should preferably match fully the target population. For sampling, in most countries a range of sampling frames is available, from a list of dwellings, address lists or a list of telephone numbers. Sampling for the latter is called random digit dialing. In case no sampling frame is available that provides adequate coverage of this group, a probability sample can be drawn from a sampling frame covering a larger population. This sample is then surveyed with only one question, the so-called screening question, to identify whether the sampled individual is part of the target population. If so, the full questionnaire can start. The major advantage of probability sampling is that sampling error can be calculated. Yet, in industrialized countries, probability samples increasingly face difficulties with non-response, which may be troublesome when the non-response is selective.

Sampling methods for web surveys are also classified as either probability or non-probability. Probability sampling is used by market research companies, who draw random or stratified samples from so-called access panels, consisting of email addresses

¹ Rogy, Michel (2005) *Broadband Technologies and Services in Sub Saharan Africa: The Case of ADSL, Opportunities for Operators and Challenges for Regulators, Communications and Strategies*, November 2005, pp. 95-105

² Dasgupta, Susmita; Lall, Somik; Wheeler, David (2005) *Policy Reform, Economic Growth and the Digital Divide*, *Oxford Development Studies*, 33/2, pp. 229-43

of volunteers for which a number of socio-demographic variables are known. Apart from discussions about this sampling method, these marketing companies exist in industrialized countries only. Non-probability methods are used in open web surveys, in particular the so-called convenience sampling. These volunteer web surveys are primarily held in marketing and in voting research, and hardly used in the field of work and employment, except for the *WageIndicator* project.

In non-probability sampling, members are selected from the population in some nonrandom manner, among others by means of convenience sampling. In non-probability sampling, the degree to which the sample differs from the population remains unknown. Non-probability sampling is used in preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample. It is used in cases where sampling frames are absent or where the target population is a rare group, as this requires expensive screening questions.

In the scientific literature about web surveys, little attention has been paid to recruitment of volunteer surveys, in contrast to extensive attention paid to panel recruitment in sampled surveys. Lee (2006) in his paper 'Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys' just notices that the panel recruitment in volunteer web surveys is done via some type of advertisement, such as banner ads, pop-up ads, or e-mails. In recent years however insight in how to reach large masses of web visitors has increased tremendously and web marketing may aim for a heterogeneous public. Volunteer web survey on websites with an attractive content for the public at large may attract large masses of visitors. Moreover, through targeting-marketing methods tailored to web surveys, underrepresented groups can be addressed. In conclusion, 'Putting a Questionnaire on the Web is not Enough', as Faas and Schoen (2006) phrase it, is indeed definitely not enough for a volunteer survey. When recruitment for volunteer web surveys is based on web marketing efforts that aim at a large and heterogeneous public, the distinct between probability and non-probability-sampling may become less sharp.

In case web marketing has become a critical feature in recruiting for volunteer web surveys, it will take time before a marketing effort pays off. Once the web-site is frequently visited, it is wise to continue and to profit from the initial investments. Thus, it becomes profitable to employ continuous web surveys. In the field of work and employment, continuous surveys are primarily found among the statistical agencies employing continuous labor force surveys, mostly for reasons of estimating seasonal fluctuations in participation rates. This insight challenges the perspective of surveys, the most common being the once-only surveys.

Volunteer web surveys: understanding web traffic

Since the start of Internet, in the early 1990s, the world has faced a continuous boost in the number of websites, web-visitors and applications. Today, the Internet has become worldwide the major medium both for providing and finding information and for communication, gradually replacing television, radio, and printed media, respectively telephone, fax and post. With many millions of visitors every day, firms have shifted parts of their advertising budgets from other media to the Internet, leading to advertisement revenues for frequently visited websites. Consequently, web marketing is increasingly deployed in attracting large numbers of visitors. In contrast to a few years ago, the major players on the Internet are aware how to attract large masses of web-visitors. Moreover, insight in web traffic has increased quickly. Providing free information for the public at large is the backbone of the Internet, in recent years encouraged by adds revenues. This has tremendously widened the variety of free information provided, a tendency recently in particular strengthened by Google.

In addition, the Internet is used for eliciting reactions from the public at large. Large audiences are using eBay for trading items, MSN for meeting friends, special websites for congratulations or mourning messages, websites for poll voting for singers or others, or

many other topics. Every time, the numbers of visitors are astonishingly high within and across national borders. The free content encyclopedia project Wikipedia contains more than 3 million articles in over 200 languages. It is written collaboratively by volunteers under the slogan: 'Imagine a world in which every single person is given free access to the sum of all human knowledge.'

Eliciting the public is increasingly also used for scientific data collection. For example, with only developing a game with a prize incentive, tremendously large numbers of web-visitors are willing to help science by naming photographs, leading to the labeling of nearly 20 million images, so that search machines can find them. This was supervised by professor Luis von Ahn, Carnegie Mellon University, Pittsburgh, USA, who is currently developing an exciting plan for the translation of all languages into all languages, with the help of large masses of web-visitors worldwide.

So, if cases prove that providing free information presents an added value to the audience all over the world and that in return web-visitors are voluntarily willing to provide information, with or without a prize incentive, could this be applied to information gathering and providing concerning wages? Yes, it can. The *WageIndicator* websites in 17 countries have shown to be able to provide information on wages and in return ask the visitors to complete a questionnaire on work and wages. The public at large in these countries has shown a desire for information about wages, and is also willing to complete a web survey.

Volunteer web surveys: trust and high wage quality information

In all participating countries a desire for free wage information is apparent, leading to high numbers of web-visitors to the national websites. Each month the *WageIndicator* website in the Netherlands has over 300,000 unique visitors. It ranks among the highest dedicated websites in the country. Within a few years most national *WageIndicator* websites become middle-sized Internet players in their countries, given proper publicity and web-marketing. The Belgian website for example started in 2004, was elected the best website of the year 2006, and it now attracts 40,000 visitors a month. The German website is growing fast. By the end of 2006 it had 100,000 visitors a month. The website in Argentina, only launched early 2006, now has already 30,000 visitors a month and the highest intake of questionnaires per visitor.

Web marketing is complicated but critical in attracting web visitors. The marketing policy unfolds along several lines. First, the information presented with the *Salary Check* and at the web pages must offer high quality. Second, visitors must trust both the information provided and volunteering their data in the survey. Trust is ensured by cooperating with top universities, with confederations of trade unions such as UNI, FNV, DGB, or TUC, with the ILO in some other countries, or with career sites. Third, cooperation with web-based partners is critical. Current cooperation includes major web-portals, such as UOL (Brazil) and Microsoft's MSN portal in UK, and The Netherlands. Fourth, cooperation with media groups or publishing houses with a strong Internet presence is critical. Current cooperation includes *Gazeta Wyborcza* (Poland), *Politiken* (Denmark), *Mail & Guardian* (South Africa), *El Pais* (Spain), *Rheinische Post*, *Stern*, *Stuttgarter Zeitung*, *Sueddeutsche* and a few other newspapers (Germany), *Business Week* (USA), *Times of India* (India, pending) and *De Telegraaf* (the Netherlands). Fifth, cooperation exists with the career site *Monster*, the world's largest job site, notably with *Job Korea*, *Monster.co.uk*, *Monster.be* and *Monster.nl*. In addition, in some countries national websites cooperate with temp agency sites or with employment agency sites. Sixth, all national teams aim for free publicity.³ Finally, each country has at least one website, but sometimes more. This is quite natural in multilingual countries, like Belgium. The US team is planning a Spanish website. Quite a number of national teams created a separate website addressing women. The Netherlands has additional websites for youth, 40+, the

³ See results at <http://www.wageindicator.org/main/WageIndicatorgazette/inthepress>.

health care and the self-employed. All these websites offer different content and look-and-feel, but the salary check and the web survey are similar.

In recent years several new web tools have been developed, such as the gross-net checker, the minimum-wage checker, the lifetime earnings checker, and the time-budget checker. These tools all contribute to attracting visitors. Whenever possible, the checkers are up-scaled to the national teams who want to include them in their website(s). A great success is the VIP-pay check, sporting wages of football players, movie stars, heads of state, etc.

Volunteer web surveys: data quality

It is sometimes assumed that a web survey cannot be taken seriously, because the Internet is associated with one-second-a-page visitors, who have neither time nor patience to complete a 20-minute questionnaire. From the consistency of the data, however, it can be concluded that this hardly applies to the *WageIndicator* survey. Item non-response is mostly below 5 percent. From the respondent's emails and their comments at the end of the questionnaire, we have learned that the vast majority of the respondents answer the questions with great care. Many do report that they enjoyed completing the questionnaire.

In addition, some people have pointed to the risk that an occupational group may systematically report earning higher wages than in reality. By doing so, the *Salary Check* in their country would come up with higher wages, which may in the long run lead to higher wages for the occupational group at large. This argument assumes that respondents within an occupation will act collectively. Of course this may happen, but we assume this to be seldom, because it assumes a highly organized, geographically not dispersed occupational group, that undertakes action to report higher wages. In case extreme wages are reported, these are excluded in the analyses for the *Salary*. Another threat might be more serious. Some countries are known for machismo bluff about earnings to family, friends and others. If this behavior is extended to wage reporting in questionnaires, it may lead to serious problems. In those countries, surveys, whatever survey mode is used, are not an adequate tool for collecting wage data. Whenever a country is likely to show this behavior, the data needs to be checked with other sources. As for the assumption that the *Salary Check* may influence national wage setting processes, two counter-arguments apply. First, for many years it is known that national wage setting processes are primarily dependent upon factors related to the economy and the industrial relations. Second, both employees and employers can and do use the *Salary Check*. Therefore, if any influences could be traced, it could be both ways.

In the case of countries with an important non-registered labor force, Internet-based surveys may allow to get better information than traditional surveys. In cases where wages are underdeclared or plainly nondeclared when the respondent -be the worker or the boss- has to answer a questionnaire face to face to an interviewer it is more likely to get the right information when the worker is sitting alone in front of the computer. Thus web surveys may capture some part of the informal labor market not covered or covered with wrong information by traditional surveys.

The traditional way of testing the quality of a questionnaire is a pilot study to identify potential problems with the survey's design while there is still time to fix them. Continuous volunteer web surveys offer two advantages in this respect, because questions can be adapted while the survey is running, and because the test population is much larger than usual in any other survey mode. Whenever an additional country joins, the Questionnaire Management System is updated, or new questions are added to the survey. Then the questionnaire is tested using the public at large for a couple of days, of course only after being tested by the *WageIndicator* team.

Continuous volunteer web surveys may use multiple client-side feedback systems for improving the questionnaire. In the case of the *WageIndicator*, visitors' emails to the

national web managers provide feedback on the questionnaire and its technical functioning. In addition, the text box at the end of the questionnaire 'what did you like about the questionnaire' allows additional input. Finally, dropout during completion provides insight in difficulties filling out the questionnaire. In the past few years, web-visitor's comments have led to several improvements in the questionnaire. Moreover, when continuous volunteer web surveys lead to large sample sizes, it is particularly profitable to invest in survey quality.

Volunteer web surveys: tackling self-selection

Apart from its many advantages, the *WageIndicator* questionnaire has one major flaw, because it is a volunteer web survey. Individuals in the target population, i.e. the labor force, do not have an equal chance to access the survey and therefore the data are not representative for the population. The selectivity is threefold. The first selectivity is associated with Internet access as this may be related to wages, the key variable in the data. Second, although the numbers of visitors of the *WageIndicator* websites are large and growing, it is still a minority of the extremely large and heterogeneous population visiting the Internet. The self-selection into the *WageIndicator* websites may be related to interest in wages, and therefore maybe high-wage earners could be under-represented. Third, once visiting the *WageIndicator* website, the visitor still has to decide whether or not to complete the questionnaire. This self-selection into the web survey may be related to availability of time, satisfaction with the website or altruism to contribute to the project, all factors which may be related to the key variable. The exact nature of these three steps of self-selection need to be studied in the years to come.

Results are available of studies in a number of countries to investigate the national *WageIndicator* bias as for socio-demographic variables.⁴ The distribution of socio-demographic variables that were assumed to be subject to bias, notably gender, age, working hours and education, in the Netherlands *WageIndicator* dataset 2000-2004 was compared in detail to the distribution of these variables in the labor force, based on annual aggregate data of the national Labor Force Survey. The main conclusions are as follows. As for working hours, the part-time labor force is under-represented in the *WageIndicator* sample, particularly for respondents working less than 20 hours a week. As for age, older workers are under-represented in the sample. This particularly applies to individuals aged 55 and over. From 2001 to 2004, however, the under-representation was reduced. In addition, the share of young workers in the labor force decreased, whereas it fluctuated in the *WageIndicator* survey, not decreasing over-representation. As for gender, the male labor force is under-represented in the sample, which may be due to the fact that the survey initially addressed women only. From 2002 to 2004, however, the under-representation reduced. As for education, the low skilled labor force is under-represented in the sample. From 2002 to 2004, however, the under-representation reduced.

WageIndicator employs several strategies to cope with the self-selection in the volunteer sample. The first strategy relates to the web marketing, whereby a broad target population is defined, including marketing aiming at sub-populations, such as women or youth. The second strategy relates to the special routing through the questionnaire, designed to address marginal groups in the labor force, as they have a higher likelihood of dropping out during completion. The third strategy relates to weighting the dataset for under- and over-represented groups, currently developed for the European countries in the project, using aggregate national labor force survey data. The fourth strategy relates to asking a few questions in the web survey, similar to those asked in other major surveys, such as the United States Labor Force Survey, the World Values Survey, and the

⁴ Kevätsalo, K. (2006) *WOLI WEB national report. Some preliminary results from Finland*. Amsterdam: WageIndicator; Tijdens, K.G. (2006) How biased is the Netherlands WageIndicator dataset? Amsterdam: WageIndicator; Depedraza, P. (2007) Weights for WageIndicator data. Amsterdam: WageIndicator

European Survey on Working Conditions, allowing to the micro-data with *WageIndicator* data and subsequent weighting. The fifth strategy relates to a full reference survey in one country, controlling for self-selection effects beyond socio-demographic variables. Such a study is foreseen in the Netherlands. These strategies combined are assumed to lead to a sample sufficiently corrected, so that it may confidently be used for analyses.

As for the fifth strategy, volunteer web surveys are primarily conducted in marketing and political voting research. According to the standard methodological literature, volunteer samples can never be representative. Therefore a number of commercial market research agencies (Harris Online, Bloomerco, McKinsey) apply a correction technique for their volunteer web surveys using an extra so-called reference survey. This reference survey is a real probability sample in which each member of the population has the same probability of selection. The volunteer sample is adjusted to the probability sample, using the Propensity Score Adjustment (PSA), which is a statistical approach for self-selection.⁵ This methodology definitely needs further exploration, but if successful, volunteer web surveys with reference surveys offer unprecedented opportunities in terms of sample sizes and country coverage.

Volunteer web surveys: advantages of large sample sizes

Having said that the Internet reaches out for millions of people, a major advantage of volunteer web surveys is said too. Volunteer web surveys allow for extremely large sample sizes, unmet in other survey modes. These large sample sizes are also not met in probability sampled web surveys. Here, the sample sizes are limited to the size of the access panel.

For several reasons, large sample sizes advantageous. First, the data from large-scale web-surveys allow for exploring small-scale units such as regional intersections, or occupations, because each unit still has sufficient data. In the field of work and employment studies, the typical small-scale studies address an organization, an industry, an occupation or a region. In large-scale data, even rare groups still have sufficient numbers of observation. These small-scale units either lack adequate sampling frames, so the research could not have been undertaken otherwise, or it is simply too expensive to use a wider sampling frame with a screenings question. When the screened population is relative small in comparison to the sampled population, the costs involved with the screening may be larger than the costs of the survey itself.

Second, large scale data allow for analyses of small-scale units. This is a passive advantage of investigating large sample sizes. Large-scale surveys allow for an active form of addressing small-scale units, as they allow for screening questions in the survey. Subsequently, follow-up questions in the web-survey can address the sub-populations, e.g. asking respondents in some occupations about early signs of specific occupational diseases.

A third major advantage of volunteer web surveys is that it can be continuous. Of course, other survey modes can be used continuously as well, and it is done so for example in labor force surveys, conducted by statistical agencies. Continuous volunteer web surveys are particularly profitable because the initial investments in the survey are relatively high, but the running costs are relatively lower, even when taken the marketing efforts in account.

Fourth, large-scale web surveys are advantageous as they offer multiple client-side feedback systems, both active by visitors' email and comments in the questionnaire's text boxes and passive by their dropout behavior. For improving web surveys, client-side feedback is extremely helpful.

⁵ Taylor, 2000; Schonlau et al 2002 and 2004; Danielsson, 2004; Varedia and Forsman, 2003; Biffignandi et al 2005; Varedian, 2005; Isaksson & Lee, 2005; Lee, 2006

Finally, web surveys offer great opportunities for temporary plug-in modules on specific items, related to the issue of work and wages.