DOING SURVEYS ONLINE

Vera Toepoel

SAGE
Los Angeles | London | New Delhi
Singapore | Washington DC

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Jai Seaman
Assistant editor: James Piper
Production editor: Victoria Nicholas
Copyeditor: Camille Bramall
Proofreader: Jennifer Hinchliffe
Indexer: Linda Haylock
Marketing manager: Sally Ransom
Cover design: Shaun Mercier
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

Library of Congress Control Number: 2015936510

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-4462-4966-6
ISBN 978-1-4462-4967-3 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the Egmont grading system. We undertake an annual audit to monitor our sustainability.

1

INTRODUCTION TO ONLINE SURVEYS

1.1 Introduction

In the past 50 years technological developments have had a great impact on society as a whole. This influence is reflected in the way present day survey research is conducted.

Advancements in technology have created a more individualized society and affected communication and accessibility between individuals. With the rise of the Internet, we have witnessed a shift to online surveying being the dominant mode of survey data collection.

The base for online surveys lies in the 1980s. It was then that the groundwork was made to allow email to be used as a new manner to collect survey data (Bethlehem and Biffignandi, 2012). This was one of the first steps in the creation of the nowadays commonly used online survey. Online surveys rapidly started to challenge the dominance of other modes of data collection, such as telephone, face-to-face and paper-and-pencil surveys, because of the advantages they offered in terms of cost and speed. Costs for hiring and training interviewers, recording equipment, travel or paper and postage could be eliminated. Furthermore, Web surveys allowed researchers to collect data while working on other tasks and reach large numbers of people in a short time, even though there might be a great geographical distance (Wright, 2005). The suitability of Web surveys as a mode of administration depends on the Internet penetration rate of the population: in other words who can (or more importantly, who cannot) be reached when using the Web as a mode of communication. For most countries, however, the Web has become the major mode of data collection.

This chapter will introduce you to Web survey research; it will illustrate the place of the Web survey in the world of survey research, and exemplify the challenges online surveys (still) face and the possibilities they offer.

1.2 Online surveys for quantitative and qualitative research

Online surveys can serve both qualitative and quantitative research questions, although they are more commonly used for quantitative research. At the start of their project, researchers should ask themselves what kind of question they wish to answer with their study. Different types of questions mean different kinds of data requirements. The first decision a researcher has to make is to choose whether to collect quantitative or qualitative data. **Quantitative research** is used to quantify a research problem by way of generating numerical data that can be used for statistical testing. **Qualitative research** is primarily used to gain an understanding of underlying reasons, opinions or motivations. It follows a less structured path compared to quantitative research and is often used for exploratory research. Qualitative and quantitative research are often combined in mixed-method research, to combine the best of both worlds.

1.2.1 Quantitative online research

First of all, the wide geographical reach of Web surveys offers the possibility to quickly and easily create a great sample of people. Quantitative Web surveys can be

used if one wishes to answer a question about great groups of people and/or generalize its results, for example, make inferences out of the sample to a general population. Because of the low costs and short time needed, Web surveys can be a great method to perform large-scale research. Quantitative research focuses more on the ability to perform statistical analyses (Eysenbach and Wyatt, 2002). In quantitative Web surveys, most questions are close-ended. This means that the respondent is restricted to a range of answering options and has to choose one. Close-ended questions are also called quantitative questions because the use of response options allows the researcher to convert the answers into numbers, thereby facilitating the statistical analyses (Epstein, 2013).

1.2.2 Qualitative online research

Not only does the Web survey enable researchers to reach a multiplicity of potential respondents, it also makes it possible to approach or analyse groups of people one would not easily find outside of the Internet. These groups are called Internet communities. Internet communities offer people a way to communicate with peers. They include chat rooms, newsgroups, mailing lists or discussion boards on websites (Eysenbach and Till, 2001). They can even be found in online games. Internet communities can serve as rich sources of qualitative data for researchers (Eysenbach and Till, 2001).

As opposed to quantitative studies, qualitative studies do not require representative samples. Researchers conducting qualitative research are not interested in the average score of a population. They want to obtain a profound understanding of particular groups or individuals and therefore have to deliberately look for specific groups or individuals (Eysenbach and Wyatt, 2002). A distinction can be made between three different research methodologies for qualitative research on the Internet:

1. ***Passive analysis***. In this type of online research, researchers analyse interactions and information on the Internet, for example, interactions in chat rooms, without actively communicating with the people from the Internet communities themselves.
2. ***Active analysis***. Just like in passive analysis, in this type of research the researcher observes the participants. The difference is that the researcher here gets actively involved in the interactions, often without revealing his or her identity as a researcher. Actively interacting with the participants does allow researchers to ask questions to which they want an answer.
3. ***Web surveys***. Web surveys are the only type of the three qualitative online research methodologies in which the researcher informs the participants of his or her identity, making it more ethical than the other two types (Eysenbach and Wyatt, 2002).

Qualitative Web surveys answer 'how' or 'why' questions. Questions in this type of survey are often open-ended, thereby offering the respondents the possibility to answer

freely in any manner they choose. Response options are not specified and responses are measured and judged by feel rather than by statistics and mathematics. This brings the advantage that, in case the researcher had no idea what the answer possibilities could be, respondents have a free rein to answer, making qualitative surveys fit for explorative research. Although Web surveys can be used for qualitative research, the majority of Web surveys have a quantitative nature. Web surveys can be divided into a range of different approaches, varying in goals and inference possibilities.

1.3 Different types of Web surveys

By now Web surveys have diffused into a variety of types that differ in aspects such as manner of recruitment and scientific value. Two categories of Web surveys that are prevalent are probability- and non-probability-based Web surveys (Couper, 2000).

1.3.1 Non-probability Web surveys

Non-probability surveys are surveys in which sampling is done in a non-random manner and there are members of the target population who have a zero chance of being selected for the survey sample (Neuman, 2012). In the group non-probability surveys, it is possible to distinguish three different types. These are: **polls as entertainment**, **unrestricted self-selected surveys** and **volunteer opt-in panels** (Couper, 2000).

1.3.1.1 Polls as entertainment

Web surveys as a form of entertainment are a type of survey popular on the Internet and are generally not considered to be a scientific method, but it may sometimes occur that people confuse these surveys with real research-oriented surveys. The primary intention of the creation of these surveys is to entertain, which they often state explicitly. The **'question of the day' poll** is a good example of a Web survey as entertainment. This is a poll that is popular among many media sites. It only polls the opinions of those people who happen to come across it and wish to take part in the poll. It does not reflect the general opinion of all Internet users. Because this kind of Web survey usually does not lead to generalizations or have a lasting value, it does not form a serious threat to scientific Web surveys (Couper, 2000).

1.3.1.2 Unrestricted self-selected surveys

Unrestricted self-selected surveys are very prevalent and sometimes claim to have scientific value, as opposed to polls as entertainment. Therefore, the self-selected Web survey might put legitimate scientific Web surveys in great jeopardy. One could consider the self-selected Web survey to be the digital equivalent to magazine insert surveys.

Participants are recruited using methods such as the placement of announcements and ads on high traffic sites. The sample that results from this recruitment is not representative because of the fact that only the people who respond to such advertisements will participate in the survey, thereby making it a self-selected survey (Couper, 2000).

1.3.1.3 Volunteer opt-in panels

The number of volunteer panels of Internet users has grown vastly in the industry of Web surveys. It does not use a random sample but rather appeals to people who volunteer in answering survey questions, thereby creating a volunteer panel. It collects basic demographic information about these volunteers at the time of their registration. This makes it possible to create a large database consisting of potential participants who could join in on later surveys. Even though this type of Web survey offers more control in the selection of participants, and panel members for particular surveys are selected from the pool of potential respondents using quota sampling or probability sampling methods, it is important to keep in mind that the initial panel does consist of a self-selected group of volunteers, as is the case for the other two types of non-probability surveys (Couper, 2000).

Proponents of volunteer opt-in panels frequently claim that these panels are equal to or even better than types of survey data collection that are based on probability sampling. This makes the volunteer opt-in panels of concern. Some volunteer panels even claim to be representative of the general population. It is necessary to evaluate their methods empirically, openly and thoroughly to establish their validity (Couper, 2000).

1.3.2 Probability-based Web surveys

As opposed to non-probability Web surveys, in **probability surveys** the recruitment process is random (from a list) and all members of the target population have a chance to be selected to participate in the survey (Neuman, 2012). This means that every individual of the population has a chance to express their opinion. Probability-based surveys can be divided into five different subtypes. These are the **intercept surveys**, **list-based samples**, **Web option in mixed-mode surveys**, **pre-recruited panels of Internet users** and **pre-recruited panels of the full population**. Two approaches can be distinguished: one approach restricts the sample to people with access to the Internet, and the other employs alternative methods to aim for a sample for which all people in the population can be selected.

1.3.2.1 Intercept surveys

Every 'nth' person who visits a particular website is invited to participate in the Web survey. Cookies are often used in intercept surveys to prevent one person from being

invited to the survey multiple times. While generalizations based on this type of survey can only be fairly limited (due to the fact that inference is only possible based on the people who visit the website), intercept surveys prove to be very useful to evaluate sites and to measure customer satisfaction (Couper, 2000).

1.3.2.2 List based samples

Another type of probability-based survey is the list based sample of high-coverage populations. In this type of Web survey it is taken into account that Web surveys are not omnipresent because of the fact that they only cover the population using the Internet. The usual approach in list based samples is to begin with a list of people who have access to the Internet and to send them invitations to participate in the survey by email. A good example of this approach is the student survey.

1.3.2.3 Web option in mixed-mode surveys

In mixed-mode designs with a choice of completion method, the Web is viewed as an alternative that can be used in mixed-mode designs. People who do not have access to the Internet have the possibility of completing the survey in a different mode, for example, telephone or paper-and-pencil. This possibility is very useful in panel surveys of establishments where lack of access to or a preference for a particular mode should not be a barrier to inclusion in the survey.

1.3.2.4 Pre-recruited panels of Internet users

Pre-recruited panels of Internet users are quite similar to the non-probability type 'volunteer panels of Internet users'. The difference between the two types is that in pre-recruited panels of Internet users the panel members are solicited using other (offline) modes, for example, **random digit dialling** (RDD, by random dialling telephone numbers) or **address-based sampling** (ABS, by using addresses as the sampling frame), to enforce probability-sampling methods. In this method potential panel members who emanate from the probability sampling using other modes are asked whether they have Internet access, and only the ones who do are selected for the panel.

1.3.2.5 Pre-recruited panels of full population

Last, but certainly not least, is the pre-recruited panel of the full population. This probability-based survey is the only type to even offer the potential of obtaining a

probability sample for the entire population, and not only for people with access to the Internet. This Web survey type enables non-Internet users, who would otherwise have been excluded from the sample, to participate in the survey. To make this possible, people without prior access to the Internet are provided with alternative equipment, for example, an Internet connection and a (simple) computer. This type of survey offers the opportunity to compare the non-response bias in Web surveys with that of surveys outside of the Web. Moreover, this is the only type of Web survey that can be generalized beyond the Internet population (Couper, 2000). The difference from a mixed-mode sample is that respondents complete the survey in the same online mode, because the right equipment is provided.

Although these probability methods start with probability samples, in which the probability of each member of the population being selected for the sample can be determined accurately, representativeness cannot be guaranteed in these samples, since non-response could harm the validity of the survey. This is a serious threat in times of falling response rates. However, in contrast to non-probability surveys, these methods do enable researchers to measure the sources of non-response and provide them with information about the group of interest and knowledge on the process of recruitment.

1.4 Total survey error

Any survey has to deal with different types of associated errors. The term '**total survey error**' (TSE) refers to a conceptual framework, which describes statistical error properties of sample surveys. The framework offers a central structure to the field of survey methodology and attempts to display a balance of costs and errors at the design stage of the survey (Groves and Lyberg, 2010).

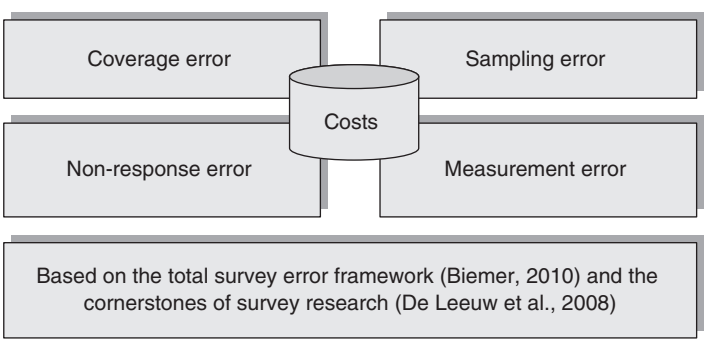

Figure 1.1 Total survey error framework: Cornerstones of survey research

Survey fact: In 2005, workshops called 'The International Total Survey Error Workshops (ITSEW)' were established to encourage research on error sources and interactions between these errors (Biemer, 2010).

Survey researchers can use the total survey error paradigm as a criterion to plan their study. To assess the survey quality and evaluate sources of error in the implementation, the 'fitness for use' perspective can be used as an estimate. It acknowledges that survey quality can be perceived from a lot of different perspectives by producers and users of survey data (Biemer, 2010). While producers often prioritize qualities like high sample size and a high response rate, data users often place a higher priority on qualities such as accessibility and usability of the data (Biemer, 2010). Survey researchers should choose the design that brings the smallest total survey error while taking the cost of that design into account (Groves and Lyberg, 2010). To establish a quality questionnaire, methodologists and statisticians strive for accuracy as the main criterion. There are four fundamental concepts to design a good and accurate survey. These are 'accurate measurements', 'high response rate', 'high coverage' and 'proper sampling' (de Leeuw, 2012). Errors in each of these cornerstones can heighten the total survey error.

1.4.1 Measurement error

Measurement error can be one of the most harmful sources of error for many surveys (Biemer, 2010). It refers to the deviation of the answers retrieved from the respondents from their true value (Couper, 2000). This type of error includes errors that arise from survey questions, from respondents and from multifarious conduction factors. The design of a questionnaire can be a major source of error. Confusing instructions, ambiguous questions, poor visual design and complicated terms can all lead to measurement error. Respondents may intentionally or unintentionally respond incorrectly to the survey questions or may be faced with different conditions to other respondents (Biemer, 2010).

As well as noting that the display of the questionnaire can differ amongst respondents, researchers should also take into account the fact that online surveys allow respondents to complete the questionnaire at a time and place of their own choice, meaning that they cannot control for environmental factors which might influence the respondents' answering behaviour. For instance, when collecting data on sensitive subjects, for example, sexual behaviour or drug use, honest disclosure of the respondents could be compromised were they to complete the questionnaire with other members of the household present (Biemer, 2010). In addition, there is no interviewer present to guide the response process.

1.4.2 Non-response error

Non-response error basically consists of three types of error due to the missing of answers. One is the **unit non-response error**, which refers to the occurrence that a sampled unit, such as a household or an individual, does not respond to the questionnaire (Biemer, 2010). The other type of non-response error is **item non-response error**. This occurs when respondents fail to complete an individual item within the survey. In addition, respondents might prematurely end their participation, leading to **partial non-response error**.

Types of non-response:

1. **Unit non-response:** individuals not participating in the survey.
2. **Partial non-response:** individuals who drop out during the survey.
3. **Item non-response:** individuals who fail to provide the answer to an individual question or item.

Non-response does not always have to cause problems for the outcome of the data. Non-response only leads to error when the non-respondents differ from the respondents precisely in the features that are central to the questionnaire. If this is not the case, non-response is random and not a threat to the validity or reliability of the results.

Non-response appears to be a greater problem in online surveys in comparison to other methods. Recent meta-analyses, in which different methods were compared, show that the non-response rate of online surveys is generally approximately 10 per cent lower than the average of other methods (de Leeuw, 2012). The largest amount of non-response occurs at the first approach to the respondents. People might be unwilling to participate in a survey or confuse the invitation to participate in the survey with spam advertisement.

1.4.3 Coverage error

Coverage error is an error of non-observation in which there are discrepancies between the **frame population**, that is, a list (or another mechanism) of the target population used to draw the sample, and the actual target population (Groves and Lyberg, 2010). Online surveys have become a popular method to study samples from great populations (Schonlau et al., 2002). If researchers only mean to investigate the online population, non-coverage is not necessarily a problem; however, researchers often aim to make generalizations that go beyond the online population, which makes non-coverage a threat to the validity of the survey.

Coverage error mostly occurs because the demographic groups that do not have access to the Internet can differ substantially from the groups of people who do have access. This discrepancy is called the **digital divide**. The degree to which coverage error has an impact on the validity of the survey data is not only influenced by the extent of the digital divide in the subject of the study, but also by the **Internet penetration**, i.e. the percentage of the population that has access to the Internet, of the target population (Mohorko et al., 2013).

Survey fact: People who do have access to the Internet are often richer, younger and more educated, which means respondents with these demographic features are more likely to be over-represented in online surveys (Rookey et al., 2008).

1.4.4 Sampling error

Just like in the case of coverage error, in **sampling error** the respondents in the sample do not accurately represent the frame population and therefore the validity of the survey can be harmed. The difference between these two sources of error is that in coverage error the frame does not represent the target population correctly, while in sampling error the frame might be accurate, but the sample that is drawn from the frame population is not. Sampling error occurs in the selection of members of the frame population to participate in the sample (Couper, 2000). Sampling error often relates to the size of the sample relative to the size of the population. Sampling error can be problematic if the sample size is too small. For example, if you have a bag of candy as your population, chances are that if you draw three sweets out of the bag, you get different ones than you would if you took out three more. You might have two pink ones and one white one at the first draw, and a yellow, white and a pink one at the second draw. There are ways to calculate optimal sample size (see Chapter 4) in order to reduce sampling error. Selecting members to participate in online surveys can be problematic in any case, since there is no frame list of all Internet users from which a random sample could be drawn.

1.5 Dimension differences between modes

Survey **modes** – such as online surveys, telephone surveys, face-to-face interviews and paper-and-pencil surveys – can cause different degrees and types of error. Modes can differ in a variety of dimensions. We will discuss the position of Web surveys in comparison to other modes with respect to five different dimensions. These are the degree of interviewer involvement, the degree of interaction with the respondent, degree of privacy, channels of communication and the use of technology (see Groves et al., 2009).

1.5.1 Degree of interviewer involvement

The degree to which the interviewer is involved in the completion of the questionnaire differs between modes. For example, interviewers administer face-to-face surveys and telephone surveys. They read the questions to the respondents and record their answers, resulting in great interviewer involvement. In other modes, the respondents can administer the questionnaire themselves (e.g. in an online or a paper-and-pencil survey). Self-administered surveys are less influenced by **social desirability bias** (respondents giving social desirable answers rather than the truth) than interviewer-administered questionnaires (Kreuter et al., 2008). Moreover, self-administrated questionnaires save researchers time and are generally less costly than interviewer administration. There are also disadvantages to the absence of an interviewer. A low degree of interviewer involvement can lead to errors of non-observation and measurement errors, for example, due to the inability of a researcher to probe, motivate and assist the respondent (Couper, 2011).

1.5.2 Degree of interaction with the respondent

The degree of interaction with the respondent is related to the degree to which the interviewer is involved. In a face-to-face interview, the interviewer has direct contact with the respondent and is able to exchange both verbal and non-verbal cues. Since telephone interviews only allow verbal cues, there is less direct contact involved in telephone surveys (Couper, 2011). In Web surveys direct contact can be completely absent. Still, Web surveys can vary in the amount of contact with the respondents. Modes in which there are videos or pictures of the interviewers reading the survey questions, or the usage of animated conversational agents or avatars representing the interviewer, incorporate a higher degree of interaction with the respondent than Web surveys without such features (Couper, 2011).

Survey fact: The connection between social desirability bias and interaction

Even though the presence of humanizing cues in Web surveys (such as animated conversational agents) could be expected to result in greater degrees of social desirability bias, computer administration of surveys appears to reduce the effects of social desirability on sensitive topics, even when humanizing cues are featured. The difference between social desirability bias in the actual presence of an interviewer and the presence of the interviewer on a video in Web surveys can be explained by the absence of direct contact in Web surveys (Fuchs and Funke, 2007).

1.5.3 Degree of privacy

A third dimension in which modes can vary is the degree of privacy the respondent is granted during the completion of the survey. A medium level of privacy can be found in interviewer-administered surveys, since there is the possibility of other people listening in to the interview. Web surveys allow people to complete the questionnaire on any computer

they want in total privacy, reducing the chance that people might try to see or overhear their answers, and offering respondents more anonymity than most other modes can.

1.5.4 Channels of communication

Typically, telephone surveys only offer the aural channel of communication and paper-and-pencil surveys (provided that they are completed without the presence of the researcher) only the visual channel. Face-to-face surveys can incorporate both these channels, since this mode makes it possible to deliver questions and answers orally, as well as allowing visual support, for example, by the use of non-verbal cues or the showing of cards. Even though most Web surveys are presented in written words, thus using the visual channel, Web surveys offer more options of presentation since this mode allows the incorporation of audio or video support to enhance the survey experience (Couper, 2011).

1.5.5 The use of technology

Lastly, modes can vary in the degree to which computer technology is used. In some survey modes no technology is used. This can be the case in paper based surveys. It is also possible that only the interviewer will use technology, for example in Computer-Assisted-Telephone-Interviewing (**CATI**) or Computer-Assisted-Personal-Interviewing (**CAPI**) (Bronner and Kuijlen, 2007). Further along the continuum we will find Computer-Administered-Self-Interviewing (**CASI**), where the respondent uses a computer, provided by the survey organization, to complete the questionnaire without having an interviewer administer the questionnaire (Couper, 2011). The Web survey finds itself at the end of the continuum. Both the researcher and the respondent use technology and the respondents can administer the questionnaire from their personal device (Couper, 2011).

Table 1.1 Mode differences among different dimensions

	Web	Face-to-face	Telephone	Paper-and-pencil
Interviewer involvement	-	+	+	-
Interaction	+/-	++	+	-
Privacy	+	-	-	+
Channel of communication	Visual (aural)	Aural (visual)	Aural	Visual
Technology	++	+	+	-

1.6 Internet penetration over time and place

Access to technology is very important in the implementation of Web surveys. The usability of an online survey is affected by the Internet penetration of the target population. There is a wide variability of Internet penetration across different countries and areas.

Survey fact: History reflected in Internet penetration

Even historical events can be reflected in Internet penetration. When looking at the German population, one can observe a lingering effect of the East–West division. While the Internet penetration in the former West Germany is approximately 61 per cent, only 48 per cent of the people living in the former East Germany have access to the Internet (de Leeuw, 2012).

The average Internet penetration is currently highest in North America, with a percentage of 87.7 of the population (Internet World Stats, 2014). Figure 1.2 displays the Internet penetration in the world regions.

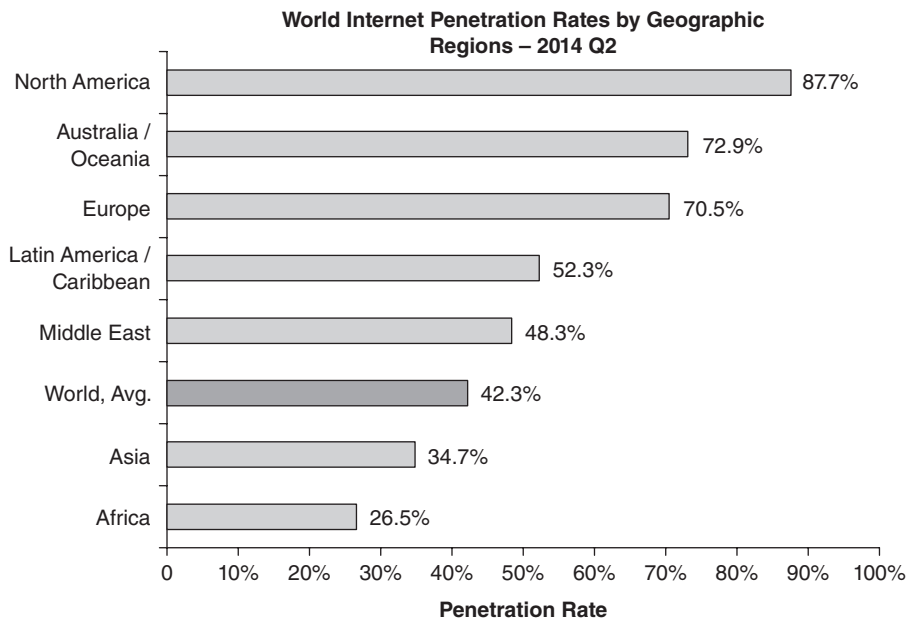

Figure 1.2 Statistics on the global Internet penetration

Source: Internet World Stats – www.internetworldststs.com/stats.htm Penetration Rates are based on a world population of 7,182,406,565 and 3,035,749,340 estimated internet users on June 30, 2014.

Even though the Internet penetration is only 34.7 per cent on average in Asia, if one were to draw a probability sample of the entire worldwide Internet population, it would result in a sample in which the greater part would be Asian citizens because of the high numbers of people living in Asia (Internet World Stats, 2014). This distribution is exhibited in Figure 1.3.

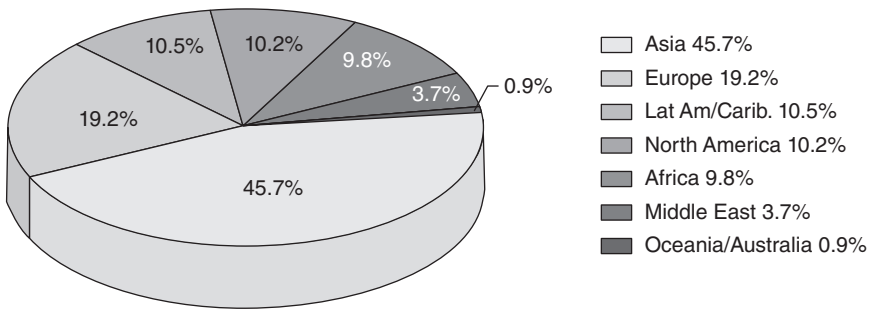

Figure 1.3 Internet users in the world, distributed by world regions (Internet World Stats, 2014)

Source: Internet World Stats – www.internetworldstats.com/stats.htm

Basis: 3,035,749,340 Internet users on June 30, 2014

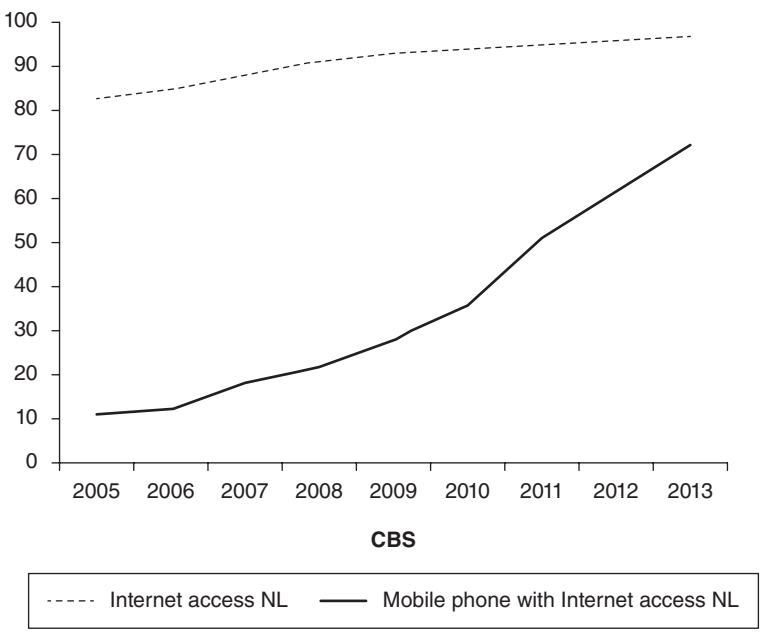

Figure 1.4 Development of Internet access and mobile phone with Internet access in the Netherlands

Internet penetration has increased rapidly over the past decade. Especially in regions where the Internet penetration was still at a low percentage, the increase has been tremendous. For example, between the years 2000 and 2014, Internet penetration in

Africa has grown with a quantity as high as 6500 per cent (Internet World Stats, 2014). Even though the Internet penetration is still relatively low in third world countries (as can be seen in Figure 1.2), the rapid growth offers positive prospects for their access to the Internet and thereby for their ability to participate in Web surveys. Following a vast increase in Internet penetration rates, we see now a vast increase in the number of mobile phones with Internet access, as can be seen for the Netherlands in Figure 1.4. Mobile phones suggest new challenges in online surveys.

1.7 Web surveys on mobile devices

As well as the increasing Internet penetration on computers, there is also a considerable growth in the number of people using the Internet on mobile smart phones and other mobile devices with a high-speed connection to the Web, thereby bringing new challenges to Web survey researchers.

FEATURED STORIES:

THE WALL STREET JOURNAL.

Europe Tops Global Smartphone Penetration

'Europe is projected to dominate global smartphone penetration by 2017, according to a report published Wednesday. The eMarketer report shows seven of the top 10 countries by penetration will be in Europe (in order: Norway (1), Denmark (2), the Netherlands (3), Sweden (4), Finland (6), the UK (7) and Germany (9). South Korea (5), Australia (8) and Japan (10) are the only [countries outside of Europe] to break in.The US is 11th.'

Figure 1.5 Wall Street Journal report on smart phone use

Source: ecommerce.com

An important challenge for researchers arises from the fact that mobile phone screens are of a small size in comparison to a computer screen and are operated by either a small keyboard or a touch screen. The growth of Web survey research means that researchers should take the size and other features of these devices into account in the design of their surveys (Callegaro, 2010). This increase in Internet use on mobile phones has already led to surveys being deployed using mobile browsers of an app-like design (Buskirk and Andres, 2012).

If Web research on mobile devices were to become more commonly used, it could offer a variety of potential benefits. First of all, it offers the possibility to use the RDD technique, which could bring about a solution to the problem that there are no random based sampling methods for Web surveys that are analogous to methods used in telephone surveys to generate telephone numbers. Furthermore, being able to complete questionnaires on mobile devices brings more freedom to the respondents to choose a convenient place and time. This could give mobile Web research the potential to reduce

non-response rates. The Internet is an evolving field and developments such as the growth of mobile Internet use keep challenging researchers to refine the Web survey in all its possibilities.

Summary

The online survey has grown to be the most prominent method used for data collection. It offers advantages such as low cost, quick completion and the possibility to easily contact people over great geographical distances. Web surveys can be used to collect great amounts of data in little time using quantitative methods as well as to collect qualitative data one would not easily find outside of the Internet. Two main types of Web surveys can be distinguished: the probability- and the non-probability-based survey. In probability-based surveys, the entire target population has a random chance of selection for the sample. This is not the case in non-probability-based samples, for example, where respondents volunteer. Both types can be divided into a variety of subtypes. Most types struggle with the problem of representativeness and their results cannot be generalized beyond the Internet population. The exclusion of people without Internet access is one of the main threats to the validity of Web surveys by causing coverage error. Coverage error is one of the bases that can heighten the total survey error, next to measurement error, non-response error and sampling error. Other modes struggle with these errors to different degrees. This can be caused by their discrepancies across a variety of dimensions. The degree of interviewer involvement, interaction with the respondent, privacy, channels of communication and the use of technology can all influence the reactions and interpretations of the respondents as well as the researcher. Lately, new challenges have arisen for researchers, as Web surveys can also be completed on mobile devices, bringing the challenge of designing the surveys to work on a variety of devices. This is only one of the challenges and possibilities likely to come.

Key terms

Address-based Sampling
CAPI
CASI
CATI
Coverage error
Digital divide
Frame population
Intercept surveys
Internet penetration
Item non-response error
List based samples
Measurement error
Modes
Non-probability surveys
Non-response error
Partial non-response error
Polls as entertainment
Pre-recruited panels of the full population
Pre-recruited panels of Internet users
Probability-based surveys
Qualitative research

Quantitative research

Question of the day poll

Random digit dialling

Sampling error

Social desirability bias

Total survey error

Unit non-response error

Unrestricted self-selected surveys

Volunteer opt-in panels

Web option in mixed-mode surveys

Exercises

1. In what type of research would it be preferable to ask qualitative survey questions instead of quantitative questions?
2. Explain the difference between probability-based Web surveys and non-probability Web surveys and relate this to the problem of generalizability.
3. Explain the relation between non-coverage, error, digital divide and Internet penetration.
4. How is self-administration related to social desirability bias, and what does this mean for Web surveys?
5. Explain the current state and developments of Internet penetration, and its relation to population size.

References

Bethlehem, J. and Biffignandi, S. (2012) *Handbook of Web Surveys*. Hoboken, NJ: John Wiley and Sons, Inc.

Biemer, P.P. (2010) 'Total survey error: Design, implementation, and evaluation', *Public Opinion Quarterly,* 74 (5): 817–48. doi: 10.1093/poq/nfq058 (For more information about total survey error)

Bronner, A.E. and Kuijlen, T. (2007) 'The live or digital interviewer: A comparison between CASI, CAPI and CATI with respect to differences in response behaviour', *International Journal of Market Research*, 49 (2): 167–190.

Buskirk, T.D. and Andrus, C. (2012) 'Smart surveys for smart phones: Exploring various approaches for conducting online mobile surveys via smartphones', *Survey Practice*. Retrieved from: http://surveypractice.wordpress.com/2012/02/21/smart-surveys-for-smart-phones

Callegaro, M. (2010) 'Do you know which device your respondent has used to take your online survey?' *Survey Practice*, 3 (6).

Couper, M.P. (2000) 'Web surveys: A review of issues and approaches', *Public Opinion Quarterly*, 64: 464–94. (For more information about types of Web surveys)

Couper, M.P. (2011) 'The future of modes of data collection', *Public Opinion Quarterly*, 75 (5): 889–908. doi: 10.1093/poq/nfr046

De Leeuw, E.D. (2012) 'Counting and measuring online: The quality of internet surveys', *Bulletin de Méthodologie Sociologique*, 114 (1): 68–78. doi: 10.1177/0759106312437290

Epstein, L. (2013) 'Get better data and simplify analysis: Qualitative vs. quantitative questions. Retrieved 10 August 2015, from: www.surveymonkey.com/blog/en/blog/2013/04/10/qualitative-vs-quantitative

Eysenbach, G. and Till, J.E. (2001) 'Information in practice: Ethical issues in qualitative research on internet communities', *British Medical Journal*, 323: 1103–5. doi: http://dx.doi.org.proxy.library.uu.nl/10.1136/ bmj.323.7321.1103

Eysenbach, J. and Wyatt, J. (2002) 'Using the internet for surveys and health research', *Journal of Medical Internet Research*, 4 (2). doi: 10.2196/jmir.4.2.e13

Fuchs, M. and Funke, F. (2007) 'Multimedia Web surveys: Results from a field experiment on the use of audio and video clips in Web surveys', in Trotman M., Burrell, T., Gerrard, L., Anderton, K., Basi, G., Couper, M., Morris, K., Birks, D., Johnson, A.J., Baker, R., Rigg, M., Taylor, S. and Westlake, A. (eds), *The Challenges of a Changing World. Proceedings of the Fifth International Conference of the Association for Survey Computing*. Berkeley: ASC, pp. 63–80.

Groves, R.M. and Lyberg, L. (2010) 'Total survey error: Past, present, and future,' *Public Opinion Quarterly*, 74 (5): 849–79. doi: 10.1093/poq/nfq065

Groves, R.M., Fowler, F.J., Jr, Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2009) *Survey Methodology.* Hoboken, NJ: John Wiley and Sons, Inc.

Internet World Stats (2014) Retrieved from: www.internetworldstats.com/stats.htm (accessed 14 April 2015).

Kreuter, F., Presser, S. and Tourangeau, R. (2008) 'Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity', *Public Opinion Quarterly*, 72 (5): 847–65. doi: 10.1093/poq/nfn063

Mohorko, A., de Leeuw, E. and Hox, J. (2013) 'Internet coverage and coverage bias in Europe: Developments across countries and over time', *Journal of Official Statistics*, 29 (4): 609–22. doi: 10.2478/jos-2013-0042

Neuman, W.L. (2012) *Understanding Research*. Boston, MA: Pearson Education, Inc.

Rookey, B.D., Hanway, S. and Dillman, D.A. (2008) 'Does a probability-based household panel benefit from assignment to postal response as an alternative to internet-only?' *Public Opinion Quarterly*, 72 (5): 962–84. doi: 10.1093/poq/nfn061

Schonlau, M., Fricker, R.D., Jr and Elliott, M.N. (2002) *Conducting Research Surveys via E-mail and the Web*. Santa Monica, CA: RAND.

Wright, K.B. (2005) 'Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services', *Journal of Computer-Mediated Communication*, 10 (3). doi: 10.1111/j.1083-6101.2005.tb00259.x

*** Items in bold are suggested readings.**