This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method

Contents

General section

1. Summary

The chapter gives an overview of factors to consider when choosing data collection method. It also gives a short presentation of different modes available, modes suitable for business surveys, advantages and disadvantages with each mode and a brief description about how to mix modes.

2. General description

2.1 Factors to consider when choosing data collection method

There are several factors to consider when choosing data collection method and each method has its pros and cons. A general idea is to choose the method that minimises the total survey error (TSE) given the budget constraints. Some factors affecting the choice of mode and data collection instrument are response burden, desired data quality (e.g., in terms of nonresponse and measurement error), available resources (budget and staff, but also IT-resources and technical conditions), topic of the survey and the questionnaire content, sampling frame, properties of the target population (e.g., type of industry) and timetable for the survey (e.g., Biemer et al., 1991; Groves et al., 2004).

For instance, response burden can be reduced by good questionnaire design, extracting files automatically or by pre-printing information from previous reporting periods in the questionnaire. Lower response burden may also be achieved by sample coordination and sample rotation. For long surveys with complex calculations, an electronic self-administered questionnaire that guides the respondent through the form with built-in helps and logic checks might be an appropriate alternative. Some electronic questionnaires might also allow the reporting person to save data temporarily and continue later on if figures have to be looked up in other systems or files. Regardless what method is chosen, a contact strategy must also be defined when planning the data collection; how and when the respondents will be contacted.

One major difference between household surveys and business surveys is that in business surveys (most often) many employees cooperate in the reporting task, something that makes the response situation more complex; see the module "Response – Response Process". We do not know much about how the tasks are divided or communicated internally within the businesses, we can only suppose this complexity makes questionnaire design even more important. Some employees might forward the whole questionnaire including instructions to a colleague; while others might interpret the question themselves and just ask the colleague for a figure (i.e., the colleague will never see or read neither the question nor the instructions). In some businesses only a few persons are authorised to report, but this does not necessarily mean that the authorised person has the knowledge to report. The questionnaire might be sent around to different employees within the business who partially fill out and report the figures they have knowledge on. In some businesses paper questionnaires are preferred, because "paper walks". Other businesses find electronic self-completion questionnaires easier to handle in the reporting situation. The differences in preferences are often related to factors like for instance business size, organisation levels (hierarchy) and type of industry.

Business surveys are also a bit special in the sense that business populations have distinct frame problems. Often they vary quite much in size and they are highly dynamic. Small businesses are born and die rapidly. Medium-sized or large businesses merge with others or split up into several units. The

business population also demonstrates a distinction between a legally defined entity and physical location (Groves et al., 2004). These are also factors to consider when designing data collection and choosing mode.

Another important step in planning the data collection is to consider how the final result, the statistics should be presented. Which variables should be reported and how detailed should they be? How shall we get hold of this information; shall the variables be collected from a register, shall they be collected directly through a questionnaire or are the variables so complex that they have to be created by compound calculations? These kinds of choices will not only affect the level of response burden in the survey, but also the level of accuracy during the data collection which is also an important design feature which should be reflected in the choice of mode. In an interview, the interviewer can give the respondent more support than in a postal questionnaire, where there are limited opportunities to help the respondent to fulfil the task. In electronic self-administered questionnaires, controls can be built in which can be both an advantage and a disadvantage for the respondent. When designing the data collection instrument, research problems have to be translated into questions in the questionnaire without creating a mismatch opening up for specification- and measurement errors. One also has to ensure that all topics are covered in the questionnaire, i.e., no variables are missing. The planning and design process is a continuous process where improvements are made by iterations. Instrument design and testing questionnaires are dealt with more in detail in the topic "Questionnaire Design".

Each survey has its own conditions, specific errors and how to treat them. In general, little is known about the relationship between quality, time, costs and response burden and it is hard to implement measures to reduce the burden without the expense of quality. Too few quantitative before-after studies are at present documented and actions intended to reduce response burden should be monitored, reviewed, documented and published better in order to gain more insight (Giesen, 2011).

2.2 *Different modes*

The *mode of data collection* refers to what medium is used for contacting the sample members to get their responses to the survey questions. The principal modes for data collection are: *face-to-face surveys*, *telephone survey, mail surveys* and *web surveys.* Face-to-face surveys and telephone surveys are often referred to as *interviewer-administered modes*, whereas mail surveys and web surveys are referred to *self-administered modes*.

The data collection can also be divided into direct and indirect data collection, referring to the level of contact with the respondent. For instance, administrative records are an indirect form for data collection with no contact with the respondent and a low data collector involvement; this in contrast to many of the other modes which are methods for direct data collection. The table below gives an overview over different modes, the level of data collection involvement from the data collector and level of contact with the respondent.

Table 1. Modes to choose from when planning the data collection.

	High Data Collector Involvement		Low Data Collector Involvement	
	Paper	Computer	Paper	Computer
Direct Contact with Respondent	Face-to-face (PAPI)	CAPI	Diary	CASI, ACASI
Indirect contact with Respondent	Telephone (PAPI)	CATI	Mail, fax, e-mail	TDE, e-mail, Web, DBM, EMS, VRE
No Contact with Respondent	Direct observation	CADE	Administrative records	EDI

ACASI, audio CASI; CADE, computer-assisted data entry; CAPI, computer-assisted data interviewing; CASI, computer-assisted self-interviewing, CATI, computer-assisted telephone interviewing; DBM, disc by mail; EDI, electronic data interchange; EMS, electronic mail survey; PAPI, paper-and-pencil interviewing; T-ACASI, telephone ACASI; TDE, touch-tone data entry; VRE, voice recognition entry. *Source: Biemer and Lyberg (2003).*

The modes have different advantages and disadvantages when it comes to costs, measurement errors, nonresponse and coverage, flexibility and timeliness. Questionnaire complexity and the respondents' possible reporting preferences are also important factors to consider, something that sometimes leads to a mixed mode solution when collecting data for the survey. Mixed-mode design might help in satisfying the respondent's preferences and hereby the response burden might be lowered. Even if lower response burden is highly desirable, it might sometimes be wise not to offer too many different modes at the same time. This is because too many computer systems to look after for the national statistical institute (hereafter called NSI) will be costly in the long run. Mixed mode also opens up for possible different error sources that might be difficult to combine and handle later on in the statistical process.

Below follows a short review of some of the modes presented in Table 1. The review primarily focuses on the modes relevant for business surveys, but as always there are exceptions and differences between countries depending on domestic conditions, which might have the greatest impact on the choice of mode at the end.

2.2.1 Mail surveys

The mail survey is carried out by a paper questionnaire sent to the sample respondents by mail. The data collector has no control over the response process or who is actually responding to the survey (e.g., Biemer et al., 1991). The response process is as previously mentioned even more complex in business surveys and sometimes it is a challenge just to find the right person within the business to mail the questionnaire to.

Mail surveys are quite inexpensive to implement, which make them the preferred mode for low-budget surveys. At the same time, mail surveys often require a long field period with at least one reminder to achieve acceptable response rates (Biemer and Lyberg, 2003). The respondent deals with the survey on its own and there is no interviewer present who can provide support or explain difficult questions. Some NSIs have chosen to have a support centre or help desk for business surveys, which the business representatives can call and ask for help when reporting. It is also common to include a telephone

number to the person who is responsible for the publication or statistical analysis in the questionnaire or in the advance letter.

The potential problem with complicated questions can be eased by a well-designed questionnaire that motivates and guides the respondent through the questionnaire by good navigation, help texts and visual support (e.g., Groves et al., 2004). Visual support and technical facilities can be made extra efficient in electronic self-completion questionnaires (see next section 2.2.2).

The quality of the answers in a mail questionnaire is to a greater extent depending on the questionnaire design than in interviews. However, it has been shown that response order and question order is less important in a mail survey, as the respondent can easily navigate back and forth in the questionnaire (Biemer et al., 1991). There is also less risk of social desirable responses for sensitive issues in mail surveys than in the interviewer-respondent situation (Biemer et al., 1991). For mail questionnaires there is a greater risk of *primacy effects*, i.e., the respondent chooses one of the first response categories when answering the question (e.g., de Leeuw, Hox and Dillman, 2008). *Open-ended questions*, where the respondent has to formulate the response on his/her own are less suitable for mail questionnaires. The respondents have proven to give less and less thoughtful answers to such questions in mail surveys than in an interview situation where the interviewer can help the respondent in formulating the answer by probing. In business surveys open-ended questions might lead to a situation where the data collector does not know what is included in the numbers reported. Without the interviewer directly motivating the respondent to participate, mail surveys typically have lower response rates than interviews and the risk of item nonresponse is also bigger in mail surveys (Biemer and Lyberg, 2003). However, the nonresponse rate is in general not the biggest problem in business surveys, since reporting most often is mandatory and failure to report will lead to fines.

2.2.2 Web surveys

Web surveys are based on self-administered electronic questionnaires which are often viewed upon as a technical version of the mail questionnaire. Logic checks and visual guidelines can be built in, but advanced solutions cost hours of programming and there is a risk of ending up with higher response burden due to all the technical features if they are not well specified and tested.

Web surveys are perhaps the most common mode for business surveys today. Many NSIs introduce electronic versions of the survey due to aims in cutting the costs for data collection and/or data editing, with the intention to improve data quality, in order to offer safe communication with businesses or in order to make it easier to respond and thereby aiming to lower the response burden (Giesen, 2011, Chapter 5).

Web surveys might also be offered for specific surveys or specific groups of surveys where reporting on the web has been found to suit the survey topic well, or where different versions of the questionnaire are sent to different subgroups in the population (e.g., small businesses).

Computerisation allows lots of built-in features like customised wording, mouse-over-help, skips and jumps, edit checks and randomised question order. These features or refinements can be said to replace the role of an interviewer that helps the respondent through the survey. Visual elements like brightness, color, shape and position can be used in order to guide the respondent through the questionnaire (Groves et al., 2004). These features have shown to lead to less measurement error and less item non-response (ibid). The visual potential might also lower the response burden.

A factor to be considered when choosing the most suitable mode is that web surveys can be run on-line or off-line. As described in the module "Data Collection – Techniques and Tools", these two ways offer the respondents the opportunity to compile the questionnaire directly on the survey web site or to download it, fill it out and send it back later on when finished.

Some examples of web-surveys in Europe: Statistics Norway introduced electronic reporting for all business surveys as a new data collection strategy; the primary data collection mode is nowadays the web (e.g., Haraldsen et al., 2011). Statistics Lithuania introduced web-surveys to create a favourable environment for the businesses in order to prepare statistical data at lower costs (e.g., Lapeniene, 2008). At Statistics Netherlands, more than half of the business surveys are available in electronic forms (e.g., Beukenhorst and Giesen, 2010) and in the latest years, work has been targeted on an electronic version of the annual Structural Business Survey (e.g., Snijkers et al., 2007) on the web. Further examples can be found in Raymond-Blaess (2011).

No matter the reason behind an electronic version of a self-completion questionnaire, there is no clear evidence that web-surveys do imply higher data quality and decreased response burden, even if some measurements suggests something in that direction (Snijkers et al., 2007: Giesen et al., 2009). Electronic data collection adds complexity to the response process which is already complicated within a business, and the respondent has to interact not only with the questions, but also with their internal records and the electronic instrument itself. Initially, switching from paper to electronic questionnaire might actually increase the (perceived) response burden and how well an electronic instrument will work in a business survey depends on several factors, such as the organisational structure, the size of the business, what industry the business operates in and the kind of products or services it sells (e.g., Goddeeris and Bruynooghe, 2011; Gravem, Haraldsen and Löfgren, 2011). Not all survey topics are suitable for electronic reporting. Sometimes a paper questionnaire is more convenient for the respondent because it is easier to handle in the reporting situation. On the other hand, electronic questionnaires can be designed to offer the same flexibility the respondent perceives it has with a paper questionnaire, something that can be achieved by creating a web-portal. The portal is not only a place to gather the surveys; it is also a system for survey administration - both for the respondents and the NSI. An example of a web portal is the AltInn-portal in Norway, where different informants can log-on and report on the parts they can contribute to and have knowledge on. This kind of web-portal solution is getting more and more common in Europe.

2.2.3 Administrative records

If existing administrative records can be used, there is not only money to save but also response burden since the respondents will not have to cope with another survey request. The error structure for administrative data is similar to those of other modes, this because the administrative records are produced on data collected somehow originally (Biemer and Lyberg, 2003). Administrative records might consist of data collected by some other institution than the NSI, but might also be data already collected by the NSI in a different survey. A good property with administrative records is that they most often cover the whole population. On the other hand, the drawbacks with administrative records is mainly that they may relate to a somewhat different population than the target population of the survey, leading to calls for further measures to achieve coverage. The content of the records is not always adapted to the wishes of statistics users and statisticians sometimes have no control over the record or how the record is updated (Biemer and Lyberg, 2003). Definitions, boundaries and variable

content may differ from those desired, so the parameters cannot be estimated easily and the NSI sometimes has to rely on model-based estimates. It is not unusual that the statistical purpose of a record comes in second hand, after the administrative ones which often are of primary interest. Different records have different data quality and this goes back to the main data collection or how the record is updated. Conceptual problems are common, especially when it comes to business surveys where there often is a mismatch between what data the businesses have and what data the NSIs ask for (Giesen, 2011).

2.2.4 Touch-tone Data Entry (TDE)

TDE is an alternative to mail collection and is a method where the respondent calls a computer linked to an automatic answering machine and reports by pressing the touchtone phone buttons. Usually, the answers are also read back for the respondent for verification (Biemer and Lyberg, 2003). TDE is only a good option in very short surveys with few questions where the answers are related to numerical information. There are, unfortunately, not many surveys that meet these requirements and there are also some up-front costs associated with using TDE in a survey, e.g., to program the hardware. The possibilities for editing during the process are also limited under this mode (Cox et al., 1995).

2.2.5 Electronic Data Interchange (EDI)

Electronic exchange of information is nowadays standard in the business world as many businesses are moving towards a paperless environment. EDI offers businesses an electronic way to exchange common standard information like order forms, shipping notes and other documents (Cox et al., 1995). The possibility to submit data by removing a file from the system and sending it to the NSI has many advantages. The respondents extract the needed data in a pre-specified format from their computer systems and transfer them to the NSI. Sophisticated EDI systems also offer direct on-line editing by the respondent (Cox et al., 1995). There is a minimal effort for the respondent, except for the first time when the base file has to be created, and response burden is therefore low. The quality of the data is dependent on the file but if it is created and updated correctly the quality might be good. The EDI technique may be used to collect large volumes of data and information from businesses.

2.2.6 Data provided by automatically extracted files (e.g., XBRL)

eXtensible Business Reporting Language (XBRL) is a technical standard for electronic communication of business and financial data and is based on the XML and Link technical standards. The idea of the XBRL language is to identify each concept (e.g., turnover) and add it into a "taxonomy", which works like a dictionary. Once defined, they can be re-used by other users. The technique has potentials in reducing response burden (Allen and Junker, 2008) and offers flexibility to the businesses. XBRL might be a good solution for businesses of large size and/or businesses that do not report themselves, but use an external accountant that have to report on the same survey on a regular basis (Goddeeris and Bruynooghe, 2011).

The relationship between computerisation and quality is not straight forward. The main strength of computers is not that they do things right, but that they do things consistently. This means that in case of incorrect programming or linkage between the statistical need and the source of information, the computer program will consistently produce errors as a result.

The XBRL-technology also struggles with two kinds of updating problems. The first is linked to when questions in the survey are changed and the second is more related to changes in staff. When questions are changed, the software company has to develop a new version and implement it at the customers, which might be a diminishing problem as more and more software updates are available on Internet. Still, this fact implies that automatic data capture will work best in stable environments with fixed survey contents. The second problem is the transfer of competence when people leave a workplace; ensuring the knowledge and experience to link the administrative systems with the statistical ones will be transferred to someone else within the company (e.g., Haraldsen et al., 2011).

Many NSIs are active in this field with different development projects; for instance Statistics Finland developed an automated data capture procedure for hotel accommodations in 2005 (Savolainen and Vertanen, 2007; Orjala, 2010). Destatis in Germany developed the eSTATISTIK.core (2008) which uses the XML file format, and the statistical bureau in Spain – Instituto Nacional de Estadistica – developed a XML based system for the hotel occupancy survey 2008 (INE, 2008). Another successful project that shows the potentials within this area is the Simplified Business Information system (Portuguese acronym IES) developed in partnership with different public entities, including Statistics Portugal. The system makes it possible to acquire administrative and statistical information in a coordinated manner, conducted electronically on one single occasion for the whole population of enterprises. The IES system also represents an improvement on the quality dimensions; coverage, coherence, punctuality, timeliness, comparability and reliability for business statistics (Pereira, 2011).

2.2.7 Face-to-face interviewing – PAPI and CAPI

Face-to-face (PAPI) interview is the oldest mode of interview since it does not rely on modern technology. The mode involves direct contact with the respondent and the data collector is highly involved. When a computer is used instead of paper-and-pencil in the interview situation, the mode is often referred to as CAPI.

PAPI and CAPI are not very common modes in business surveys; however they are used in some countries that for instance lack a business register and/or have problems in locating or contacting the businesses. There might also be some survey specific circumstances when the modes might be a good choice; e.g., when the respondent clearly would benefit the support from an interviewer (e.g., help in recalling events, amounts or frequencies of some phenomenon) or has no access to Internet.

PAPI and CAPI are by far the most expensive data collection methods especially when the respondents are spread over large geographic areas; mainly because of travel and lodging expenses for interviewers as well as interviewer training. In the case of CAPI the interviewer also has to be equipped with a computer. The mode has traditionally been associated with high quality, mainly due to the interviewer's presence and the positive effects from that. Besides for CAPI, the pc-support has the same advantages mentioned for web surveys.

This view has changed in recent decades due to the discovery of measurement error and the problems face-to face interviewing potentially brings, especially for questions on sensitive topics (Biemer et al., 1991). Personal contact is efficient when persuading respondents to participate, something often mirrored in the high response rates for face-to-face interviewing compared to other modes. A face-to-face interview may be longer and cove more complex issues than a telephone interview or a questionnaire sent by mail. At the interview the interviewer can control the response situation; that the

respondent has understood the question and ensure that the response is not influenced by other persons, or that it is the intended respondent who responds to the survey and not someone else. The latter is for instance something out of the NSIs control when sending out a questionnaire by mail.

Another advantage with the face-to-face interview is that the interviewer can use visual aids in the field work, e.g., cards with response categories; something that would not be possible in a telephone interview situation (Biemer et al., 1991). The presence of an interviewer can also have a negative effect on the responses and the quality of the data collected; interviewers affect the respondents' answers in a way similar to the clustering effect in cluster sampling. The responses are affected through the individual interviewers' behaviour and performance pattern during the interview. Different interviewers have different behaviour patterns and they ask the questions in their own style and pace and the question wording might not always be exactly as in the questionnaire. The interviewer effect is strongest particularly in face-to-face interviews and especially on sensitive issues where the interviewer's influence can lead to so called *social desirability bias* (e.g., Biemer and Lyberg, 2003). Social desirability bias is probably more common in household surveys, but can occur in business surveys too depending on industry covered and topic of the survey. For instance, businesses within an industry known for air pollution might report strategic or "brushed up" figures when it comes to environmental investments in cleaning technology or environmental protection with the intention to make them look better in public.

2.2.8   Telephone interviewing (CATI)

Telephone interviewing is the fastest data collection mode to implement from start to completion of data collection and is often used in combination with other modes in mixed-mode surveys (Biemer and Lyberg, 2003). The mode is not so common in business surveys in the data collection phase, but rather when it comes to call-backs, the editing phase when trying to fill out missing values or to reduce nonresponse. However, the mode is still used in business surveys in some countries, e.g., in agricultural surveys, and therefore it is included in this review.

By building common survey procedures directly into CATI systems, or into pre-packed setup modules, surveys with similar designs can be conducted more efficiently, even by staff with limited survey experience (Groves et al., 1988). The telephone interview shares some of the advantages and disadvantages with the face-to-face mode concerning the interviewer presence, as well as some of the advantages of the electronic questionnaires mentioned before. The interviewer effects and risks of social desirability bias are however lower than in face-to-face interviews. A disadvantage with telephone interviews is they are less flexible. Visual aids cannot be used, and neither the survey topic nor the survey questions (or the response categories) can be too many or too complicated in a telephone interview situation (e.g., Biemer et al., 1991). With too many response categories the respondent might forget and systematically pick the last response category read; something called *recency effect* (e.g., Biemer et al., 1991). The respondent might also interrupt the interviewer after the first response category has been read and say "yes" to that one, not letting the interviewer finish the job with reading the other response alternatives. This phenomenon is often referred to as *top-of-the-head-responses*. Top-of-the-head responses occur in all modes, but are perhaps more frequent in telephone interviews (see Biemer and Lyberg, 2003). The influence on data caused by recency effects and top-of-the-head-responses can be diminished by some programming if questions and response categories are allowed to be randomised within the questionnaire. If the survey questions require some

extra effort from the respondent like a check-up in computer systems or calculations, both face-to-face interviews and telephone interviews are less suitable modes. A growing problem in general with telephone interviews is that parts of the population may be difficult to reach since they are not listed in the telephone book (e.g., Biemer and Lyberg, 2003). This phenomenon is increasing as more and more people use only their mobile telephones and do not have fixed land line (e.g., Lepkowski et al., 2008). Naturally, finding the telephone number of the business is in general not a huge problem when conducting business surveys; the issue lies more within finding the right person within the business.

2.2.9 Direct observation

Direct observation in the field means that data are collected without direct involvement of a particular respondent; the observer assumes the role of the respondent (Biemer and Lyberg, 2003). The mode is often used in biology and qualitative research (de Leeuw, Hox and Dillman, 2008) but can also be used in data collections in business surveys for official statistics. An example of direct observation is when the goal is to estimate the proportion of trucks in traffic on a ring-road around a city, where observers register the number of trucks travelling at a random place during a randomly selected time period. Measurement errors for this mode may be introduced by the recording of observations by the observers in ways similar to the errors introduced by interviewers. The measurement errors may also relate to the instrument or device used to gather information. Large scale data collections using direct observation as mode are found in most agricultural surveys (Biemer et al., 1991).

2.3 How to mix modes

With all these mode possibilities there is a good opportunity to combine the strong points of each mode offering the respondent several modes for reporting. Such mode decision has to be planned carefully because it implies a more complicated, more expensive, longer and probably more challenging survey implementation. The usual goal is to find an optimal mix for data collection given the research question and the population under study given the restrictions (Biemer and Lyberg, 2003). The reason for mixing modes might be to collect follow-up panel data from the same respondent at a later time, but also to collect data from same respondents during a single data collection period. Mixed mode can be carried out to meet the respondents' preferences, but usually the main reason for mixed mode surveys is to battle the nonresponse. The general idea of mixed mode is to start with one main mode and when all possibilities are emptied for that mode a switch to another often more expensive mode is made, and so on. Allowing mixed modes or letting for instance businesses completely choose and define the agenda how they want to report might not be the best approach in the long run. Different modes have different ways of contacting the respondent which affects the answers, something that might cause problems in comparative surveys if the instruments are not well designed. It might also be costly to develop and maintain the data collection systems for each mode (e.g., de Leeuw, Hox and Dillman, 2008). A more detailed description on mixed modes can be found in the module "Data Collection – Mixed Mode Data Collection".

3. Design issues

4. Available software tools

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Allen, J. and Junker, C. (2008), How far can IT standards and tools help to reduce response burden? Paper presented at the 94th DGINS conference, Vilnius, 25th-26th of September 2008.

Beukenhorst, D. and Giesen, D. (2010), Internet Use for Data Collection at Statistics Netherlands. Paper presented at the 2nd International Workshop on Internet Survey Methods, Statistics Korea, Daejeon, South Korea, September 8 & 9, 2010.

Biemer, P. P. et al. (1991), *Measurement Error in Surveys*. Wiley Series in Probability and Mathematical Statistics, New York.

Biemer, P. P. and Lyberg, L. E. (2003), *Introduction to Survey Quality*. Wiley Series in Survey Methodology, New Jersey.

Cox, B. G. et al. (1995), *Business Survey Methods*. Wiley Series in Probability and Mathematical Statistics, New York.

e.Statistik.core (2008), Neue Wege zur Entlastung der Unternehmen. Statistische Bundesamt, Wiesbaden. (www.statistik-portal.de)

de Leeuw, E. D., Hox, J. J., and Dillman, D. A. (2008), *International Handbook of Survey Methodology*. European Association of Methodology, Lawrence Erlbaum Associates, New York.

Giesen, D. (ed.) (2011), Response Burden in Official Business Surveys: Measurement and Reduction Practices of National Statistical Institutes. BLUE-Enterprise and Trade Statistics, Small or medium-scale focused research project. (http://www.blue-ets.istat.it/)

Giesen, D., Morren, M., and Snijkers, G. (2009), The effect of survey redesign on response burden: and evaluation of the redesign of the SBS questionnaires. Draft paper presented at the 3rd European Survey Research Association Conference 2009, Warsaw, June 29-July 3 2009.

Goddeeris, O. and Bruynooghe, K. (2011), Administrative Simplification of the Structural Business Statistics. In: D. Giesen and M. Bavdaž (eds.), *Proceedings of the BLUE-ETS Conference on Burden and Motivation in Official Business Surveys*, Statistics Netherlands, Heerlen.

Gravem, D., Haraldsen, G., and Löfgren, T. (2011), Response Burden Trends and Consequences. In: D. Giesen and M. Bavdaž (eds.), *Proceedings of the BLUE-ETS Conference on Burden and Motivation in Official Business Surveys*, Statistics Netherlands, Heerlen.

Groves, R. M. et al. (2004), *Survey Methodology*. Wiley Series in Survey Methodology, New Jersey.

Groves, R. M. et al. (1988), *Telephone Survey Methodology*. Wiley Series in Probability and Mathematical Statistics, New York.

Haraldsen, G. et al. (2011), Utilizing Web Technology in Business Data Collection: Some Norwegian, Dutch and Danish Experiences. Paper presented at the New Techniques and Technologies for Statistics (NTTS) Conference, 22-24 February 2011, Brussels, Belgium.

INE (2008), The response burden in business statistics - The Spanish experience. Paper presented at the 94th DGINS conference, Vilnius, 25th-26th September 2008.

Lapeniene, V. (2008), Reduction of Data Collection Burden. Paper presented at the 17th Statistical Days, Radenci (Slovenia), 5-7 November 2007.

Lepkowski, J. M. et al. (2008), *Advances in Telephone Survey Methodology*. Wiley Series in Survey Methodology, New Jersey.

Orjala, H. (2010), Reducing the administrative burden in official statistics – Enterprise respondents in focus. Paper presented at the SIMPLY 2010 conference, Ghent 2nd-3rd, December 2010.

Pereira, H. J. (2011), Simplified Business Information (IES) – Is coordination between public entities really possible? In: D. Giesen and M. Bavdaž (eds.), *Proceedings of the BLUE-ETS Conference on Burden and Motivation in Official Business Surveys*, Statistics Netherlands, Heerlen.

Raymond-Blaess, V. (2011), Overview of measures used by NSIs to reduce response burden as reported in the literature between 2005 and 2010. In: D. Giesen and V. Raymond-Blaess (eds.), *Response burden measurement and reduction in official business statistics. A literature review of national statistical institutes' practices and experiences*, BLUE-ETS project, deliverable 2.1, 27–42.

Savolainen, A. and Vertanen, V. (2007), Statistics Finland's measures to reduce enterprises' response burden. Paper presented at the Seminar, NordStat 2007, Reykjavik, June 2007.

Snijkers, G., Onat, E., and Vis-Visschers, R. (2007), The Annual Structural Business Survey: Developing and Testing an Electronic Form. *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association, Alexandria, VA, 456–463.

Interconnections with other modules

8.   **Related themes described in other modules**

1. Questionnaire Design – Main Module

2. Data Collection – Mixed Mode Data Collection

3. Data Collection – Techniques and Tools

4. Response – Response Process

9.   **Methods explicitly referred to in this module**

1.

10.   **Mathematical techniques explicitly referred to in this module**

1.

11.   **GSBPM phases explicitly referred to in this module**

1.

12.   **Tools explicitly referred to in this module**

1.

13.   **Process steps explicitly referred to in this module**

1.

Administrative section

14. Module code

Data Collection-T-Design of Data Collection (Part 1)

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	25-01-2012	first version	Tora Löfgren	Statistics Norway
0.2	24-04-2012	second version – changes, added text and additional references according to review	Tora Löfgren	Statistics Norway
0.3	19-05-2012	third version – with some minor changes	Tora Löfgren	Statistics Norway
0.4	04-07-2012	fourth version with some minor changes according to review	Tora Löfgren	Statistics Norway
0.5	06-07-2012	last version	Tora Löfgren	Statistics Norway
0.6	21-11-2013	revised after EB comments	Tora Löfgren	Statistics Norway
0.6.1	21-11-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:48