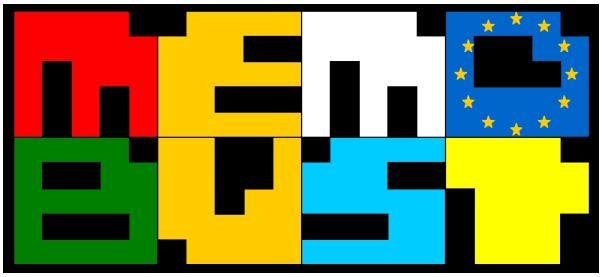

This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Repeated Surveys

Contents

General section

1. Summary

A repeated survey is a survey carried out more than once, mostly with regular frequency, for example monthly, quarterly, or annually. Most surveys in a statistical office are repeated. Samples in a repeated survey may be independent over time, or the sample design may deliberately involve a unit at several occasions. The sample design should balance accuracy requests, which often imply considerable overlap between samples over time, and response burden. A sample design, where a business is selected each time during a period and then is not selected for a time period, has advantages for both accuracy and the respondent. Panel surveys and longitudinal surveys are particular cases of repeated surveys. There are other arrangements, for instance based on permanent random numbers. A repeated survey may use administrative data, either only or in combination with directly collected data.

Measures of change are normally an important part of the statistical output of a repeated survey, for example indices and in many cases also time series. Seasonal adjustment may be used for short-term statistics to make comparisons easier for the users. Usually there are time-related requests on the output, such as comparability over time and high accuracy in estimates of change. The changes over time are due both to population changes and changes in values of variables. The requests have implications for the survey design. Differences in definitions and methods between two points in time mostly have a negative effect on the comparability between the two sets of statistics. Considering comparability over time only, such differences should be avoided. It is in the nature of a repeated survey to use the same definitions, methods etc.

A break in the time series may become unavoidable for external or internal reasons, and it can be justified when the advantages outweigh the disadvantages. It is important to measure its size, if possible, and to inform the users in advance about the introduced changes and the break. When statistics from repeated surveys are published it is often the case that the statistics for one or more of the earlier time periods are revised. It is recommended to have a revision policy, preferably aligned with other statistics, both nationally and internationally.

The repetitive character of the survey gives possibilities to improve the statistical production process and the quality of the output by utilising both previously collected data and process data (paradata). These possibilities should be taken into account before the first production round and incorporated in the design to ensure that appropriate data, paradata, and metadata are collected and saved for future use. An imbedded experiment is an example of a method to study effects of a suggested change in advance and possibly avoid time series breaks or at least reduce the effects.

There are three major reasons for a separate description in the handbook of repeated surveys: the possibilities to make improvements over time, the possibilities to utilise previous data if a unit is included repeatedly in the survey, and issues related to time series breaks. This topic provides an overview of the specifics of repeated surveys. Most methods are already described in other parts of the handbook, for example methods for sampling and estimation. References are given to relevant modules and also to the general literature, mainly on specific issues. There are few books dedicated to repeated business surveys, perhaps a bit surprisingly, since business surveys often are regularly repeated surveys. The overview edited by Cox et al. (1995) has good coverage. Snijkers, Haraldsen, Jones, and Willimack (2013) describe how to design and conduct business surveys; a recent book.

2. General description

2.1 Introduction

Business statistics are largely based on regularly repeated surveys. One reason is that measures of change are important, often more important than measures of level. Comparability over time, which then is an essential property, depends highly on concepts being the same. Stability of methodology is also essential – methodology needs to be unchanged if it has systematic influence on the output. These two facts are important to consider and handle when working with a repeated survey. It may be desirable or necessary to change concepts or methodology in order to improve quality (mostly the contents or the accuracy) or to reduce costs. Some resulting time series issues are mentioned here, and they are further discussed below in Section 2.5. Effects of the considered changes on the statistics should be measured, if possible, to avoid or reduce a break in the time series. It may be possible to adjust or extrapolate a time series forwards or backwards, depending on the knowledge about the break and an assessed likely behaviour over time. It is also important to inform the users. See for instance OECD (2007), a handbook on presentation.

A repeated survey can have samples that are independent over time or samples that are deliberately overlapping. Here focus is on the latter type. There are several advantages of including the same statistical unit in several rounds of the survey. Accuracy in measures of change is usually improved – but the sampling design needs to balance between accuracy gains and response burden. When a statistical unit has provided survey data, these data can be used in later rounds of the repeated survey, for example in editing or by providing them to the respondent as a support in a later round. In the latter case it sometimes happens that a statistical unit corrects earlier data, for example when seeing the previous data or realising for other reasons that it has provided erroneous data to the survey. This may occur for instance when there is a new respondent in the business, looking with new eyes. These new data (correction on micro level) can be used to revise the earlier published statistics (macrodata) later on; or even to correct the statistics outside the regular publishing scheme, if needed. Units that are outliers – values are correct but extreme and influential in regular estimation – should be studied to understand reasons; such knowledge may be used to improve size measures and estimation procedures in forthcoming rounds of the survey. It is typical for business statistics that the population changes quickly due to births, deaths, splits, and other re-organisations. It is important to have a frame that is regularly updated and gives access to the survey population, directly or in several steps. Some care is needed, though, when using previous information from the sample in the sampling frame, in order not to introduce bias; see further Section 3.3 below.

Experiences and more formal evaluations can be used to draw conclusions from earlier to later production rounds, thus improving the cost-efficiency of the survey. If the possible consequences of a methodological change are not known it may be useful and relatively cheap to use an embedded experiment, where the sample is divided into two or more groups, to compare the different methodologies (similar to study and control groups in other experiments).

A repeated survey may be based on direct data collection or on other types of data sources, such as administrative data, or possibly a combination. Some parts of this module are relevant only for surveys with direct data collection; surveys which often are sample surveys. This is particularly the case for sample co-ordination, providing previous values to the respondents, estimation with weights, outliers, and variance estimation. Other parts are relevant for all types of surveys. This is the case for updates

of frames and other information on populations, comparability over time, time series breaks, evaluation, and successive improvements. Use of administrative data means work with administrative and statistical units, populations, variable definitions, possibly models to make "transformations", estimation models etc. See, for instance, the handbook modules "Data Collection – Collection and Use of Secondary Data" and "Overall Design – Overall Design".

Most of the methods that are used for repeated surveys are described fully or partly in other topics of the handbook. The sub-sections 2.2–2.6 below have the headings Frames and sampling, Data collection and data processing, Estimation, Time series issues, and finally Tests, experiments, and evaluation. Even if the phases of the Generic Statistical Business Process Model (GSBPM) are not mentioned as such, most of them are discussed here: Specify needs, Design, Build, Collect, Process, Analyse, Disseminate, and Evaluate.

2.2 Frames and sampling

In repeated surveys considerable attention should be given to frame construction, sample design, and estimation. Some general remarks are given in Sections 2.2 and 2.4, and Section 3 describes design issues in some more detail. With administrative data the main issue in this context is to be careful with updates and reference times for population information.

When estimating change over time a large overlap of samples between occasions is desirable. Large businesses will be selected with probability 1. Small businesses may, for instance, be in the sample for one, a few, or some years. It is easier both for the statistical office and the respondent if there are longer periods of being inside and outside the sample, rather than frequently jumping in and out. This is part of the sampling design, as described below in Section 3.3. Often some sample co-ordination is used to increase or decrease the overlap of samples in comparison with independent samples. Simply expressed, negative sample co-ordination between two surveys means that samples for these have as few businesses in common as possible or reasonable. Conversely positive sample co-ordination between two surveys means that these surveys have as many businesses in common as possible or reasonable. Similarly, positive co-ordination over time for a survey means a high overlap between the samples of the two periods. As stated previously, accuracy and response burden are the two main reasons for such positive or negative co-ordination of samples.

Since business populations usually change rapidly, frames and samples need to be updated with some frequency, rather than retaining the same sample for a long period of time. The frequencies of updating need to be decided. The frame and the sample could be updated at the same times. Alternatively, the frame could be updated more often, with just small changes of the sample. For instance a sample corresponding to new parts of the population could be drawn, as a complement. There is a balance between improved information and the work with new frames and samples (also for the affected respondents). See further Section 3.3 below. Another way of handling coverage problems due to using 'old' samples is in the estimation; see Section 2.4 below and the overview in the handbook module "Weighting and Estimation – Main Module".

Maintaining the business register is essential to capture births, deaths, and other changes of the units before sampling to avoid problems in estimation. Updating registers and frames with data from samples in repeated surveys is, however, not straightforward. In repeated surveys in which there is a controlled overlap of samples between occasions, survey feedback can lead to bias in estimates, see

Section 3.2. The problem is discussed also in the handbook module "Sample Selection – Sample Co-ordination". See Cox et al. (1995), too.

2.3 Data collection and data processing

One of the characteristics of a repeated survey is that many statistical units are included several times, often every month or quarter during a period of one or several years. Values for one or several previous periods are available when the unit has responded, and that information can be utilised in several ways. Such data, with measurements of the same variable at several times, are sometimes called longitudinal data.

Some particular comments for surveys with administrative data follow. There are often large amounts of data, and re-contacts are in general not possible. See the specific editing module "Statistical Data Editing – Editing Administrative Data". The comments below on possible corrections are valid but probably not very frequent.

2.3.1 Providing previous data

A provided value may be printed or filled in when sending out the next questionnaire (paper or other mode). This gives support to the respondent, who fills in the questionnaire, and it reduces the response burden. Also, if a mistake was made, the respondent has an opportunity to correct the erroneous figure. On the other hand – if there was misinterpretation of which information to provide – showing these data may conserve this misinterpretation.

The layout of a printed or electronic questionnaire can take the form of two parallel columns beside the question. One column refers to a previous period and is pre-filled with the values from that period. The other column refers to the current period, where the information is to be filled in by the respondent. The respondent may be asked to correct the previous value in case it is in error. Such an error may be due to an earlier mistake of the respondent when providing the information, an update of that information, or a mistake when registering the information (for instance scanning).

The situation is similar when interviewing the respondent instead of using a questionnaire. When previous information is utilised in the questions of the interview, it is called "dependent interviewing". The previous information can be used in different ways, such as using a value without re-asking, probing in case of an unexpected change, verifying in case of unlikely combinations, and checking the time of a rare event. The advantages and disadvantages of dependent interviewing are similar to those of making previous information visible in questionnaires.

Holmberg (2004) carried out an experimental study, one of the few in business statistics, for a survey with pre-printed self-administered questionnaires for an establishment population. He lists several reasons for the use of such questionnaires: respondent support (reducing the response burden, questionnaire guidance, memory support, anchoring, and feedback purposes), improved efficiency in the data collection, and reduction of measurement errors and improved data quality. He also gives reasons against pre-printing: risk of bias due to underreporting of changes and conservation of errors, loss of confidence and goodwill (if the pre-printed data are of poor quality or if they are not recognised by the respondent), and disclosure risk. He used three treatments in the experiment: no pre-printing, pre-printing for one period, and pre-printing for two time periods. The survey that was studied was complex and asked for data during fourteen months. He considered the effects of pre-printing on

various aspects: response variability, presence and size of outliers, effects on other months, and experiences in general. In this study there were advantages: fewer and smaller problems with outliers and less spurious variation. Care has to be taken before generalising from one survey to another, though.

Overall, from the relatively few reported studies, it seems as if the advantages for preprinting are stronger than the disadvantages. When planning or using previous data, different factors behind the advantages and disadvantages should be noted and taken care of, as far as possible. Morrison (2009) provides some guidelines and further references on questionnaire design. Snijkers et al. (2013) provide broad information on questionnaire design.

2.3.2 Corrections of previous microdata – and macrodata

Corrections from the respondent of previous data mean that the statistical office can improve the output of earlier periods. For statistics with regular revisions this is a natural action in the next publication round. If the statistics are already final when additional corrections are made on the micro level, the effects on the estimates have to be considered. It may happen that the effect is great enough to motivate an unplanned correction of the statistics published. Specifically, two aspects are the size of the effect(s) and the status of the statistics, preliminary or final.

The size of the effect should be considered together with several factors, such as the accuracy and the aggregation level of the statistics. This leads to the following cases:

- If the size is considerable, a correction may be necessary.
- If the statistics are preliminary and the next planned revision date is close, it may be better to wait until this regular revision, even if the correction size is not very small.
- If the size is small and the statistics are preliminary, the correction and its effects simply go into the next revision.
- If the size is small in comparison with the accuracy of the statistics and the statistics are final, it is probably better to refrain from correction.

2.3.3 Using previous values in editing and imputation

Previous values are useful for both editing and imputation. This – longitudinal data – is an inherent strength of repeated surveys with some overlap between survey rounds in comparison with surveys that are made just once.

In editing a comparison with a previous period may show a value to be spurious due to an unrealistic change – and thus lead to the detection of an error that had otherwise gone unnoticed. It may also happen that the previous value was wrong, as discussed above. A re-contact may be motivated in case of a surprising change. Perhaps, there has been a mistake or, as a possible alternative, a change in organisation of the enterprise. See further the handbook module “Statistical Data Editing – Editing for Longitudinal Data”.

A further use of previous values is in imputation. If the unit is late in responding this period it may be reasonable to utilise the value of one or more previous periods and bring them forward in time, with regard taken also to possible seasonality – unless there are signals about the unit undergoing some

change. It may be reasonable to assume that a group of similar units (for example in a stratum) have similar sizes of change since that period. This method, which uses previous information of the unit, may be better than using only information from a group of similar units in the current period. See further the handbook module "Imputation – Imputation for Longitudinal Data".

2.4 Estimation

This section is mainly relevant for sample surveys with direct data collection. When administrative data are used, estimation issues will involve coverage deficiencies and possibly late data, especially for small enterprises. See for instance the handbook module "Weighting and Estimation – Estimation with Administrative data", which has a description and provides references to the ESSnet project AdminData with many relevant deliverables.

Data from previous runs of the survey are available in repeated surveys, at least to some extent. This means that further and more advanced estimation methods are available. For instance, there are different imputation methods based on previous values, as mentioned above.

The changes over time for a variable, like turnover, have two basic causes: the target population changes due to births, deaths, splits etc. and the "stable" units change in size resulting in larger or smaller variable values. The estimation has to take into account these causes based on the sample design, the frame, sample information, and possibly auxiliary information. The frequency of new register and frame information has an influence on the estimation procedure, and so has the frequency of new or partly renewed samples.

In short-term statistics, using a frame with information that is already somewhat old due to reporting delays and a sample for a longer period leads to coverage problems and problems with businesses that merge or split. These problems are, of course, present in all business surveys, but in short-term surveys with sample selection perhaps once or a few times a year they grow and become more serious at the end of the period that the sample is used. Mergers and splits mean that the units are not the same as they were at the sampling occasion. Sometimes data on the sampled units can still be collected, in case the data providers can report values for the old units. If that is not possible the problem with unit changes has to be handled in the estimation. There are several principal and practical issues to consider in the estimation (not only for repeated surveys), see the handbook module "Weighting and Estimation – Main Module".

Coverage problems may be handled if there is an updated register or frame, which can be used to "adjust" the estimates to correspond to the updated information. One or more auxiliary variables, which are known on both sample and population levels, are used for such a calibration; see the handbook methodological module "Weighting and Estimation – Calibration". There may be some practical issues with merging, comparing and handling large businesses.

Problems with outliers are common in business surveys of economic variables, as indicated above. Usually there is stratification by some size measure or a sampling design using a probability proportional to size. However, since the size measure is not always up to date or the size measure is not perfectly correlated with the study variable(s), it is difficult to order the businesses by size. For some study variables this is more difficult than for others, for instance an investment variable with a skewed distribution. There is always a risk of observations with a high influence on the estimates, because the size measure in the sampling design is not up-to-date or the variable has a highly skewed

distribution. Some type of outlier treatment must be in place in the estimation and included already in the design. See Section 3.2 and the handbook module "Weighting and Estimation – Outlier Treatment (Robust Estimation)".

When samples are co-ordinated over time variance estimation for measures of change is not as straightforward as variance estimation for parameter estimates based on one sample. Some methods are in place; see for instance Nordberg (2000), who suggests a variance estimator when samples are co-ordinated by using permanent random numbers, and Knottnerus and Van Delden (2012) for a more general setting. This is discussed also in the handbook module "Sample Selection – Sample Co-ordination".

The population parameters to be estimated may be indices, both price indices and volume indices of different types. The choices should be made from the start in agreement with important users.

2.5 Time series issues

An important reason for repeated surveys is to measure population changes over time. Such interests from users have implications for both production and dissemination. The quality component comparability over time is important, but it has, of course, to be balanced with other quality components. A change may occur, for instance by introducing new methodology for accuracy reasons, even if such a change implies a break in the time series. Alternatively, a break may be caused by external changes, for instance new tax rules. Use of new concepts in the survey may be justified because they describe the current situation better than the old concepts. Hence contents and relevance may motivate a disturbance to comparability over time. This section is relevant for all types of surveys.

2.5.1 Comparability over time

There is always inaccuracy in statistics: random variation and systematic variation. The random variation makes comparisons less "sharp" or conclusive, but the comparisons are still meaningful with just random variation. However, if there is some systematic deviation, this normally disturbs or destroys the comparability. Examples where such possibly systematic influence is introduced are:

- A different way of updating the business register, for example a new source or a different time schedule. This influences for instance the frame coverage and the accuracy in classifications.
- A new data collection mode influences possibly, but not necessarily, the data that the respondent provides.
- Changes in the editing procedure influence possibly, but not necessarily, the output in a systematic way.

This is simply expressed as follows. With "everything unchanged" in production, response processes, and the society context, comparability over time is not affected. Otherwise possible systematic influence has to be investigated.

Some systematic errors may have different effects on estimators of levels and estimators of change measures. For example coverage deficiencies mainly affect the level of estimation. However, care has to be taken also for estimators of change measures with regard to over-coverage and under-coverage. Such coverage deficiencies have different sizes over time for instance between different parts of a

business cycle (up-and-down movements in economic activity). Hence, a simple assumption in the estimation procedure of over-coverage and over-coverage being equal may work when the economy is stable but be quite misleading in times of change.

Comparability over time means that a presentation of the time series in a graph or in a table is meaningful. There may still be calendar and seasonal effects. Removing such effects makes the study over time easier. Sometimes just simple comparisons with the corresponding period in the previous year are made.

Comparability defects, if any, should be clearly stated and explained in all presentation modes. There are two major situations where breaks occur. The causes indicate at least partly which methods to use in order to handle the problems that occur. Such different methods are discussed in the next two sections.

2.5.2 Methods to overcome breaks, for instance caused by redesign

There are typically two time series, the old and the new one. They cover different time periods, possibly with a "double" period. The two series show "the same thing", but there is a difference, which for instance is caused by using another method. Hence, there is a "jump".

One possibility to quantify the effect of a change without a full implementation is to make an experiment, as described in Section 2.6.2 below. At best, use of experiments can reduce, or even avoid, time series breaks. If the change is introduced without a controlled experiment, there should be a double period in order to have some possibility to estimate the effects of the change.

A further possible situation is to have two separate time series that describe the same or nearly identical phenomena but without a clear connection between the series. This may for instance be the case when the National Accounts replace one source (the old one) with another source (the new one). There may be one or a few time periods where both sources exist. There is often a "gap" between the two time series. De la Fuente (2009) describes a set of methods that is more flexible than a simple "vertical movement" of the old series. An error term is introduced to describe the difference between the two time series, and some appropriate assumption is made about this error; this assumption may not be obvious. Then the adjustment is derived.

Van den Brakel and Roels (2010) describe an approach where state-space models and intervention analysis are used to estimate the discontinuities, typically caused by survey redesign. Theory and illustrations are provided.

When several related time series are involved, they have to be considered together so that consistency is preserved, for example so that parts of a sum add up to the total.

2.5.3 Revision of a classification

Much methodological and practical work has been devoted to the situation where a classification is revised, for instance a new version of NACE. The revised classification is introduced because it is now a better description of society than the old classification. The users like long time series and often ask for "back-casting". This may be reasonable for a high aggregation level and for a moderate time period. Care has to be taken in computations, presentations, and interpretations of the back-casted series. Perhaps some industries in the revised classification did not exist ten years back in time.

There is a user need to extend the time series, especially backwards, in spite of the break. The relationship between the old and the new versions of the classification is complex in many cases. The Business Register is normally double-coded, that is coded according to both classification versions, usually for a brief period of time.

There are two major approaches to extend time series: the micro and the macro approach. They are briefly described below.

The micro approach is based on business units. The new coding is extended backwards for units in the survey, and estimates are made for these new domains of estimation. These estimates will be somewhat inaccurate. There is uncertainty and difficulty with coding backwards. Some units no longer exist when the double-coding is made, and they have to be coded with limited knowledge. Moreover, the survey was designed for the old classification. There may be relatively few units in some domains of estimation.

The macro approach builds on relationships between the two versions of classification. The double-coded business register is an important source of knowledge about the relationship in that period. Choosing some appropriate level of detail, a cross-classification can be made. Imagine a matrix showing the industries in the old system row-wise and the industries in the new system column-wise. Each cell shows some quantity, like the number of employees or the turnover. Each row will show the "flow" from the old version to the new version. Each column will similarly show, for each new industry on that level, which old industries contributed and to what extent.

In the macro approach such a matrix is used for "conversion" of the time series in the old classification version to time series in the new classification version. The computations are fairly easy as such for many surveys. The difficulties lie in choosing appropriate information for the conversion matrix and finding appropriate levels of detail.

There are many papers, both from national statistical offices and from Eurostat as well as other international organisations. Only two references are given here.

Brunauer and Haitzmann (2010) describe a case at Statistics Austria for the Structural Business Statistics. They compare the micro and the macro approaches. The former may seem more accurate at a first glance, but it is difficult to code statistical units backwards in time and to code units no longer existing. It is also resource-consuming.

Van den Brakel (2010) describes several aspects, especially sampling and estimation techniques with regard to the two classifications and back-casting procedures. Both the micro and the macro approaches are included. Conversion factors are discussed in some detail. Time dependency and indices are included. This is a methodological paper.

2.6 Tests, experiments, and evaluation

Embedded experiments and pilot studies mainly refer to surveys with direct data collection. Studies and evaluations are relevant for all types of surveys.

2.6.1 Tests and experiments

Testing is mostly an investment that prevents future work with corrections and other problems. Some types of mistakes mean that one or more processes have to be run again. In some cases it is not

possible to re-run the process, though, in spite of the mistake. This is not specific for repeated surveys, but a general observation.

In a repeated survey much is the same from round to round. A new reference period may require new settings in the production system, by parameters (preferably) or manual settings. Changes may be needed for data set names etc. Every change means a potential error. Changes should be tested early.

A main goal of a repeated survey is to measure changes in population parameters, so comparability over time is mostly important. Changes in production may be motivated for a number of reasons, for example indications about desirable or possible improvements from earlier production rounds or other internal changes in the production environment. Cost-effectiveness is always essential, and small improvements may mean considerable savings in the long run for a survey that is conducted many times.

However, consequences of changes should be foreseen before embarking on them. A pilot study is always a possibility to make investigations on a small scale. It may be a good and cheap way to discover weak points and unexpected implications. Repeated surveys provide a setting that enables more than separate pilot studies, because experiments can be made within the survey setting: so-called embedded experiments.

In an embedded experiment the sample is randomly divided into two or more subsamples according to an experimental design. There is normally a control group and one or more treatment groups (subsamples). The treatment(s) may be, for example, new advance letters, new data collection modes, or new contact strategies. Care has to be taken in planning in order to avoid confounding with other influencing factors and also to make sure that the experiment is feasible in practice. Moreover, the hypotheses should be formulated in advance, and the powers of the tests should be estimated at the same time to understand which differences can be considered as significant. Otherwise the experiment may turn out to be an inconclusive disappointment.

A methodological description of how to make an embedded experiment is given by Van den Brakel and Renssen (2005), where also further references can be found. Another description with both an experiment and a time series perspective is given by Van den Brakel, Smith, and Compton (2008). They describe quality procedures for survey transitions; see also Section 2.5 above. A practical study dealing with response burden is presented by Hedlin, Lindkvist, Bäckström, and Erikson (2008). A practical study about pre-printing effects is presented by Holmberg (2004); see also Section 2.3 above.

2.6.2 Evaluation

An evaluation should always be made after a production round to learn for the future, considering both this survey (if repeated, as described here) and other related surveys. The ambition level should vary with regard to survey frequency and findings. In short-term statistics it is natural to catch the most urgent matters after each single round. All findings can be summarised in more detail after a longer period and then acted upon. The regular annual planning may be a suitable point in time to consider modifications and redesign.

An example of a qualitative type of evaluation is a brief summary of staff experiences made during the production. Debriefing with staff may be a good way both to collect findings and to get suggestions

for improvements. Staff working on editing is such a group where difficulties regarding data collection can be detected and summarised. These include observed difficulties, for instance in filled-in questionnaires and in re-contacts with respondents about spurious values. Suggestions for better wordings in questions and instructions may be obtained here, and also suggestions to improve the editing procedure.

Other evaluations are more quantitative. It is recommended to collect process data (paradata) as a basis for evaluations. Two important types are measures/indicators of quality and data about costs. A quality indicator may be useful for process quality or product quality or both. The goal of an evaluation is both quality assessment for this round and improvements in future rounds. The quality assessment is, of course, used in quality reporting. Measures of response rate over time provide an example of a quality indicator, where the conclusions drawn about the output quality have to be careful and restricted. Similarly, there are rates in editing that are indicators of the process, for instance how many "suspicious" cases turn out to be influential errors. This is not a good indicator of output quality, though.

Analysis of the response rate together with dates of reminders is one way to improve the strategy for contacting respondents: how often, in what way etc. There are several ways to improve the data collection. See Section 3 below for a brief description and several handbook modules: the module "Data Collection – Main Module", two modules on design of data collection and one on mixed mode.

There is also a handbook module devoted to the topic evaluation: "Evaluation – Evaluation of Business Statistics".

3. Design issues

3.1 Introductory remarks

There are two basic situations: a new repeated survey and an ongoing repeated survey. The latter situation is typical for statistical offices and their business statistics. The former situation is rarer, but sometimes a new repeated survey is launched. Several design issues are mentioned already in Section 2 in their respective contexts. Here in Section 3 the focus is on design, and especially frame, sampling, and estimation are discussed concentrating on repeated surveys. Most of the design issues are relevant for all types of surveys. Responsive design, sampling, and estimation from a sample are exceptions, useful only for surveys with direct data collection.

3.2 A new repeated survey – some issues

Even if the design of a repeated survey is in many respects similar to the design of a one-time survey, there are some additional issues and possibilities to consider. Many of these should be considered before the first round to achieve the potential gains from the start and to eliminate risks of undoing some work.

It is, as always, important to consider the user needs early. One of the special issues here is to discuss and verbalise the priorities among the accuracies of different estimators. One such example is how to balance between the accuracy of estimation of levels and the accuracy of estimation of change over time.

Coherence and comparability also need to be considered. These considerations may have implications for the choice of statistical unit(s), population delineation, variable definitions, reference times/periods, frames etc. They may also influence a possible co-ordination with other surveys. This is discussed in the next section.

The fact that the survey is made repeatedly means that previous data can be used in data collection, editing, and imputation as discussed above in Section 2. This should preferably be designed from start, since the production system needs to include these previous data and additional procedures. Editing and imputation procedures are, of course, needed also for units without previous data. Hence, several different procedures need to be included with priority rules.

Possibilities for continuous improvements should be built into the system, preferably from the beginning. This means both collecting appropriate paradata and using them. This is considered in Section 3.4 below and in the handbook module "Overall Design – Overall Design".

3.3 Frame, sampling, and estimation

There are some essential choices in the design. The description is simplified here, mostly because there is information in other topics and modules. Some essential decisions are included, which are typical for repeated surveys. They are, of course, not to be made one by one, but together.

One important decision is about the frequency of frame updates and of new (or renewed) samples. In annual surveys the procedure is usually to update the frame and select a sample once a year. In short term statistics the value of new information and the additional work with new samples have to be balanced. Since business populations change rapidly using a sample for a year (or many months) leads to coverage problems especially at the end of the period. Births, deaths, splits, and mergers are frequent and lead to population changes. Moreover, changes in size measure and economic activity, typically used for stratification, are common, too. If the business register and other sources of data, which are used for frame construction and sample selection, are updated more frequently than annually, a new frame can also be constructed more frequently than annually. See "Statistical Registers and Frames – Main Module" and its directions to further modules in that topic.

Selecting new samples often leads to increased workload in the data collection: contacting new enterprises etc. The first time a new sample is used, the non-response is sometimes larger, since some of the units are included in the survey for the first time. There may also be more problems with measurement errors, since new data providers are included more often. Furthermore, selecting samples more often also leads to increased burden for enterprises that are included and excluded from the survey more often. It may also be possible to continue with the previous sample but utilise the new frame information to handle movements into and out of target population. The sample may then be "complemented" with regard to units that belong to the target population according to the new but not the previous information. Conversely, parts that no longer belong to the target population are removed. Both theoretical and practical issues have to be considered. The sampling procedure and the estimation method with design weights must be known. Some further comments are given below.

Another important design decision is if and how to co-ordinate samples, both over time and with other surveys. The idea of co-ordination between samples over time is to obtain a 'large' overlap part of the samples between survey rounds. Large overlap between samples over time normally ensures smaller variance of estimated change. Parts of the sample should, however, be replaced to capture changes in

the population, for example to include new units, and to spread the response burden among the businesses. There may, for instance, be a system with permanent random numbers (PRNs), where:

- Positive co-ordination is used over time. This increases the accuracy of estimators of change. The respondent may find it relatively good to be "in" for a period and then "out" for period.
- Positive co-ordination is used between some related surveys. This makes comparisons easier on the micro level (co-editing) and on the macro level.
- Negative co-ordination is used between many non-related surveys, mainly to reduce response burden.

See the handbook module "Sample Selection – Sample Co-ordination" and several related method modules.

Stratification is usually done by domains, such as kind of activity, and size. The largest size classes, one or more, in each domain are usually totally enumerated while samples are selected from strata with small or medium-sized enterprises. In repeated surveys the choice of size stratification variable is maybe not the seemingly optimal one. A variable that is more stable over time can be preferable to one that leads to a more efficient design but which changes more rapidly. Using size in the design is, of course, important to avoid, as far as possible, problems with outliers.

A further important decision is about principles and handling of outliers: Some type of outlier treatment must be in place in the estimation and included already in the design. There are several methods to handle outliers in the estimation. Many methods modify the weight or the value. Often this means that the variance of the estimator is considerably reduced, but that a bias is introduced; together this normally means that the mean squared error is reduced. With a repeated survey data from previous runs of the survey can be used for identifying outliers, for outlier treatment in the estimation, and possibly for a better sampling design.

A further important estimation decision, which is related to updates of the frame and the sample, is about handling changes of units: Principles are needed for taking merging, splitting, deaths, changed industry etc. into the estimation. Some of these changes will be known in later versions of the register and frame, possibly at the time of estimation but possibly not until later. Hence, the changes may be known for the sample only. It is possible to work with the weights or the values of the statistical units. As mentioned above, there is also the decision how often to update the sample and in which way (fully or partly). See the handbook module "Weighting and Estimation – Main Module".

Another decision for estimation concerns the possible use of auxiliary information to improve accuracy, for instance with regard to non-response and coverage deficiencies. If the first production round (for preliminary statistics) is very early, specific estimation procedures may be needed, more model-based. If there is cut-off sampling, a model-based estimation is needed. See the handbook module "Weighting and Estimation – Main Module", where design is discussed.

Finally, in repeated surveys in which there is a controlled overlap of samples between occasions, survey feedback into the business register and the survey frames has to be considered. There is a dependency between samples. For instance, consider the case where information about deaths is quicker from the survey than from the regular sources. If such information is fed into the register and the new frame, the next sample that is drawn will have more updated information than the frame and population as a whole, due to the positive co-ordination with the previous and updating sample. With

standard estimation procedures, survey feedback can lead to bias in estimators. The problem with survey feedback is further discussed in the handbook module "Sample Selection – Sample Co-ordination". See also Cox et al. (1995). It is still important to use the updated information in communication with the respondents and to design the handling of the updated information in all surveys in a consistent way.

3.4 Improvements of repeated surveys

Improvements are based on evaluation, mostly of previous production rounds. Some typical warning signals are a high item non-response rate, a lot of editing work for some variables, and lower measures of accuracy than expected. Causes should be searched, for instance some small mistake in the questionnaire or a computer program. The estimation procedure should perhaps be improved, for instance through more auxiliary information. The working procedures should perhaps be adjusted by allocating resources differently. There are many possibilities for improvement. Warnings of quality deficiencies should be traced backwards. Such studies to learn and improve are relevant for editing, for instance, where an editing method, including parameters, can be studied with respect to numbers of suspicious cases, detected errors etc.

There is also a possibility of evaluation and actions within a production round. The design may, for instance, include points in time where the situation is considered, typically with respect to non-response. A method to adjust the data collection procedures (for instance the contact strategy based on achieved response rates) based on the data so far can be included in the design. Such a design is often called a responsive or adaptive design and should be planned in advance. See the handbook modules "Data Collection – Design of Data Collection Part 2: Contact Strategies" and "Overall Design – Overall Design".

The second aspect is the inclusion of measures or indicators of resources used, both in paradata and analyses for improvement. The aim is to try to increase the efficiency not only for single methods but also for the allocation of resources used in the production. See also the handbook module "Overall Design – Overall Design".

3.5 Redesign and other considerable changes of a repeated survey

There may be internal reasons to make considerable changes or to redesign a survey, for example due to integration with other surveys or new data collection possibilities. Then evaluations and possibly other information from the survey hitherto should be combined with user needs. Contacts with stakeholders and users about possible changes and about current needs and priorities may be wise. Embedded experiments may be valuable, as described above in Section 2.6, as a start of a redesign, to study possible, desirable, and non-desirable effects.

There are a few different situations with substantial changes. Time series breaks provide an example, for instance when there is a new classification or some other unavoidable external change. Switching from direct data collection to an administrative data source, partially or fully, may be such a substantial change, where output quality components have to be considered carefully. It is important to be pro-active as described above in Section 2.5.

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

Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S. (eds.) (1995), *Business Survey Methods*. John Wiley and Sons, New York.

Brunauer, M. and Haitzmann, M. (2010), *Backcasting Methods at Statistics Austria*. (Comparison of the Micro and Macro Approach on basis of the Structural Business Statistics 2005 to 2007. Results within the scope of NACE Revision 2.) Paper presented at the conference Q2010, Statistics Finland.

De la Fuente, A. (2009), *A mixed splicing procedure for economic time series*. Barcelona Economics Working Paper Series, Working Paper No. 415.

Hedlin, D., Lindkvist, H., Bäckström, H., and Erikson, J. (2008), An Experiment on Perceived Survey Response Burden among Businesses. *Journal of Official Statistics* **24**, 301–318.

Holmberg, A. (2004), Pre-printing Effects in Official Statistics: An Experimental Study. *Journal of Official Statistics* **20**, 341–355.

Knottnerus, P. and Van Delden, A.(2012), On variances of changes estimated from rotating panels and dynamic strata. *Survey Methodology* **38**, 43–52.

Morrison, R. L. (2009), Writing and Revising Questionnaire Design Guidelines. Proceeding of Statistics Canada Symposium 2008. Data Collection: Challenges, Achievements and New Directions.

Nordberg, L. (2000), On Variance Estimation for Measures of Change When Samples are Coordinated by the Use of Permanent Random Numbers. *Journal of Official Statistics* **16**, 363–378.

OECD (2007), Data and Metadata Reporting and Presentation Handbook. OECD Publishing. doi: 10.1787/9789264030336-en.

Snijkers, G., Haraldsen, G., Jones, J., and Willimack, D. K. (2013), *Designing and Conducting Business Surveys*. John Wiley and Sons, Inc.

Van den Brakel, J. A. (2010), Sampling and estimation techniques for the implementation of new classification systems: the change-over from NACE Rev. 1.1 to NACE Rev. 2 in business surveys. *Journal for Survey Research Methods* **4**, 103–119.

Van den Brakel, J. A. and Renssen, R. H. (2005), Analysis of experiments embedded in complex sampling designs. *Survey Methodology* **31**, 23–40.

Van den Brakel, J. A. and Roels, J. (2010), Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics* **4**, 1105–1138.

Van den Brakel, J. A., Smith, P. A., and Compton, S. (2008), Quality procedures for survey transitions, experiments, time series and discontinuities. *Journal for Survey Research Methods* **2**, 123–141.

Interconnections with other modules

8. Related themes described in other modules

1. Overall Design – Overall Design
2. Statistical Registers and Frames – Main Module
3. Sample Selection – Sample Co-ordination
4. Data Collection – Main Module
5. Data Collection – Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method
6. Data Collection – Design of Data Collection Part 2: Contact Strategies
7. Data Collection – Mixed Mode Data Collection
8. Data Collection – Collection and Use of Secondary Data
9. Statistical Data Editing – Editing Administrative Data
10. Statistical Data Editing – Editing for Longitudinal Data
11. Imputation – Imputation for Longitudinal Data
12. Weighting and Estimation – Main Module
13. Weighting and Estimation – Estimation with Administrative Data
14. Evaluation – Evaluation of Business Statistics

9. Methods explicitly referred to in this module

1. Weighting and Estimation – Calibration
2. Weighting and Estimation – Outlier Treatment

10. Mathematical techniques explicitly referred to in this module

1.

11. GSBPM phases explicitly referred to in this module

1. Phase 1. Specify Needs
2. Phase 2. Design
3. Phase 3. Build (especially regarding tests)
4. Phase 4. Collect
5. Phase 5. Process
6. Phase 6. Analyse
7. Phase 7. Disseminate

8. Phase 9. Evaluate

12. Tools explicitly referred to in this module

1.

13. Process steps explicitly referred to in this module

1.

Administrative section

14. Module code

Repeated Surveys-T-Repeated Surveys

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	10-03-2012	all modules together, with sections, from NL review	Eva Elvers and Tiina Orusild	Statistics Sweden
0.1.5	27-03-2012	from GR review	Ditto	Statistics Sweden
0.2	21-06-2012	from Sander, mainly	Ditto	Statistics Sweden
0.2.5	11-03-2013	from HU, templ. glossary	Eva Elvers	Statistics Sweden
0.3	24-05-2013	updates	Eva Elvers	Statistics Sweden
0.3.2	26-06-2013	NL review and glossary	Eva Elvers	Statistics Sweden
0.3.3	23-08-2013	a few adjustments	Eva Elvers	Statistics Sweden
0.3.5	06-11-2013	some clarifications	Eva Elvers	Statistics Sweden
0.3.6	15-11-2013	references, …	Eva Elvers	Statistics Sweden
0.3.7	22-11-2013	clarifications, NL review	Eva Elvers	Statistics Sweden
0.4	20-12-2013	after EB review; preliminary release	Eva Elvers	Statistics Sweden
0.4.1	08-01-2014	add admin data; harmon.	Eva Elvers	Statistics Sweden
0.4.2	14-01-2014	clarification for admin.	Eva Elvers	Statistics Sweden
0.5	10-02-2014	updates, glossary	Eva Elvers	Statistics Sweden
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 17:28