CHAPTER ONE

The Road to Web Surveys

1.1 Introduction

Web surveys are a next step in the evolution of survey data collection. Collecting data for compiling statistical overviews is already very old, almost as old as humankind. All through history statistics have been used by rulers of countries to take informed decisions. However, new developments in society always have had their impact on the way the data were collected for these statistics.

For a long period, until the year 1895, statistical data collection was based on complete enumeration of populations. The censuses were mostly conducted to establish the size of the population, to determine tax obligations of the people, and to measure the military strength of the country. The idea of sampling had not emerged yet.

The year 1895 marks a fundamental change. Populations had grown bigger and bigger. It was the period of industrialization. Centralized governments required more and more information. The time was ripe for sample surveys. The first ideas emerged around 1895. A lot of discussion took place between 1895 and 1934 about how samples should be selected: by means of probability sampling or some other sample selection technique.

By 1934, it was clear that only surveys based on probability sampling could provide reliable and accurate estimates. Such surveys were accepted as a scientific method of data collection. In the period from the 1940s to the 1970s, most sample surveys were based on probability sampling. Questionnaires were printed on paper forms. They were completed in face-to-face, telephone, or mail surveys.

Somewhere in the 1970s another significant development began. The fast development of microcomputers made it possible to introduce computer-assisted

Handbook of Web Surveys, First Edition. Jelke Bethlehem and Silvia Biffignandi.

interviewing. This made survey data collection faster, cheaper, and easier, and it increased data quality. It was a time when acronyms like CATI (computer-assisted telephone interviewing) and CAPI (computer-assisted personal interviewing) emerged.

The next major development was the creation of the Internet around 1982. When more and more persons and companies received access to the Internet, it became possible to use this network for survey data collection. The first Internet surveys were e-mail surveys. In 1989, the World Wide Web was introduced. This software allowed for much friendlier graphical user interfaces for Internet users. The first browsers emerged, and the use of the Internet exploded. In the middle of 1990s, the World Wide Web became widely available and e-mail surveys were increasingly replaced by web surveys.

Web surveys are attractive because they have several advantages. They allow for simple, fast, and cheap access to large groups of potential respondents. Not surprisingly, the number of conducted web surveys has increased rapidly over time. There are, however, also potential methodological problems. Ample examples of web surveys are not based on probability sampling. Therefore, generalization of survey results to the population is questionable.

This chapter describes the historical developments that have led to the emergence of web surveys. As an illustration, Section 1.3 shows how these developments were implemented at Statistics Netherlands and led to new software for survey data collection.

1.2 Theory

1.2.1 THE EVERLASTING DEMAND FOR STATISTICAL INFORMATION

The history of data collection for statistics goes back in time thousands of years. As far back as Babylonian times, a census of agriculture was carried out. This already took place shortly after the art of writing was invented. The same thing happened in China. This empire counted its people to determine the revenues and the military strength of its provinces. There are also accounts of statistical overviews compiled by Egyptian rulers long before Christ. Rome regularly took censuses of people and of property. The collected data were used to establish the political status of citizens and to assess their military and tax obligations to the state.

Censuses were rare in the Middle Ages. The most famous one was the census of England taken by the order of William the Conqueror, King of England. The compilation of his *Domesday Book* started in the year 1086 AD. The book records a wealth of information about each manor and each village in the country. Information was collected about more than 13,000 places. More than 10,000 facts were recorded for each county.

To collect all this data, the country was divided into several regions. In each region, a group of commissioners was appointed from among the greater lords. Each county within a region was dealt with separately. Sessions were organized in

each county town. The commissioners summoned all those required to appear before them. They had prepared a standard list of questions. For example, there were questions about the owner of the manor; the number of free men and slaves; the area of woodland, pasture, and meadow; the number of mills and fishponds, to the total value; and the prospects of getting more profit. The *Domesday Book* still exists, and many county data files are available on CD-ROM and the Internet.

Another interesting example of the history of official statistics can be found in the Inca Empire that existed between 1000 and 1500 AD. Each Inca tribe had its own statistician, called the *Quipucamayoc*. This man kept records of the number of people, the number of houses, the number of llamas, the number of marriages, and the number of young men that could be recruited for the army. All these facts were recorded on *quipus*, a system of knots in colored ropes. A decimal system was used for this. At regular intervals, couriers brought the quipus to Cusco, the capital of the kingdom, where all regional statistics were compiled into national statistics. The system of Quipucamayocs and quipus worked remarkably well. The system vanished with the fall of the empire.

An early census also took place in Canada in 1666. Jean Talon, the intendant of New France, ordered an official census of the colony to measure the increase in population since the founding of Quebec in 1608. Name, age, sex, marital status, and occupation were recorded for every person. It turned out 3,215 people lived in New France.

The first censuses in Europe were conducted in the Nordic countries: The first census in Sweden-Finland took place in 1749. Not everyone welcomed the idea of a census. In particular, religious people believed that people should not be counted. They referred to the census ordered by King David in biblical times, which was interrupted by a terrible plague and never completed. Others said that a population count would reveal the strengths and weaknesses of a country to foreign enemies. Nevertheless, censuses were conducted in more and more countries. The first census in Denmark-Norway was done in 1769. In 1795, at the time of the Batavian Republic under Napoleon's influence, the first count of the population of the Netherlands took place. The new centralized administration wanted to gather quantitative information to devise a new system of electoral constituencies (see Den Dulk & Van Maarseveen, 1990).

In the period until the late 1880s, there were some applications of *partial investigations*. These were statistical inquiries in which only part of a complete human population was investigated. The way the persons were selected from the population was generally unclear and undocumented.

In the second half of the 19th century, so-called *monograph studies* became popular. They were based on Quetelet's idea of the average man. According to Quetelet, many physical and moral data have a natural variability. This variability can be described by a normal distribution around a fixed, true value. He assumed the existence of something called the *true value*. Quetelet introduced the concept of *average man* (*l'homme moyenne*) as a person of which all characteristics were equal to the true value, (see Quetelet, 1835, 1846).

The period of the 18th and 19th centuries is also called the era of the Industrial Revolution. It led to important changes in society, science, and

technology. Among many other things, urbanization started from industrialization and democratization. All these developments created new statistical demands. The foundations for many principles of modern social statistics were laid. Several central statistical bureaus, statistical societies, conferences, and journals were established soon after this period.

1.2.2 THE DAWN OF SAMPLING THEORY

The first ideas about survey sampling emerged in the world of official statistics. If a starting year must be chosen, 1895 would be a good candidate. Anders Kiaer, the founder and first director of Statistics Norway, started in this year a fundamental discussion about the use of sampling methods. This discussion led to the development, acceptance, and application of sampling as a scientific method.

Kiaer (1838–1919) was the founder and advocate of the survey method that is now widely applied in official statistics and social research. With the first publication of his ideas in 1895, he started the process that ended in the development of modern survey sampling theory and methods. This process is described in more detail by Bethlehem (2009).

It should be noted that earlier examples of scientific investigations have been based on samples, but they were lacking proper scientific foundations. The first known attempt of drawing conclusions about a population using only information about part of it was made by the English merchant John Graunt (1662). He estimated the size of the population of London. Graunt surveyed families in a sample of parishes where the registers were well kept. He found that on average there were three burials per year in 11 families. Assuming this ratio to be more or less constant for all parishes, and knowing the total number of burials per year in London to be about 13,000, he concluded that the total number of families was approximately 48,000. Putting the average family size at eight, he estimated the population of London to be 384,000. As this approach lacked a proper scientific foundation, John Graunt could not say how accurate his estimates were.

More than a century later, the French mathematician Pierre Simon Laplace realized that it was important to have some indication of the accuracy of his estimate of the French population. Laplace (1812) implemented an approach that was more or less similar to that of John Graunt. He selected 30 departments distributed over the area of France in such a way that all types of climate were represented. Moreover, he selected departments in which accurate population records were kept. Using the central limit theorem, Laplace proved that his estimator had a normal distribution. Unfortunately, he disregarded the fact that sampling was purposively and not at random. These problems made application of the central limit theorem at least doubtful.

In 1895, Kiaer (1895, 1997), the founder and first director of Statistics Norway, proposed his *representative method*. It was a partial inquiry in which a large number of persons was questioned. Selection of persons was such that a "miniature" of the population was obtained. Anders Kiaer stressed the importance of *representativity*. He argued that, if a sample was representative with respect to variables for which the population distribution was known, it would also be representative with respect to the other survey variables.

EXAMPLE 1.1 The representative method of Anders Kiaer

Anders Kiaer applied his representative method in Norway. His idea was to survey the population of Norway by selecting a sample of 120,000 people. Enumerators (hired only for this purpose) visited these people and filled in 120,000 forms. Approximately 80,000 of the forms were collected by the representative method and 40,000 forms by a special (but analogue) method in areas where the working class people lived.

For the first sample of 80,000 respondents, data from the 1891 census were used to divide the households in Norway into two strata. Approximately 20,000 people were selected from urban areas and the rest from rural areas.

Thirteen representative cities were selected from the 61 cities in Norway. All five cities having more than 20,000 inhabitants were included, as were eight cities representing the medium-sized and small towns. The proportion of selected people in cities varied: In the middle-sized and small cities, the proportion was greater than in the big cities. Kiaer motivated this choice by the fact that the middle-sized and small cities did not represent only themselves but a larger number of similar cities. In Kristiania (nowadays Oslo) the proportion was 1/16; in the medium sized towns, the proportion varied between 1/12 and 1/9; and in the small towns, it was 1/4 or 1/3 of the population.

Based on the census, it was known how many people lived in each of the 400 streets of Kristiania, the capital of Norway. The streets were sorted in four categories according to the number of inhabitants. A selection scheme was then specified for each category: The whole adult population was enumerated in 1 out of 20 for the smallest streets. In the second category, the adult population was enumerated in half of the houses in 1 out of 10 of streets. In the third category, the enumeration concerned 1/4 of the streets and every fifth house was enumerated; and in the last category of the biggest streets, the adult population was enumerated on half of the streets and in 1 out of 10 houses in them.

In selecting the streets, their distribution over the city was taken into account to ensure the largest possible dispersion and the "representative character" of the enumerated areas. In the medium-sized towns, the sample was selected using the same principles, although in a slightly simplified manner. In the smallest towns, the whole adult population in three or four houses was enumerated.

The number of informants in each of the 18 counties in the rural part of Norway was decided on the basis of census data. To obtain representativeness, municipalities in each county were classified according to their main industry, either as agricultural, forestry, industrial, seafaring, or fishing municipalities. In addition, the geographical distribution was taken into account. The total number of the representative municipalities

amounted to 109, which is 6 in each county on average. The total number of municipalities was 498.

The selection of people in a municipality was done in relation to the population in different parishes, and so that all different municipalities were covered. The final step was to instruct enumerators to follow a specific path. In addition, enumerators were instructed to visit different houses situated close to each other. That is, they were supposed to visit not only middle-class houses but also well-to-do houses, poor-looking houses, and one-person houses.

Kiaer did not explain in his papers how he calculated estimates. The main reason probably was that the representative sample was constructed as a miniature of the population. This made computations of estimates trivial: The sample mean is the estimate of the population mean, and the estimate of the population total could be attained simply by multiplying the sample total by the inverse of sampling fraction.

A basic problem of the representative method was that there was no way of establishing the precision of population estimates. The method lacked a formal theory of inference. It was Bowley (1906, 1926) who made the first steps in this direction. He showed that for large samples, selected at random from the population, estimates had an approximately normal distribution. From this moment on, there were two methods of sample selection:

- Kiaer's representative method, based on purposive selection, in which representativity played an essential role, and for which no measure of the accuracy of the estimates could be obtained;
- Bowley's approach, based on simple random sampling, and for which an indication of the accuracy of estimates could be computed.

Both methods existed side by side until 1934. In that year, the Polish scientist Jerzy Neyman published his famous paper (1934). Neyman developed a new theory based on the concept of the confidence interval. By using random selection instead of purposive selection, there was no need any more to make prior assumptions about the population. The contribution of Neyman was not only that he proposed the confidence interval as an indicator for the precision of estimates. He also conducted an empirical evaluation of Italian census data and proved that the representative method based on purposive sampling could not provide satisfactory estimates of population characteristics. He established the superiority of random sampling (also referred to as *probability sampling*) over purposive sampling. Consequently, use of purposive sampling was rejected as a scientific sampling method.

Gradually probability sampling found its way into official statistics. More and more national statistical institutes introduced probability sampling for

official statistics. However, the process was slow. For example, a first test of a real sample survey using random selection was carried out by Statistics Netherlands only in 1941 (CBS, 1948). Using a simple random sample of size 30,000 from the population of 1.75 million taxpayers, it was shown that estimates were accurate.

The history of opinion polls goes back to the 1820s, in which period American newspapers attempted to determine the political preference of voters just before the presidential election. These early polls did not pay much attention to sampling. Therefore, it was difficult to establish the accuracy of the results. Such opinion polls were often called *straw polls*. This expression goes back to rural America. Farmers would throw a handful of straws into the air to see which way the wind was blowing.

It took until the 1920s before more attention was paid to sampling aspects. Lienhard (2003) describes how George Gallup worked out new ways to measure interest in newspaper articles. Gallup used *quota sampling*. The idea was to investigate a group of people that could be considered representative for the population. Hundreds of interviewers across the country visited people. Interviewers were given a quota for different groups of respondents. They had to interview so many middle-class urban women, so many low-class rural men, and so on. In total, approximately 3,000 interviews were conducted for a survey.

Gallup's approach was in great contrast with that of *Literary Digest* magazine, which was at that time the leading polling organization. This magazine conducted regular "America Speaks" polls. It based its predictions on returned questionnaire forms that were sent to addresses taken from telephone directory books and automobile registration lists. The sample size for these polls was in the order of two million people. So the sample size was much larger than that of Gallup's polls.

The presidential election of 1936 turned out to be decisive for both methods. This is described by Utts (1999). Gallup correctly predicted Franklin D. Roosevelt to be the new president, whereas *Literary Digest* predicted that Alf Landon would beat Franklin D. Roosevelt. The prediction based on the very large sample size turned out to be wrong. The explanation was that the sampling technique of *Literary Digest* did not produce representative samples. In the 1930s, cars and telephones were typically owned by middle-and upper class people. These people tended to vote Republican, whereas lower class people were more inclined to vote Democrat. Consequently, Republicans were overrepresented in the *Literary Digest* sample.

As a result of this historic mistake, opinion researchers learned that they should rely on more scientific ways of sample selection. They also learned that the way a sample is selected is more important than the size of the sample.

The classic theory of survey sampling was more or less completed in 1952. Horvitz and Thompson (1952) developed a general theory for constructing unbiased estimates. Whatever the selection probabilities are, as long as they are known and positive, it is always possible to construct a useful estimate. Horvitz and Thompson completed this classic theory, and the random sampling approach was almost unanimously accepted. Most classic books about sampling also were published by then (Cochran, 1953; Deming, 1950; Hansen, Hurwitz, & Madow, 1953; Yates, 1949).

1.2.3 TRADITIONAL DATA COLLECTION

There were three modes of data collection in the early days of survey research: face-to-face interviewing, mail interviewing, and telephone interviewing. Each mode had its advantages and disadvantages.

Face-to-face interviewing was already used for the first censuses. So, it is not a surprise it was also used for surveys. Face-to-face interviewing means that interviewers visit the persons selected in the sample. Well-trained interviewers will be successful in persuading reluctant persons to participate in the survey. Therefore, response rates of face-to-face surveys are usually higher than surveys not involving interviewers (for example, mail surveys). Interviewers can also assist respondents in giving the right answers to the questions. This often results in better data. However, the presence of interviewers can also be a drawback. Research suggests that respondents are more inclined to answer sensitive questions properly if no interviewers are present.

Survey agencies often send a letter announcing the visit of the interviewer. Such a letter also can give additional information about the survey, explain why it is important to participate, and assure that the collected information is treated confidentially. As a result, the respondents are not taken by surprise by the interviewers.

The response rate of a face-to-face survey is usually high and so is the quality of the collected data. But a price has to be paid literally: Face-to-face interviewing is much more expensive. A team of interviewers has to be trained and paid. They also have to travel, which costs time and money.

Mail interviewing is much less expensive than face-to-face interviewing. Paper questionnaires are sent by mail to persons selected in the sample. They are invited to answer the questions and to return the completed questionnaire to the survey agency. A mail surveys does not involve interviewers. Therefore, it is a cheap mode of data collection. Data collection costs only involve mailing costs (letters, postage, and envelopes). Another advantage is that the absence of interviewers can be experienced as less threatening for potential respondents. As a consequence, respondents are more inclined to answer sensitive questions properly.

The absence of interviewers also has several disadvantages. There are no interviewers to explain questions or to assist respondents in answering them. This may cause respondents to misinterpret questions, which has a negative impact on the quality of the collected data. Also, it is not possible to use show cards. A *show card* is typically used for answering closed questions. Such a card contains the list of all possible answers to a question. Respondents can read through the list at their own pace and select the answer corresponding to their situation or opinion. Mail surveys put high demands on the design of the paper questionnaire. For example, it should be clear to all respondents how to navigate through the questionnaire and how to answer questions.

As the persuasive power of the interviewers is absent, the response rates of mail surveys tend to be low. Of course, reminder letters can be sent, but this is often not very successful. More often survey questionnaire forms end up in the pile of old newspapers.

In summary, the costs of a mail survey are relatively low, but often a price has to be paid in terms of data quality: Response rates tend to be low and the quality

of the collected data is also often not very good. Dillman (2007) believes, however, that good results can be obtained by applying his tailored design method. This set of guidelines is used for designing and formatting mail survey questionnaires. It pays attention to all aspects of the survey process that may affect response rates or data quality.

Face-to-face interviewing was preferred in the early days of survey interviewing in the Netherlands. The idea was in the 1940s that poor people had poor writing skills, and moreover, they were not interested in the topics of the surveys. Therefore, they had a smaller probability of completing mail questionnaires. People completing and returning questionnaire forms were assumed to be more interested in the survey topics because their intelligence and social-economic position was above average.

A third mode of data collection is *telephone interviewing*. Interviewers are needed to conduct a telephone survey, but not as many as for a face-to-face survey because they do not have to travel from one respondent to the next. They can remain in the call center of the survey agency and can conduct more interviews in the same amount of time. Therefore, interviewer costs are less. An advantage of telephone interviewing over face-to-face interviewing is that respondents may be more inclined to answer sensitive questions because the interviewer is not present in the room. A drawback in the early days of telephone surveys was that telephone coverage in the population was small. Not every respondent could be contacted by telephone.

EXAMPLE 1.2 The first telephone survey in the Netherlands

The first telephone survey was conducted in the Netherlands on June 11, 1946. See NIPO (1946) for a detailed description. A few hundred owners of telephones in Amsterdam were asked to answer a few questions about listening to the radio. The people were called between 20:00 and 21:30 hours on a Tuesday night. Some results are given in Table 1.1.

TABLE 1.1 The first telephone survey in the Netherlands

Are you listening to the radio at this moment?	Percentage
Was listening	24%
Was not listening	38%
Line busy	5%
No answer	31%
Did not have a radio	2%

People listening to the radio also were asked which program they were listening to. It turned out that 85% was listening the "Bonte Dinsdagavondtrein," a very famous radio show at that time.

Telephone interviewing has some limitations. Interviews cannot last too long, and questions may not be too complicated. Another problem may be the lack of a proper sampling frame. Telephone directories may suffer from severe undercoverage because many people do not want their phone number to be listed in the directory. Another, new, development is that increasingly people replace their landline phone by a mobile phone. Mobile phone numbers are not listed in directories in many countries. For example, according to Fannic Cobben and Jelke G. Bethlehem (2005), only between 60% and 70% of the Dutch population can be reached through a telephone dictionary. For more information about the use of mobile phones for interviewing, see Kuusela, Vehovar and Callegaro (2006).

A way to avoid the undercoverage problems of telephone directories is to apply *random digit dialing* (RDD) to generate random phone numbers. A computer algorithm computes valid random telephone numbers. Such an algorithm can generate both listed and unlisted numbers. So, there is complete coverage. An example of an algorithm used in the United Kingdom is to take a number from a directory and replace its last digit by a random digit. Random digit dialing also has drawbacks. In some countries, it is not clear what an unanswered numbers means. It can mean that the number is not in use. This is a case of overcoverage. No follow-up is needed. It also can mean that someone simply does not answer the phone, which is a case of nonresponse, that has to be followed up. Another drawback of RDD is that there is no information at all about nonrespondents. This makes correction for nonresponse very difficult (see also Chapter 10 about weighting adjustments).

The choice of the mode of data collection is not any easy one. It is usually a compromise between quality and costs. In large countries (like the United States) or sparsely populated countries (like Sweden), it is almost impossible to collect survey data by means of face-to-face interviewing. It requires so many interviewers that have to do so much traveling that the costs would be very high. Therefore, it is not surprising that telephone interviewing emerged here as a major data collection mode. In a very small and densely populated country like the Netherlands, face-to-face interviewing is much more attractive. The coverage problems of telephone directories and the low response rates also play a role in the choice for face-to-face interviewing. More about data collection issues can be found in the study by Couper et al. (1998).

1.2.4 THE ERA OF COMPUTER-ASSISTED INTERVIEWING

Collecting survey data can be a costly and time-consuming process, particularly if high-quality data are required, the sample is large, and the questionnaire is long and complex. Another problem of traditional data collection is that the completed paper questionnaire forms may contain many errors. Substantial resources must therefore be devoted to cleaning the data. Extensive data editing is required to obtain data of acceptable quality.

Rapid developments in information technology since the 1970s have made it possible to reduce these problems. This was accomplished by introducing microcomputers for data collection. The paper questionnaire was replaced by a

computer program asking the questions. The computer took control of the interviewing process, and it checked answers to the questions. Thus, *computer-assisted interviewing* (CAI) emerged.

Computer-assisted interviewing comes in different modes of data collection. The first mode of data collection that emerged was *computer assisted telephone interviewing* (CATI). Couper and Nicholls (1998) describe its development in the United States in the early 1970s. The first nationwide telephone facility for surveys was established in 1966. The idea at that time was not implementation of computer-assisted interviewing but simplification of sample management. The initial systems evolved in subsequent years into full featured CATI systems. Particularly in the United States, there was a rapid growth of the use of these systems. CATI systems were little used in Europe until the early 1980s.

Interviewers in a CATI survey operate a computer running interview software. When instructed to do so by the software, they attempt to contact a selected person by telephone. If this is successful and the person is willing to participate in the survey, the interviewer starts the interviewing program. The first question appears on the screen. If this is answered correctly, the software proceeds to the next question on the route through the questionnaire.

Call management is an important component of many CATI systems. Its main function is to offer the right telephone number at the right moment to the right interviewer. This is particularly important in cases in which the interviewer has made an appointment with a respondent for a specific time and date. Such a call management system also has facilities to deal with special situations like a busy number (try again after a short while) or no answer (try again later). This all helps to increase the response rate. More about the use of CATI in the United States can be found in the study by Nicholls and Groves (1986).

Small portable computers came on the market in the 1980s. This made it possible for the interviewers to take computers with them to the respondents. This is the computer-assisted form of face-to-face interviewing. It is called *computer-assisted personal interviewing* (CAPI). After interviewers have obtained cooperation from the respondents, they start the interviewing program. Questions are displayed one at a time. Only after the answer has been entered, will the next question appear on the screen.

At first, it was not completely clear whether computers could be used for this mode of data collection. There were issues like the weight and size of the computer, the readability of the screen, battery capacity, and the size of keys on the keyboard. Experiments showed that CAPI was feasible. It became clear that computer-assisted interviewing for data collection has three major advantages:

- It simplifies the work of interviewers. They do not have to pay attention any more to choosing the correct route through the questionnaire. The computer determines the next question to ask. Interviewers can concentrate more on asking questions and on helping respondents give the proper answers.
- It improves the quality of the collected data. Answers can be checked by the software during the interview. Detected errors can be corrected immediately. The respondent is there to provide the proper information. This is much

more effective than having to do data editing afterward in the survey agency and without the respondent

- Data are entered into the computer immediately during the interview. Checks are carried out straightaway, and detected errors are corrected. Therefore, the record of a respondent is "clean" after completion of the interview. No more subsequent data entry and/or data editing is required. Compared with the old days of traditional data collection with paper forms, this considerably reduces the time needed to process the survey data. Therefore, the timeliness of the survey results is improved.

More information about CAPI in general can be found in the study by Couper et al. (1998).

The computer-assisted mode of mail interviewing also emerged. It was called *computer-assisted self-interviewing* (CASI) or sometimes also *computer assisted self-administered questionnaires* (CSAQ). The electronic questionnaire program is sent to the respondents. They run the software, which asks the questions and stores the answers. After the interview has been completed, the data are sent back to the survey agency. Early CASI applications used diskettes or a telephone and modem to transmit the questionnaire and the answers to the question. Later it became common practice to use the Internet as a transport medium.

A CASI survey is only feasible if all respondents have a computer on which they can run the interview program. As the use of computers was more widespread among companies than households in the early days of CASI, the first CASI applications were business surveys. An example is the production of Fire Statistics in the Netherlands in the 1980s. Because all fire brigades had a microcomputer at that time, data for these statistics could be collected by means of CASI. Diskettes were sent to the fire brigades. They ran the questionnaire on their MS-DOS computers. The answers were stored on the diskette. After having completed the questionnaire, the diskette was returned to Statistics Netherlands.

An early application in social surveys was the *Telepanel*, which was set up by Saris (1998). The Telepanel started in 1986. It was a panel of 2,000 households that agreed to complete questionnaires regularly with the computer equipment provided to them by the survey organization. A home computer was installed in each household. It was connected to the telephone with a modem. It also was connected to the television set in the household so that it could be used as a monitor. After a diskette was inserted into the home computer, it automatically established a connection with the survey agency to exchange information (downloading a new questionnaire or uploading answers of the current questionnaires). Panel members completed a questionnaire each weekend. The Telepanel was in essence very similar to the web panels that are frequently used nowadays. The only difference was the Internet did not exist yet.

1.2.5 THE CONQUEST OF THE WEB

The development of the Internet started in the early 1970s. The first step was to create networks of computers. The U.S. Department of Defense decided to connect computers across research institutes. Computers were expensive. A

network made it possible for these institutes to share each other's computer resources. This first network was called ARPANET.

ARPANET became a public network in 1972. Software was developed to send messages over the network. Thus, e-mail was born. The first e-mail was sent in 1971 by Ray Tomlinson of ARPANET.

The Internet was fairly chaotic in the first decade of its existence. There were many competing techniques and protocols. In 1982, the TCP/IP set of protocols was adopted as the standard for communication of connected networks. This can be seen as the real start of the Internet.

Tim Berners-Lee and scientists at CERN, the European Organization for Nuclear Research in Geneva, were interested in making it easier to retrieve research documentation over the Internet. This led in 1989 to the *hypertext* concept. This is text containing references (hyperlinks) to other texts the reader can immediately access. To be able to view these text pages and navigate to other pages through the hyperlinks, Berners-Lee developed computer software. He called this program a *browser*. This first browser was named the *World Wide Web*. This name is now used to denote the whole set of linked hypertext documents on the Internet.

In 1993, Mark Andreesen and his team at the National Center for Supercomputing Applications (NCSA, IL) developed the browser *Mosaic X*. It was easy to install and use. This browser had increased graphic capabilities. It already contained many features that are common in current browsers. It became a popular browser, which helped to spread the use of the World Wide Web across the world.

The rapid development of the Internet led to new modes of data collection. Already in the 1980s, prior to the widespread introduction of the World Wide Web, e-mail was explored as a new mode of survey data collection. Kiesler and Sproull (1986) described an early experiment conducted in 1983. They compared an e-mail survey with a traditional mail survey. They showed that the costs of an e-mail survey were much less than those of a mail survey. The response rate of the e-mail survey was 67%, and this was somewhat smaller than the response rate of the mail survey (75%). The turnaround time of the e-mail survey was much shorter. There were less socially desirable answers and less incomplete answers. Kiesler and Sproull (1986) noted that limited Internet coverage restricted wide-scale use of e-mail surveys. In their view, this type of data collection was only useful for communities and organizations with access to, and familiarity with, computers. These were relatively well-educated, urban, white collar, and technologically sophisticated people.

Schaefer and Dillman (1998) also compared an e-mail surveys with mail surveys. They applied knowledge about mail surveys to e-mail surveys and developed an e-mail survey methodology. They also proposed mixed-mode surveys for populations with limited Internet coverage. They pointed out some advantages of e-mail surveys. In the first place, e-mail surveys could be conducted very fast, even faster than telephone surveys. This was particularly the case for large surveys, where the number of available telephones and interviewers may limit the number of cases that can be completed each day. In the second place, e-mail surveys were inexpensive because there were no mailing, printing, and interviewers costs.

The experiment of Schaefer and Dillman (1998) showed that the response rates of e-mail and mail surveys were comparable, but the completed questionnaires of the e-mail survey were received much quicker. The answers to open questions were, on average, longer for e-mail surveys. This did not come as a surprise because of the relative ease of typing an answer on a computer compared with writing an answer on paper. There was lower item nonresponse for the e-mail survey. A possible explanation was that moving to a different question in an e-mail survey is much more difficult than moving to a different question on a paper form.

Couper, Blair, and Triplett (1999) found lower response rates for e-mail surveys in an experiment with a survey among employees of statistical agencies in the United States. They pointed out that nonresponse can partly be explained by delivery problems of the e-mails and not by refusal to participate in the survey. For example, if people do not check their e-mail or if the e-mail with the questionnaire does not pass a spam filter, people will not be aware of the invitation to participate in a survey.

Most e-mail surveys could not be considered a form of computer-assisted interviewing. It was merely the electronic analogue of a paper form. There was no automatic routing and no error checking. See Figure 1.1 for a simple example of an e-mail survey questionnaire. It is sent to the respondents. They are asked to reply to the original message. Then they answer the questions in the questionnaire in the reply message. For closed questions, they do that by typing an "X" between the brackets of the option of their choice. The answer to an open question is typed between the corresponding brackets. After completion, they send the e-mail message to the survey agency.

Use of e-mail imposes substantial restrictions on the layout. Because of the e-mail software of the respondent and the settings of the software, the

1. What is your age?
[ ]
2. Are you male or female?
[ ] Male
[ ] Female
3. What is your marital status?
[ ] Married
[ ] Not married
4. Do you have a job?
[ ] Yes
[ ] No
5. What kind of job do you have?
[ ]
6. What is your yearly income?
[ ] Less than $20,000
[ ] Between $20,000 and $40,000
[ ] More than $40,000

FIGURE 1.1 Example of an e-mail survey questionnaire

questionnaire may look different to different respondents. For example, to avoid problems caused by line wrapping, Schaefer and Dillman (1998) advise a line length of at most 70 characters.

Schaefer and Dillman (1998) also noted another potential problem of e-mail surveys: the lack of anonymity of e-mail. If respondents reply to the e-mail with the questionnaire, it is difficult to remove all identifying information. Some companies have the possibility to monitor the e-mails of their employees. If this is the case, it may become difficult to obtain high response rates and true answers to the questions asked.

Personalization may helps to increase response rates in mail surveys. Therefore, this principle should also be applied to e-mail surveys. An e-mail to a long list of addresses does not help to create the impression of personal treatment. It is probably better to send a separate e-mail to each selected person individually.

EXAMPLE 1.3 The first e-mail survey at Statistics Netherlands

The first test with an e-mail survey at Statistics Netherlands was carried out in 1998. At the time, Internet browsers and HTML where not sufficiently developed and used to make a web survey feasible.

The objective of the test was to explore to what extent e-mail could be used to collect data for the Survey on Short Term Indicators. This was a noncompulsory panel survey, where companies answered a small number of questions about production expectations, order-books, and stocks.

The traditional mode of data collection for this survey was a mail survey.

The test was conducted in one wave of the survey. A total of 1,600 companies were asked to participate in the test. If they did, they had to provide their e-mail address. Approximately 190 companies agreed to participate. These were mainly larger companies with a well-developed computer infrastructure.

A simple text form was sent to these companies by means of e-mail. After activating the reply-option, respondents could fill in answers in the text. It was a software-independent and platform-independent solution, but it was primitive from a respondent's point of view.

The test was a success. The response rate among the participating companies was almost 90%. No technical problems were encountered. Overall, respondents were positive. However, they considered the text-based questionnaire old-fashioned and not very user-friendly.

More details about this first test with an e-mail survey at Statistics Netherlands can be found in the study by Roos, Jaspers, and Snijkers (1999).

It should be noted that e-mail also can be used in a different way to send a questionnaire to a respondent. An electronic questionnaire can be offered as an executable file that is attached to the e-mail. The respondents download this interview program on their computers and run it. The advantage of this approach is that such a computer program can have a much better graphical user interface. Such a program also can include routing instructions and checks. This way of data collection is sometimes called CASI.

EXAMPLE 1.4 The production statistics pilot at Statistics Netherlands

In October 2004, Statistics Netherlands started a pilot to find out whether a CASI approach could be used to collect data for yearly production statistics.

One of the approaches tested is denoted by electronic data reporting (EDR). It was a system for responding companies to manage interviewing programs (generated by the Blaise System) on their own computers. The EDR software could be sent to respondents on CD-ROM, or respondents could download the software from the Internet. After the software had been installed, new survey interviews could be sent to respondents by e-mail. These electronic questionnaires were automatically imported in the EDR environment. A simple click would start the interview. After off-line completion of the interview, the entered data were automatically encrypted and sent to Statistics Netherlands.

The pilot made clear that downloading the software was feasible. It should be preferred over sending a CD-ROM because it was simpler to manage and less expensive. Some companies experiences problems with downloading and installing the software because security settings of their computer systems and networks prevented them of doing so. User-friendliness and ease of navigation turned out to be important issues for respondents.

For more information about this pilot, see Snijkers, Tonglet, and Onat (2004, 2005).

This form of CASI also has disadvantages. It requires respondents to have computer skills. They should be able to download and run the interviewing program. Couper et al. (1999) also note that problems may be caused by that fact that different users may have different operating systems on their computers or different versions of the same operating system. This may require different versions of the interviewing program. And it must be known in advance which operating system a respondent has. Moreover, the size of an executable file may be substantial, which may complicate sending it by e-mail.

E-mail surveys had the advantages of speed and low costs. Compared with computer-assisted interviewing, it had the disadvantages of a poor user interface

and lack of adequate editing and navigation facilities. An e-mail questionnaire was just a paper questionnaire in an e-mail. The Internet became more interesting for survey data collection after HTML 2.0 was introduced in 1995. HTML stands for HyperText Markup Language. It is the markup language for web pages. The first version of HTML was developed by Tim Berners-Lee in 1991. Version 2 of HTML included support for forms. This made it possible to transfer data from a user to the web server. Web pages could contain questions, and the answers could be collected by the server.

EXAMPLE 1.5 Designing questions in HTML 2.0

Version 2.0 of HTML made it possible to implement questions on a web page. The `<input>` tag can be used to define different types of questions. With `type=radio`, this tag becomes a *radio button*. A *closed question* is defined by introducing a radio button for each possible answer. See Figure 1.2 for an example. Not more than one radio button can be selected. This corresponds to a closed question for which only one answer must be selected.

FIGURE 1.2 A closed question in HTML

Sometimes respondents must be offered the possibility to select more than one answer, like in Figure 1.3. Respondents are asked for their means of transport to work. Some people may use several transport means. For

FIGURE 1.3 A check-all-that-apply question in HTML

example, a person may first take a bicycle to the railway station, and then continues by train. Such a closed question is sometimes also called a *check-all-that-apply* question. It can be implemented in HTML by means of a series of *checkboxes*. A checkbox is obtained by stetting the type of the `<input>` tag to `checkbox`.

Figure 1.4 shows the implementation of an open question. Any text can be entered in the input field. A limit may be set to the length of the text. An open question is defined with `type=text` for the `<input>` tag.

FIGURE 1.4 An open question in HTML

If an input field is preferred that allows for more lines of text to be answered, the `<textarea>` tag can be used for this.

There are no specific types of the `<input>` tag for other types of questions. However, most of these question types can be implemented with the input field of an open question. For example, Figure 1.5 shows an numeric question. The question is basically an open question, but extra checks on the answer only allow numbers to be entered within certain bounds.

FIGURE 1.5 A numeric question in HTML

Date questions can be specified as a set of three input fields: one for the day, one for the month, and one for the year.

In the first years of the World Wide Web, use of web surveys was limited by the low penetration of the Internet. Internet penetration was higher among establishments than among households. Therefore, it is not surprising that first experiments tested the use of web business surveys. Clayton and Werking (1998)

describe a pilot carried out in 1996 for the Current Employment Statistics (CES) program of the U.S. Bureau of Labor Statistics. They expected the World Wide Web to offer a low-cost survey environment. Because it was a form of true on-line data collection, an immediate response to the answers of the respondents was possible. This could improve data quality. They also saw the great flexibility of web survey questionnaires. They could be offered in a form layout or in a question-by-question approach. The drawback was the limited number of respondents having access to the Internet. Only 11% of CES respondents had access to Internet and a compatible browser.

Roos and Wings (2000) conducted a test with Internet data collection at Statistics Netherlands for the construction industry. Respondents could choose among three modes:

- Completing a form off-line. The form was sent as an HTML-file that was attached to an e-mail. The form is downloaded, completed off-line, and returned by e-mail.
- Completing a form on-line. The Internet address of an on-line web form was sent by e-mail. The form was completed online.
- Completing an e-mail form. An e-mail is sent containing the questionnaire in plain text. Respondents clicked the reply button, answered the questions, and sent the e-mail back.

A sample of 1,500 companies was invited to participate in the experiment. Overall, 188 companies were willing and able to participate. Of those, 149 could surf the Internet and 39 only had e-mail. Questionnaire completion times of all three modes were similar to that of a paper form. Respondents preferred the form-based layout over the question-by-question layout. The conclusion of the experiment was that web surveys worked well.

General-population web surveys were rare in the first period of existence of the Internet. This was attributed to the low Internet penetration among households. This prevented conducting representative surveys. However, there were polls on the Internet. Recruitment of respondents was based on self-selection and not on probability sampling. Users could even create their own polls on websites like *Survey Central*, *Open Debate* and *Internet Voice*; see O'Connell (1998).

Also in 1998, the *Survey 2000* project was carried out. This was a large self-selection web survey on the website of the National Geographic Society. This was a survey on mobility, community, and cultural identity. In a period of two months over 80,000 respondents completed the questionnaire. See Witte, Amoroso, and Howard (2000), for more details about this project.

It seems to be typical for this type of self-selection web surveys that they make it possible to collect data about a large number of respondents in a relatively short time. Other examples are given by Bethlehem and Stoop (2007). The survey *21minuten.nl* has been conducted several times in the Netherlands. This survey was supposed to supply answers to questions about important

problems in Dutch society. Within a period of 6 weeks in 2006, approximately 170,000 people completed the online questionnaires. A similar survey was conducted in Germany. It is called *Perspektive Deutschland*. More than 600,000 participated in this survey in 2005/2006.

It should be noted that these large sample sizes are no guarantee for proper statistical inference. Because of undercoverage (not everyone has access to the Internet) and self-selection (no proper random sampling) estimates can be biased. This bias is independent of the sample size.

Internet penetration is still low in many countries, making it almost impossible to conduct a general population web survey. Because data collection costs can be reduced if the Internet is used, other approaches are sought. One such approach is *mixed-mode data collection*. A web survey is combined with one or more other modes of data collection, like a mail survey, a telephone survey, or a face-to-face survey. Researchers first attempt to collect as much data as possible with the cheapest mode of data collection (web). Then, the nonrespondents are reapproached in a different (next cheapest) mode, and so on.

EXAMPLE 1.6 An experiment with a mixed-mode survey

Beukenhorst and Wetzels (2009) describe a mixed-mode experiment conducted by Statistics Netherlands. They used the Dutch Safety Monitor for this experiment. This survey asks questions about feelings of security, quality of life, and level of crime experienced. The sample for this survey was selected from the Dutch population register. All sample persons received a letter in which they were asked to complete the survey questionnaire on the Internet. The letter also included a postcard that could be used to request a paper questionnaire. Two reminders were sent to those that did not respond by web or mail. If still no response was obtained, nonrespondents were approached by means of CATI, if a listed telephone number was available. If not, these nonrespondents were approached by CAPI.

To be able to compare this four-mode survey with a traditional survey, also a two-mode survey was conducted for an independent sample. Sampled persons were approached by CATI if their telephone number was listed in the directory, and otherwise, they were approached by CAPI.

The response rate for the four-mode survey turned out to be 59.7%. The response rate for the two-mode survey was 63.5%. So, introducing more modes did not increase the overall response rate. However, more than half of the response (58%) in the four-mode survey was obtained with a self-administered mode of data collection (web or paper). Therefore, the costs of the survey were much lower. Interviewers were deployed in only 42% of the cases. For more detail, see Beukenhorst and Wetzels (2009) or Bethlehem, Cobben, and Schouten (2011).

1.3 Application

The historic developments with respect to surveys as described in the previous section also took place in the Netherlands. In particular, the rapid developments in computer technology have had a major impact on the way Statistics Netherlands collected its data. Efforts to improve the process of collecting and processing survey data in terms of costs, timeliness, and quality have led to a powerful software system called Blaise. This system emerged in the 1980s, and it has evolved over time so that it is now also able to conduct web surveys and mixed-mode surveys. The section gives an overview of the developments at Statistics Netherlands leading to Internet version of Blaise.

All statistics published by Statistics Netherlands in the first half of the 20th century were based on a complete enumeration. Either data were collected by means of a population census or the data were obtained from a population register. One of the first real applications of sampling took place in 1947 with respect to income statistics. In 1946, a complete enumeration had been carried out. It meant processing data on 4 million tax administration cards. As the quality of the data on the cards was not very good, substantial manual editing was required. To reduce the magnitude of this immense operation, it was decided to use sampling methods for subsequent years.

Statistics Netherlands started using sample survey methods for agricultural statistics in the same period. Surveys were carried out from 1947 to estimate agricultural production. Samples were selected from a sampling frame consisting of a list of addresses of farms. These lists were compiled in the agricultural census (a complete enumeration) that was conducted every year in the month of May. A stratified sample was selected, where strata were formed based on province and size of farms. Within each stratum, systematic samples were selected. The total sample size was 10,000 to 20,000 farms. The surveys allowed for early estimates for the type and size of agricultural production.

Collecting and processing statistical data was a time-consuming and expensive process. Data editing was an important component of this work. The aim of these data editing activities was to detect and correct errors in the individual records, questionnaires, or forms. This should improve the quality of the results of surveys. Since statistical offices attached much importance to this aspect of the survey process, a large part of human and computer resources are spent on it.

To obtain more insight into the effectiveness of data editing, Statistics Netherlands carried out a Data Editing Research Project in 1984. Bethlehem (1987) describes how survey data were processed at that time. After all paper forms had been collected, subject-matter specialists checked them for completeness. If necessary and possible, skipped questions were answered, and obvious errors were corrected on the forms. Sometimes, the data on a form were manually copied to a new form to allow for the subsequent step of fast data entry. Next, the forms were transferred to the data entry department. Data typists entered the data in the computer at high speed without error checking. The computer was a dedicated data entry system. After data entry, the files were

transferred to the mainframe computer system. On the mainframe, an error detection program was run. Usually, this was a dedicated program written in the Cobol language. This program produced a list of detected errors. This list was sent to the subject-matter department. Specialists investigated the error messages, located and consulted corresponding forms, and corrected errors on the lists. Corrected forms were sent to the data entry department, and data typists entered the corrections in the data entry computer. The file with corrections was transferred to the mainframe computer. Corrected records and already present correct records were merged. The cycle of batch-wise error detection and manual correction was repeated until the number of detected errors was sufficiently small.

The Data Editing Research Project discovered several problems:

- Various people from different departments were involved. Many people dealt with the information: respondents, subject-matter specialists, data typists, and computer programmers. Transfer of material from one person/department to another could be a source of error, misunderstanding, and delay.
- Different computer systems were involved. Most data entry was carried out on Philips P7000 minicomputer systems, and data editing programs ran on a CDC Cyber 855 mainframe. Furthermore, there was a variety of desktop (running under MS-DOS) and other systems. Transfer of files from one system to another caused delay, and incorrect specification and documentation could produce errors.
- Not all activities were aimed at quality improvement. Time was also spent on just preparing forms for data entry and not on correcting errors.
- The process was going through cycles, from one department to another, and from one computer system to another. The cycle of data entry, automatic checking, and manual correction was in many cases repeated three times or more. Because of these cycles, data processing was very time consuming.
- The structure and nature of the data (the metadata) had to be specified in nearly every step of the data editing process. Although essentially the same, the "language" of this meta-data specification could be completely different for every department or computer system involved.

The conclusions of the Data Editing Research Project led to general redesign of the survey processes of Statistics Netherlands. The idea was to improve the handling of paper questionnaire forms by integrating data entry and data editing tasks. The traditional batch-oriented data editing activities, in which the complete data set was processed as a whole, was replaced by a record-oriented process in which records (forms) were completely dealt with one at a time.

The new group of activities was implemented in a so-called CADI system. CADI stands for *computer-assisted data input*. The CADI system was designed for use by the workers in the subject-matter departments. Data could be processed in two ways by this system:

- *Heads-up data entry*. Subject-matter employees worked through a pile of forms with a microcomputer, processing the forms one by one. First, they

entered all data on a form, and then they activated the check option to test for all kinds of errors. Detected errors were reported on the screen. Errors could be corrected by consulting forms or by contacting the suppliers of the information. After elimination of all errors, a "clean" record was written to file. If employees did not succeed in producing a clean record, they could write the record to a separate file of "dirty" records. A specialist could deal with these hard cases later on, also with a CADI system.

- *Heads-down data entry*. Data typists used the CADI system to enter data beforehand without much error checking. After completion, the CADI system checked in a batch run all records and flagged the incorrect ones. Then subject-matter specialists handled these "dirty" records one by one and corrected the detected errors.

To be able to introduce CADI on a wide scale in the organization, a new standard package was developed in 1986. The name of this standard package was *Blaise*. The basis of the Blaise System was the Blaise language, which was used to create a formal specification of the structure and contents of the questionnaire.

The first version of the Blaise System ran on microcomputers (or networks of microcomputers) under MS-DOS. It was intended for use by the people of the subject-matter departments; therefore, no computer expert knowledge was needed to use the Blaise system.

In the Blaise philosophy, the first step in carrying out a survey was to design a questionnaire in the Blaise language. Such a specification of the questionnaire contains more information than a traditional paper questionnaire. It did not only describe questions, possible answers, and conditions on the route through the questionnaire but also relationships between answers that had to be checked.

Figure 1.6 contains an example of a simple paper questionnaire. The questionnaire contains one route instruction: Persons without job are instructed to skip the questions about the type of job and income.

Figure 1.7 contains the specification of this questionnaire in the Blaise System. The first part of the questionnaire specification was the *Fields section*. It contains the definition of all questions that can be asked. A question consists of an identifying name, the text of the question as presented to the respondents, and a specification of valid answers. For example, the question about age has the name *Age*, the text of the question is "*What is your age?*", and the answer must be a number between 0 and 99. The question *JobDes* is an open question. Any text not exceeding 20 characters is accepted. *Income* is a closed question. There are three possible answer options. Each option has a name (for example, *Less20*) and a text for the respondent (for example, *Less than 20,000*).

The second part of the Blaise specification is the *Rules section*. Here, the order of the questions is specified and the conditions under which they are asked. According to the rules section in Figure 1.7, every respondent must answer the questions *SeqNum*, *Age*, *Sex*, *MarStat*, and *Job* in this order. Only persons with a job ($Job = Yes$) have to answer the questions *JobDes* and *Income*.

The rules section can also contain checks on the answers of the questions. Figure 1.7 contains such a check. If people are younger than 15 years ($Age < 15$), then their marital status can only be not married ($MarStat = NotMar$).

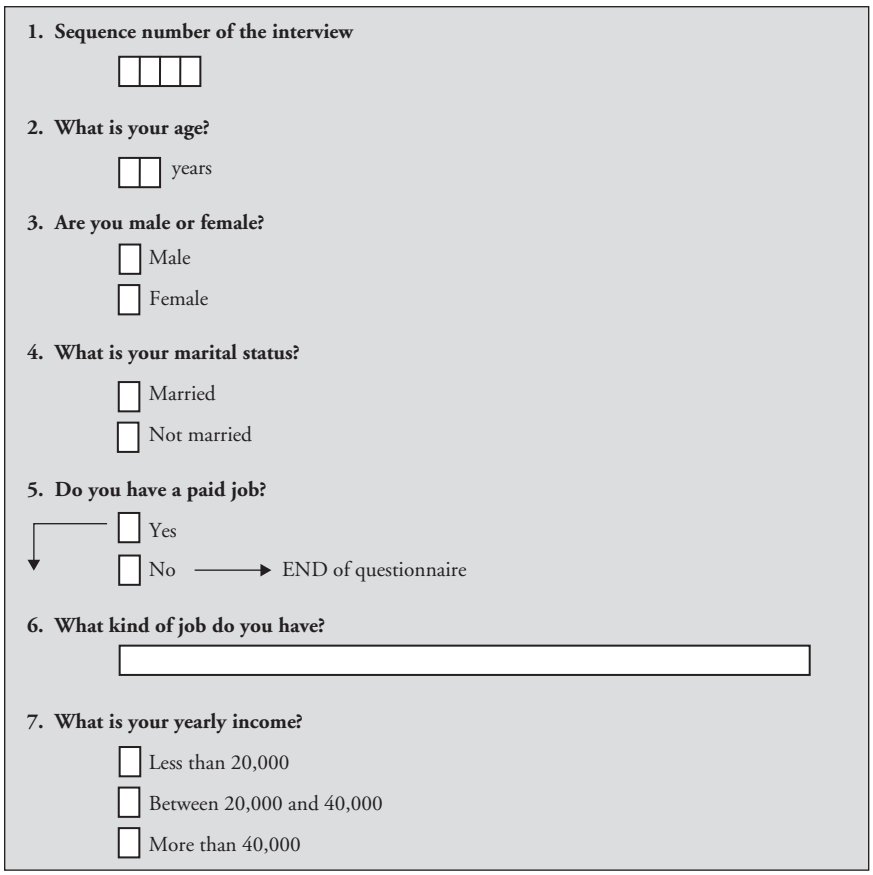

FIGURE 1.6 **A simple paper questionnaire**

The check also contains texts that are used to display the error message on the screen. (*If respondent is younger than 15, then he/she is too young to be married!*)

The rules section also may contain computations. Such computations could be necessary in complex routing instructions or checks, or to derived new variables.

The first version of Blaise used a questionnaire specification to generate a CADI-program. Figure 1.8 shows how the computer screen of this MS-DOS program looked like for the Blaise questionnaire in Figure 1.7.

As this program was used by subject-matter specialists, only question names are shown on the screen. Additional information could be displayed through special keys.

Note that the input fields for the questions *Age* and *MarStat* contain error counters. These error indicators appeared because the answers to the questions *Age* (2) and *MarStat* (*Married*) did not pass the check.

```
DATAMODEL LFS "The Labour Force Survey";

FIELDS
  SeqNum  "Sequence number of the interview?": 1..1000
  Age     "What is your age?": 0..99
  Sex     "Are you male or female?": (Male, Female)
  MarStat "What is your marital status?":
            (Married "Married",
             NotMar "Not married")
  Job     "Do you have a job?": (Yes, No)
  JobDes  "What kind of job do you have?": STRING[20]
  Income  "What is your yearly income?":
             (Less20 "Less than 20,000",
             Upto40 "Between 20,000 and 40,000",
             More40 "More than 40,000")

RULES
  SeqNum Age Sex MarStat Job
  IF Job = Yes THEN
     JobDes Income
  ENDIF

  IF Age < 15 "respondent is younger than 15" THEN
     MarStat = NotMar "he/she is too young to be married!"
  ENDIF

ENDMODEL
```

FIGURE 1.7 A simple Blaise questionnaire specification

FIGURE 1.8 A Blaise CADI program

After Blaise had been in use for a while, it was realized that such a system could be made much more powerful. The questionnaire specification in the Blaise system contained all knowledge about the questionnaire and the data needed for survey processing. Therefore, Blaise should be capable of handling computer-assisted interviewing.

Implementing computer-assisted interviewing means that the paper questionnaire is replaced by a computer program containing the questions to be asked. The computer takes control of the interviewing process. It performs two important activities:

- *Route control*. The computer program determines which question is to be asked next and displays that question on the screen. Such a decision may depend on the answers to previous questions. Hence, it relieves the interviewer of the task of taking care of the correct route through the questionnaire. As a result, it is not possible anymore to make route errors.
- *Error checking*. The computer program checks the answers to the questions that are entered. Range checks are carried out immediately after the answer has been entered and consistency checks after entry of all relevant answers. If an error is detected, the program produces an error message, and one or more of the answers concerned has to be modified. The program will not proceed to the next question until all detected errors have been corrected.

Application of computer-assisted data collection has three major advantages. In the first place, it simplifies the work of interviewer (no more route control); in the second place, it improves the quality of the collected data; and in the third place, data are entered in the computer during the interview resulting in a clean record, so no more subsequent data entry and data editing is necessary.

Version 2 of Blaise was completed in 1988. It implemented one form of computer-assisted interviewing: computer-assisted personal interviewing (CAPI). It is a form of face-to-face interviewing in which interviewers use a small laptop or notebook computer to ask the questions and to record the answers.

Figure 1.9 shows an example of a screen of a CAPI program generated by Blaise. The screen was divided into two parts. The upper part contains the current question to be answered (*What kind of a job do you have?*). After an answer had been entered, this question was replaced by the next question on the route.

Just displaying one question at a time gave the interviewers only limited feedback on where they are in the questionnaire. Therefore, the lower part of the screen displayed (in a very compact way) the current page of the questionnaire.

Statistics Netherlands started full-scale use of CAPI in a regular survey in 1987. The first CAPI survey was the Labor Force Survey. Each month, approximately 400 interviewers equipped with laptops visited 12,000 addresses. After a day of interviewing, they returned home and connected their computers to the power supply to recharge the batteries. The laptop also was connected to a telephone and modem. The collected data were automatically transmitted to the office at night. In return, new addresses were sent to the interviewers. The next

FIGURE 1.9 **A Blaise CAPI program**

morning, the batteries were recharged and the interviewing software was prepared for a new day of interviewing.

Another mode of computer-assisted interviewing was included in 1990: computer-assisted telephone interviewing (CATI). The interviewing program was installed on desktop computers. Interviewers called respondents from a central unit (call center), and they conducted interviews by telephone. The interviewing program for CATI was the same as that for CAPI. An important new tool for CATI was a call scheduling system. This system took care of proper handling of busy numbers (try again shortly), no-answers (try again later), appointments, and so on.

In the very early 1990s, nearly all household surveys of Statistics Netherlands had become CAPI or CATI surveys. Surveys using paper forms had almost become extinct. Table 1.2 lists all major and regular household surveys at that time together with their mode of interviewing.

In the middle of the 1990s, the MS-DOS operating system on microcomputers was gradually replaced by Windows (Microsoft Corporation, Redmond, WA). Particularly, the release of *Windows 95* was a success. It marked the start of the use of graphical user interfaces. Early versions of Microsoft's Internet browser *Internet Explorer* were included in this operating system.

The change of operating systems also had consequences for the Blaise System. Blaise 4 was the first production version of Blaise for Windows. It was released in 1999. The functionality of Blaise did not change, but the graphical user interface offered many more possibilities for screen layout. Figure 1.10 gives an example of a screen of the Blaise 4 CAPI program.

When more and more people and companies were connected to the Internet, web surveys became more and more a popular mode of data collection among researchers. The main reasons for this popularity were the high response

TABLE 1.2 Household surveys carried out by Statistics Netherlands in the early 1990s

Surveys	Modes	Interviews per year
Survey on Quality of Life	CAPI	7,500
Health Survey	CAPI	6,200
Day Recreation Survey	CAPI	36,000
Crime Victimization Survey	CAPI	8,000
Labour Force Survey	CAPI	150,000
Car Use Panel	CATI	8,500
Consumer Sentiments Survey	CATI	24,000
Social-Economic Panel	CATI	5,500
School Career Survey	CATI	4,500
Mobility Survey	CATI / CADI	20,000
Budget Survey	CADI	2,000

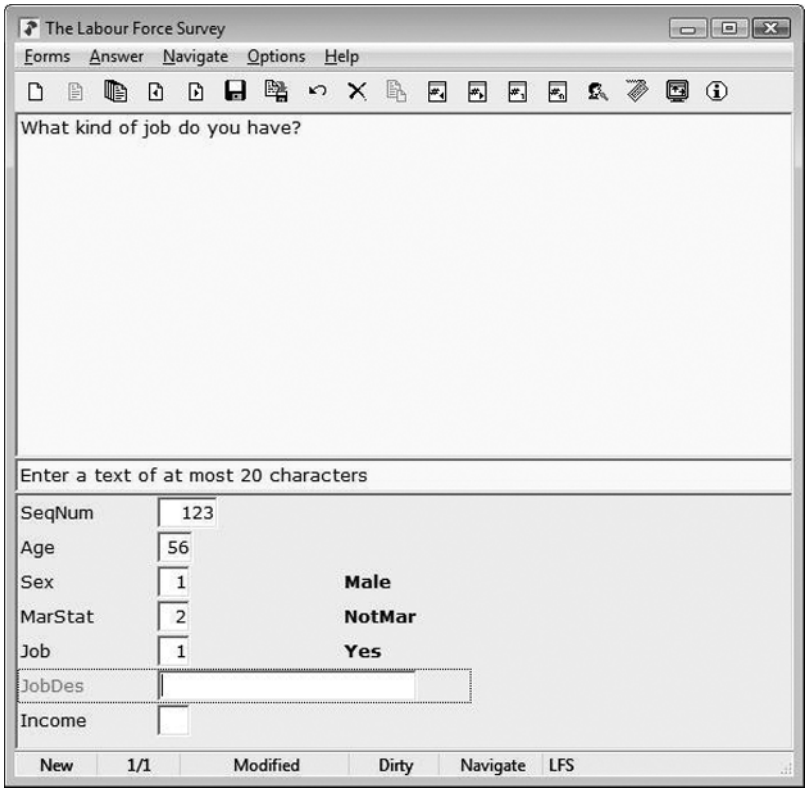

FIGURE 1.10 The screen of a CAPI program in Blaise 4

speed, the possibility to provide feedback to respondents about the meaning of questions and possible errors, and the freedom for the respondents to choose their own moment to fill in the questionnaire.

An attractive property of web surveys is that the questionnaire is completed in a browser. Respondents are familiar with browsers because they also use them for all their other activities on the Internet. So there was no need to explain the graphical user interface. The user interfaces for CAPI and CATI software were not that straightforward. Therefore, use of these software tools for computer-assisted self-interviewing was much more demanding.

The possibility of conducting web surveys was included in version 4.6 of Blaise that was released in 2003. Blaise provided two different approaches for web surveys: the interview approach and the form approach.

The *interview approach* is typically used for long and complex questionnaires that contain routing (skip) instructions and checks on the given answers. The respondent completes the questionnaire on-line because continuous interaction is required between the computer of the respondent and the software on the Internet server.

The Internet questionnaire is divided into pages. Each page may contain one or more questions. After the respondent has answered all the questions on a page, the answers are submitted to the Internet server. The answers are checked, and a new page is returned to the respondent. The contents of this page may depend on the answers to previous questions.

Figure 1.11 shows a simple example of a screen of a web survey. In this case, the page contains only one question. A progress indicator in the upper right corner keeps respondents informed about their progress in the questionnaire.

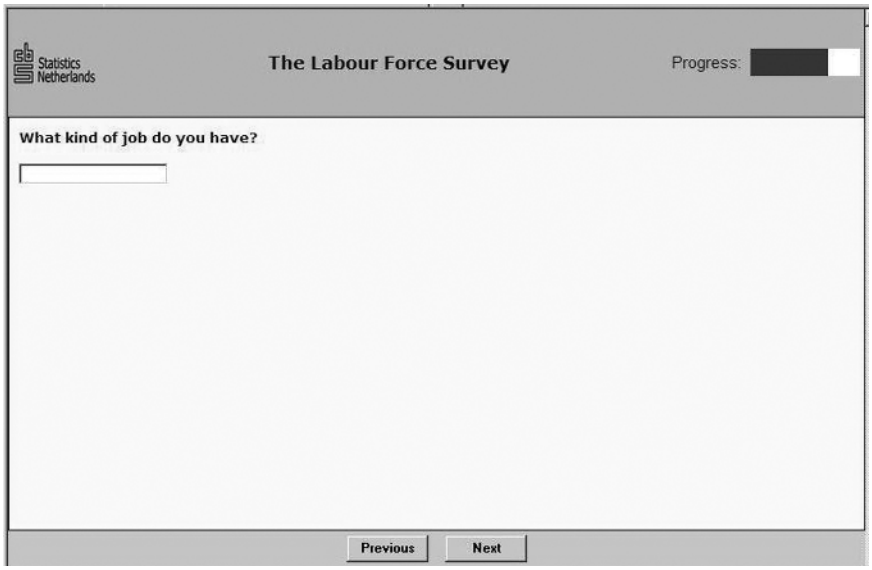

FIGURE 1.11 **A web questionnaire using the interview approach**

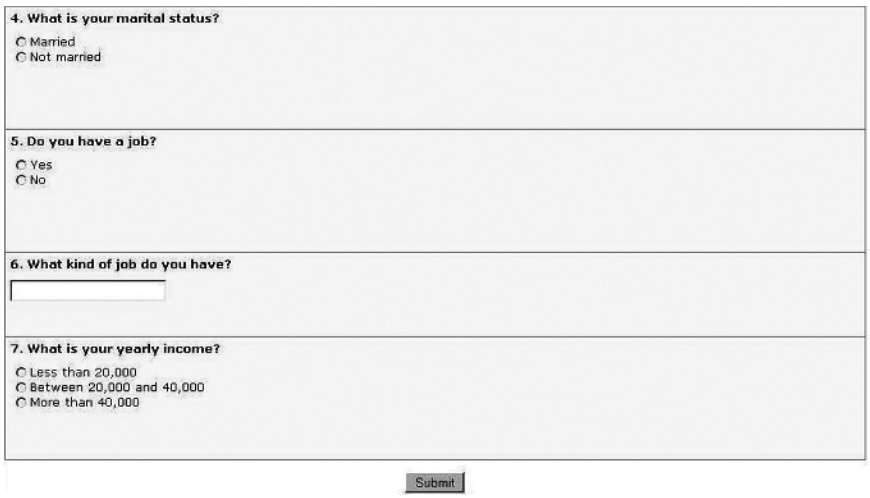

FIGURE 1.12 **A web questionnaire using the form approach**

Blaise 4.6 also supports a *form approach*. This approach is suitable for short and simple questionnaires with straightforward data entry without question routing. The Internet form is just like a paper form. There is no extra functionality. The questionnaire comprises one web page that can be scrolled up and down to answer the questions.

The web questionnaire form may be delivered in several ways. One way is to offer its URL on a website, and another way is to send a form to respondents as an attachment to an e-mail.

All questions are presented in a fixed sequence. Respondents can browse through the form and answer questions in any order. They can fill out questions off-line. There is no need for continuous contact between respondents and the Internet server while they answer questions. However, this does not mean the form cannot be filled out on-line. Figure 1.12 contains an example of a form-based web questionnaire in Blaise.

Because there is no contact between the respondent and the server, no routing or checking is possible during the interview. However, the client computer does execute range checks for entered answers.

When respondents have completed their questionnaire, they submit their form to the server (after making an Internet connection if none is present). Then the answers are sent to the Internet server. The server can execute the checking mechanism if desired and can store the data in a Blaise database. The respondent receives a confirmation from the server.

More about the development of the Blaise System an its underlying philosophy can be found in the study by Bethlehem and Hofman (2006).

1.4 Summary

Web surveys are a next step in the evolution process of survey data collection. Collecting data for compiling statistical overviews is already very old, almost as old as humankind. All through history, statistics have been used by rulers of countries to take informed decisions. However, new developments in society always have had their impact on the way the data were collected for these statistics.

For a long period, until the year 1895, statistical data collection was based on complete enumeration of populations. The censuses were mostly conducted to establish the size of the population, to determine the tax obligations of the people, and to measure the military strength of the country.

The first ideas about sampling emerged around 1895. There was a lot of discussion between 1895 and 1934 about how samples should be selected: by means of probability sampling or some other sample selection technique. By 1934 it was clear that only surveys based on probability sampling could provide reliable and accurate estimates. Such surveys were accepted as a scientific method of data collection.

Somewhere in the 1970s another significant development started. The fast development of microcomputers made it possible to introduce computer-assisted interviewing. This made survey data collection faster, cheaper, and easier, and it increased data quality. It was a time when acronyms like CATI and CAPI emerged.

The next major development was the creation of the Internet around 1982. When more and more persons and companies got access to the Internet, it became possible to use this network for survey data collection. The first Internet surveys were e-mail surveys. In 1989 the World Wide Web was developed. In the middle of the 1990s, web surveys became popular.

Web surveys are attractive because they allow for simple, fast, and cheap access to large groups of potential respondents. There are, however, also potential methodological problems. There are ample examples of web surveys that are not based on probability sampling. It is not always easy to distinguish good from bad surveys.

KEY TERMS

Blaise: A software package for computer-assisted interviewing and survey processing developed by Statistics Netherlands.

Census: A way of gathering information about a population in which every element in the population has to complete a questionnaire form.

Computer-assisted interviewing (CAI): A form of interviewing in which the questionnaire is not printed on paper. Questions are asked by a computer program.

Computer-assisted personal interviewing (CAPI): A form of face-to-face interviewing in which interviewers use a laptop computer to ask the questions and to record the answers.

Computer-assisted self-administered questionnaires (CSAQ): A form of data collection in which respondents complete the questionnaires on their own computer. See also CASI.

Computer-assisted self-interviewing (CASI): A form of data collection in which respondents complete the questionnaires on their own computer. See also CSAQ.

Computer-assisted telephone interviewing (CATI): A form of telephone interviewing in which interviewers use a computer to ask the questions and to record the answers.

E-mail survey: A form of data collection via the Internet in which respondents are sent a questionnaire that is part of the body text of an e-mail. The questionnaire is completed by returning the e-mail after answering the questions in the text.

Face-to-face interviewing: A form of interviewing where interviewers visit the homes of the respondents (or another location convenient for the respondent). Together, the interviewer and the respondent complete the questionnaire.

Mail survey: A form of data collection where paper questionnaire forms are sent to the respondents. After completion of the questionnaires, they are returned to the research organization.

Purposive sampling: A form of nonprobability sampling in which the selection of the sample is based on the judgment of the researcher as to which elements best fit the criteria of the study.

Probability sampling: A form of sampling where selection of elements is a random process. Each element must have a positive and known probability of selection.

Quota sampling: A form of purposive sampling in which elements are selected from the population in such a way that the distribution of some auxiliary variables matches the population distribution of these variables.

Random digit dialing (RDD): A form of sample selection for a telephone survey where random telephone numbers are generated by some kind of computer algorithm.

Representative method: A method proposed by Anders Kiaer in 1896 to select a sample from a population in such a way that it forms a "miniature" of the populations.

Straw poll: An informal survey conducted to measure a general feeling of a population. Sample selection is such that it usually does not allow us to draw conclusions about the population as a whole.

Survey: A way of gathering information about a population in which only a sample of elements from the population has to complete a questionnaire form.

Telephone interviewing: A form of interviewing in which interviewers call selected persons by telephone. If contact is made with the proper person, and this person wants to cooperate, the interview is started and conducted over the telephone.

Web survey: A form of data collection via the Internet in which respondents complete the questionnaires on the World Wide Web. The questionnaire is accessed by means of a link to a web page.

EXERCISES

Exercise 1.1. Which of the following options is not an advantage of computer-assisted interviewing (CAI) as compared with traditional modes of data collection

a. Data quality is higher because of included checks.

b. The software is in charge of routing through the questionnaire.

c. CAI leads to higher response rates.

d. Data can be processed quicker.

Exercise 1.2. What is an advantage of an e-mail survey over a traditional mail survey?

a. Data quality is higher because of included checks.

b. There is less undercoverage.

c. Response rates are higher.

d. It has better facilities for navigation through the questionnaire.

Exercise 1.3. Why were the first surveys on the Internet e-mail surveys and not an the web surveys?

a. E-mail surveys were cheaper.

b. The World Wide Web did not exist yet.

c. E-mail surveys are more user-friendly.

d. E-mail surveys require less data communication over the Internet.

Exercise 1.4. When should the form-based approach be preferred over the question-by-question approach in a web survey?

a. The questionnaire is very long.

b. The questionnaire contains route instructions and edits.

c. All questions fit on one screen.

d. The survey is a business survey.

Exercise 1.5. Which of the four features is typically an advantage of web surveys;

a. There is no undercoverage.

b. The sample size is always large.

c. A survey can be designed and conducted very quickly.

d. Accurate estimates can always be computed.

Exercise 1.6. How can the problem of undercoverage in a general population web survey be avoided?

a. Conduct a mixed-mode survey.

b. Increase the sample size.

c. Conduct a self-selection web survey.

d. Replace the web survey by an e-mail survey.

REFERENCES

Bethlehem, J. G. (1987), The Data Editing Research Project of the Netherlands Central Bureau of Statistics, *Proceedings of the Third Annual Research Conference of the US Bureau of the Census*, U.S. Bureau of the Census, Washington, DC, pp. 194–203.

Bethlehem, J. G., Cobben, F., & Schouten, B. (2011), *Handbook on Nonresponse in Household Surveys*. John Wiley & Sons, Hoboken, NJ.

Bethlehem, J. G. (2009), *The Rise of Survey Sampling*. Discussion Paper 09015, Statistics Netherlands, The Hague/Heerlen, the Netherlands.

Bethlehem, J. G. & Hofman, L. F. M. (2006), Blaise—Alive and Kicking for 20 Years. *Proceedings of the 10th International Blaise Users Conference*, Arnhem, the Netherlands, pp. 61–86.

Bethlehem, J. G. & Stoop, I. A. L. (2007), Online Panels—A Theft of Paradigm? The Challenges of a Changing World. *Proceedings of the Fifth International Conference of the Association of Survey Computing*, Southampton, U. K., pp. 113–132.

Beukenhorst, D. & Wetzels, W. (2009), *A Comparison of Two Mixed-mode Designs of the Dutch Safety Monitor: Mode Effects, Costs, Logistics*. Technical Paper DMH 206546, Statistics Netherlands, Methodology Department, Heerlen, the Netherlands.

Bowley, A. L. (1906), Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. *Journal of the Royal Statistical Society,* 69, pp. 548–557.

Bowley, A. L. (1926), Measurement of the Precision Attained in Sampling. *Bulletin of the International Statistical Institute*, XII, Book 1, pp. 6–62.

CBS (1948), Enige Beschouwingen over Steekproeven. Reprint from *Statistische en Economische Onderzoekingen* 3, Statistics Netherlands, The Hague, the Netherlands.

Clayton, R. L. & Werking, G. S. (1998), Business Surveys of the Future: The World Wide Web as a Data Collection Methodology. In: Couper, M. P., Baker, R. P., Bethlehem, J. G., Clark, C. Z. F., Martin, J., Nicholls, W. L., & O'Reilly, J. (eds.), *Computer Assisted Survey Information Collection*. John Wiley & Sons, New York.

Cobben, F. & Bethlehem, J. G. (2005), *Adjusting Undercoverage and Non-response Bias in Telephone Surveys*. Discussion Paper 05006. Statistics Netherlands, Voorburg/ Heerlen, the Netherlands.

Cochran, W. G. (1953), *Sampling Techniques*. John Wiley & Sons, New York.

Couper, M. P., Baker, R. P., Bethlehem, J. G., Clark, C. Z. F., Martin, J., Nicholls II, W. L., & O'Reilly, J. M. (eds.) (1998), *Computer Assisted Survey Information Collection*. John Wiley & Sons, New York.

Couper, M. P., Blair, J., & Triplett, T. (1999), A Comparison of Mail and E-mail for a Survey of Employees in U. S. Statistical Agencies. *Journal of Official Statistics,* 15, pp. 39–56.

Couper, M. P. & Nicholls, W. L. (1998), The History and Development of Computer Assisted Survey Information Collection Methods. In: Couper, M. P., Baker, R. P., Bethlehem, J. G., Clark, C. Z. F., Martin, J., Nicholls, W. L., & Reilly, J. (eds.), *Computer Assisted Survey Information Collection*. John Wiley & Sons, New York.

Deming, W. E. (1950), *Some Theory of Sampling*. John Wiley & Sons, New York.

Den Dulk, K. & Van Maarseveen, J. (1990), The Population Censuses in The Netherlands. In: Van Maarseveen, J. & Gircour, M. (eds.), *A Century of Statistics, Counting, Accounting and Recounting in The Netherlands*. Statistics Netherlands, Voorburg, the Netherlands.

Dillman, D. A. (2007), *Mail and Internet Surveys: The Tailored Design Method*. John Wiley & Sons, Hoboken, NJ.

Graunt, J. (1662), *Natural and Political Observations upon the Bills of Mortality*. Martyn, London, U. K.

Hansen, M. H., Hurvitz, W. N., & Madow, W. G. (1953), *Survey Sampling Methods and Theory*. John Wiley & Sons, New York.

Horvitz, D. G. & Thompson, D. J. (1952), A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association,* 47, pp. 663–685.

Kiaer, A. N. (1895), Observations et Expériences Concernant des Dénombrements Représentatives. *Bulletin of the International Statistical Institute*, IX, Book 2, pp. 176–183.

Kiaer, A. N. (1997 reprint), Den Repräsentative Undersökelsesmetode. *Christiania Videnskabsselskabets Skrifter. II. Historiskfilosofiske klasse*, Nr 4 (1897). English translation: *The Representative Method of Statistical Surveys*, Statistics Norway, Oslo, Norway.

Kiesler, S. & Sproull, L.S. (1986), Response Effects in the Electronic Survey. *Public Opinion Quarterly*, 50, pp. 402–413.

Kuusela, V., Vehovar, V., & Callegaro, M. (2006), *Mobile phones—Influence on Telephone Surveys*. Paper presented at Second International Conference on Telephone Survey Methodology, Miami, FL.

Laplace, P. S. (1812), *Théorie Analytique des Probabilités. Oevres Complètes*, Vol. 7, Gauthier-Villar, Paris, France.

Lienhard, J. H. (2003), *The Engines of Our Ingenuity, An Engineer Looks at Technology and Culture*. Oxford University Press, Oxford, U. K.

Neyman, J. (1934), On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society,* 97, pp. 558–606.

Nicholls, W. L. & Groves, R. M. (1986), The Status of Computer Assisted Telephone Interviewing. *Journal of Official Statistics,* 2, pp. 93–134.

NIPO, (1946), Eerste Telefonische Enquête in Nederland Verricht Door NIPO. *De Publieke Opinie*, 1e jaargang, No. 4, p. 1.

O'Connell, P. L. (1998), Personal Polls Help the Nosy Sate Curiosity. *New York Times*, June 18.

Quetelet, L. A. J. (1835), *Sur l'Homme et le Développement de ses Facultés, Essai de Physique Sociale*. Paris, France.

Quetelet, L. A. J. (1846), *Lettre à S.A.R. le Duc Régant de Saxe Coburg et Gotha sur la Théorie des Probabilités, Appliquée aux Sciences Morales at Politiques*. Brussels, Belgium.

Roos, M., Jaspers, L., & Snijkers, G. (1999), *De Conjunctuurtest via Internet*. Report H4350–99-GWM, Statistics Netherlands, Data Collection Methodology Department, Heerlen, the Netherlands.

Roos, M. & Wings, H. (2000), Blaise Internet Services Put to the Test: Web-surveying the Construction Industry. *Proceedings of the 6th International Blaise Users Conference*, Kinsale, Ireland.

Saris, W. E. (1998), Ten Years of Interviewing Without Interviewers: the Telepanel. In: Couper, M. P., Baker, R. P., Bethlehem, J. G., Clark, C. Z. F., Martin, J., Nicholls II, W. L., & O'Reilly, J. M. (eds.), *Computer Assisted Survey Information Collection*. John Wiley & Sons, New York, pp. 409–430.

Schaefer, D. R. & Dillman, D. A. (1998), Development of a Standard E-mail Methodology: Results of an Experiment. *Public Opinion Quarterly,* 62, pp. 378–397.

Snijkers, G., Tonglet, J., & Onat, E. (2004), *Projectplan Pilot e-PS*. Internal Report H3424–04-BOO, Development and Support Department, Division of Business Statistics, Statistics Netherlands, Heerlen, the Netherlands.

Snijkers, G., Tonglet, J., & Onat, E. (2005), *Naar een Elektronische Vragenlijst voor Productiestatistieken*. Internal Report, Development and Support Department, Division of Business Statistics, Statistics Netherlands, Heerlen, the Netherlands.

Utts, J. M. (1999), *Seeing Through Statistics*. Duxburry Press, Belmont, CA.

Witte, J. C., Amoroso, L. M., & Howard, P. E. N. (2000), Method and Representation in Internet-based Survey Tools. *Social Science Computer Review,* 18, pp. 179–195.

Yates, F. (1949), *Sampling Methods for Censuses and Surveys*. Charles Griffin & Co, London, U. K.