Guidelines for Survey Development

These guidelines provide advice on how to develop your collection instrument and methodology to help minimise respondent burden and generate fit-for-purpose results. More comprehensive information can be found in the ABS Form Design Manual, the NSS website and the SCH website.

Section 1. Setting up the survey

Survey managers should consider:
- the need for the proposed data; and
- that a survey of businesses is the most appropriate mechanism through which to collect the required data.

1.1 Identify Data Needs

The survey manager should consider the purpose of the survey, including:
- the survey's main objectives;
- the key outputs of the survey (which includes the estimates, commentary and recommendations);
- how the outputs will meet the survey objectives;
- identify any legal or administrative requirements that make the collection of this data critical;
- what level of detail do you require from the respondent (e.g. total employment Vs employment by full/part time status);
- the level you will be producing outputs (e.g. national, state, industry, sector);
- what level of quality is sufficient to meet the needs of users (covered in Sections 3 and 4).

1.2 Consult and Confirm Needs

A key part of the planning of a survey should be consultation with key stakeholders. This consultation will assist in gaining a thorough understanding of the users' needs regarding the data.

Key stakeholders, such as industry groups or other government agencies, may also be helpful in identifying existing data in their area, as well as related surveys that have previously been run. In addition, stakeholders may be able to identify whether it is actually possible to collect the required data, as well as identify any constraints on attaining the data.

As the survey objectives and proposed data items are developed, it's important to maintain regular contact with users to ensure the survey continues to meet their requirements.

Section 2. Respondent Burden

Respondent burden can be defined as the time, effort and cost required for respondents to complete a survey.

Survey managers should consider:
- *Actual burden*, which is quantifiable in terms of the frequency and length of the questionnaire and contact.
- *Perceived burden*, though not quantifiable, is how the requirements of a survey (such as the presentation of the survey instruments, the information provided in the approach letter and other provider information, etc.) are perceived by the respondent in terms of the level of (cognitive) effort required to participate in the survey.

2.1 Actual Burden

Actual burden can be measured as the frequency (e.g. monthly, quarterly, annual) and the length of time that the respondent is engaged with the survey, which can be further broken into:
- *contact time*, which is the number of businesses that the survey makes contact with, multiplied by the time taken to make the initial contact (usually to gauge whether the respondent is in-scope and willing to participate in the survey); and
- *responding time*, which is the number of businesses from which a response is actually received (including partial responses), multiplied by the time taken to respond to the survey. The responding time includes activities such as compiling and collecting records in order to be able to complete the survey.

2.2 Measures Taken to Reduce Respondent Burden

The Survey Manager needs to ensure that appropriate measures are taken to reduce respondent burden where possible. This includes determining that there are no alternative sources of information available and no reasonable alternative means of obtaining the required information with less respondent burden. Where possible, survey managers are encouraged to use existing data in order to minimise the amount of data that needs to be collected.

Another key mechanism for reducing the burden imposed on providers is controlling the overlap between different surveys and, where applicable, between different cycles of the same survey. This is usually only possible for surveys that share the same frame (frames are discussed in more detail in **Section 4.1 Population and Frame**), whereby selection in one survey (or one survey cycle) means that the same unit cannot be selected in another survey (or cycle).

Section 3. Data Collection Methods and the Collection Instrument

A well designed collection instrument is critical to ensuring that both the actual and perceived burden of a survey is minimised. Survey managers are encouraged to follow established questionnaire design principles and standards (for example, see the ABS Forms Design Standards Manual). Important things to consider include:

- an appropriate data collection mode is chosen;
- the questionnaire is structured in a logical fashion that is easily followed by respondents;
- the questionnaire contains all of the necessary instructions and clear definitions to reduce confusion;
- the surrounding infrastructure of the collection instrument enables ease of use; for example, for an online form it is crucial that respondents can easily access the form, and save and exit and then re-enter the form; and
- the collection instrument has been appropriately tested to ensure it will operate as expected for respondents.

3.1 Data Collection Strategy

Survey managers should consider a data collection strategy to plan the details of how the data is actually going to be collected. The strategy should be designed to maximise the response rate and achieve data that is fit-for-purpose.

The data collection strategy should include key details, such as:

- the data collection mode (covered in **Section 3.2**);
- the collection period, which should be long enough to provide respondents with sufficient time to gather relevant details and complete and submit the survey questionnaire;
- pre-approach materials that will alert providers to the upcoming survey;
- the contact strategy, including the number of attempted contacts that will be made, and follow-up procedure. The expected response rate will influence this contact strategy.

3.2 Collection Mode

Data collection mode refers to the method that is chosen to administer the survey (e.g. paper form; telephone interview; online form, etc.). Choosing the right mode to collect data can be fundamental to the success of the survey. The mode can have serious implications for the response rate and data quality for the survey.

Each mode has certain advantages and disadvantages that make them suitable for some contexts and not others. The choice of mode usually results from a process of weighing-up a number of issues relevant to the collection, in-light of the resources available.

There are a number of key issues to consider when determining whether the proposed data collection mode is suitable or not, including:

- the type of topic. Some modes are more suitable for particular topics than others. For example, surveys that require respondents to check their own records, or compile data from their systems are more suited to self-administered modes.

- the complexity of the topic and the size of the survey. The level of complexity of a survey, as well as the level of detail required will often dictate which mode is selected. For example, if a survey requires complex sequencing then an electronic format may be suitable, as it enables the sequencing to be automated. For brief surveys that require relatively straight-forward responses, a telephone interview may be a suitable option.

- respondent preferences and characteristics. Consider the types of respondents and any factors that may influence how the data collection mode fits with these respondents. Are you targeting a particular type of respondent for which one mode may be more suitable than others? Do your respondents have a particular preference for a mode? For example, electronic/online forms may be the most convenient format for some businesses to provide data, but not for those without easy access to the internet.

- the response rate. The choice of mode can have a significant impact on the achieved response rate. Face-to-face interviews tend to achieve higher response rates than other modes because respondents find it more difficult to refuse participation to a person's face compared to over the phone or a mail out/mail back survey.

- collection resources. The choice of mode affects the resources for data collection, follow-up and data processing. For example, data taken by a person, whether it's face-to-face or over the phone can be edited as it is being collected, whereas self-administered surveys will need more editing when it is returned. Some modes may typically achieve a higher level of quality than others, such as face-to-face interviews. However, they also cost a lot more to run. Therefore, it's important to weigh up what is achievable with the resources available and will still produce data that is fit-for-purpose.

3.3 Questionnaire Design

Questionnaires need to be designed in a manner that achieves the right balance between the level of burden imposed on respondents and the quality of the data they will produce.

Broadly speaking, questionnaire design should take into account:
- the type of data respondents are able to provide;
- the style of language respondents are familiar with; and
- common reporting standards (e.g. accounting standards, industry classifications).

The following sections outline the key areas of questionnaire design:

Questionnaire Structure

The questionnaire should be structured in a logical manner that minimises cognitive load on respondents. Good practice includes:

- the questionnaire should be of an appropriate length (for the proposed mode). See the [ABS Forms Design Standards Manual](#) for more detail on this.
- a logical order of questions;
- logical grouping of questions;
- sensitive questions placed towards the end of the questionnaire, when appropriate;
- if respondents are required to skip past questions, then this needs to be logical and easily understood by respondents.

Supporting Materials

The questionnaire should have a good balance of supporting materials to help, but not burden the respondent to complete the survey. In particular, the questions should be worded and the form should be designed in a way to minimise the need for instructions and definitions, whilst providing clear supporting materials when required.

Instructions and definitions should be:
- clear and concise instructions on how to use the form;
- clear and concise definitions for key, complex, or repeating terms in the questionnaire;
- no unnecessary instructions and definitions, or other superfluous information;
- use of standard definitions and classifications where applicable; and
- location in proximity to the relevant questions and modules

Design of Questions and Response Options

Questions and response options should clearly reflect the data item being collected, minimise the need for supporting materials like notes, and be designed to help the respondent easily provide good quality data.

When designing questions and response options, survey mangers should consider:
- the order of questions, considering effects of sensitive topics and logical grouping;
- layout of questions;
- appropriate types of response options (e.g. multiple vs single choice response options);
- use of filter and sequencing (skip) questions when possible to reduce the burden imposed by irrelevant data items;
- effects of un/ordered response options; and
- appropriate use and design of rating and ranking scales.

Only Necessary Data Items Collected

The survey should only collect data that:
- will be published;
- will be used to reach an estimate;
- is required to validate other data; and
- helps create a controlled context effect to help respondents correctly interpret following questions.

3.4 Infrastructure of Collection Instrument

The survey manager should consider the way the instrument will be received, used and submitted. For example, a lengthy online form seeking detailed financial and human resource data is likely to be filled in by multiple areas of a large organisation. Therefore, the online form must be easy to save, retrieve and circulate within the organisation, whilst also enabling the provider to keep a record for themselves.

Survey mangers should consider:
- appropriate delivery of the instrument;
- access to the instrument;
- ability to save and store the data temporarily before submission;
- ability to save a copy of the data for provider records;
- ability to circulate the form within the business, if required;
- provision of previously reported data in repeating surveys, when appropriate;
- ability to review data before submission;
- submission of the instrument; and
- confirmation of submission.

3.5 Questionnaire Testing

Surveys should undergo thorough testing before they commence. There are a number of types of testing that may be helpful; however, key consideration should be given to:
- the survey instrument and content being tested with respondents to ensure that the concepts are well understood and correctly interpreted, and the questions are able to be easily answered; and
- the systems and infrastructure required to run the survey being thoroughly tested to ensure that they perform as required and that data will be accurately captured.

Section 4. Survey Methodology

In order to justify the burden of surveys, the survey methodology should produce the desired outcomes for which the data is collected. The level of quality required for the survey data to be fit-for-purpose should be largely determined by the needs of the users of that data.

When considering whether a survey will be fit-for-purpose, you should consider whether the data may be:
- over-fit, meaning that the level of quality is much better than is necessary for data users, resulting in a higher level of burden than is required. For example, an Agency may want to run a census of a particular sector of businesses, when a sample of those businesses would still attain the required level of quality.
- under-fit, meaning that the level of quality of the collected data will be poorer than is required for its use.

4.1 Population and Survey Frame

The population is the group of units (e.g. businesses) about which the survey is being conducted. The target population specifically refers to the population of units about which information is required.

The survey frame is the list of units in the population from which you have to run your survey. This is sometimes referred to as your survey population. In an ideal situation, the survey population should be the same as the target population. However, it is common for these two populations to be different. This is sometimes called a "coverage" problem, as your survey frame may be missing units from the target population and/or it contains units that aren't in your target population.

It is important to assess the quality of the frame and determine whether there is a risk that the frame may be systematically different from the target population. The risk with systematic differences between the frame and the target population is that the resulting data may be biased. For example, say you wanted to run a survey of all businesses in Australia, but your frame only contains large business. In this case, there is a risk the attributes of small/medium businesses are different from large businesses. Given data/responses of small/medium businesses will be excluded from the results, this would cause your estimates to be biased and could lead to low quality results.

It is very difficult to obtain a frame that perfectly matches the target population. Instead you should identify and assess any major frame deficiencies and address these where possible. Common frame issues include:
- missing units (under-coverage);
- missing contact information (so units can't participate in the survey and are in effect "missing");
- units that are on the frame but not part of the target population;
- duplicate units; and
- dead units.

Solutions to deal with frame issues can range from:
- modifying your sample design to improve the representativeness of your sample and adjust the weights at the estimation stage;
- find an alternative frame for data collection;
- when weighting results, use population counts from a more accurate frame/source; and
- supplement the frame with a related frame to improve coverage of the target population.

As well as affecting the quality of your results, the information available on the frame (e.g. contact details; auxiliary variables such as the size of the business) may also affect which data collection modes should be considered.

4.2 Sampling Methodology

The survey manager should examine the survey's methodology (e.g. sample design and estimation methods) to ensure that as few respondents as possible need to be contacted in order to meet the objectives of the survey.

The information provided in this section is a broad overview of the issues to consider when determining an appropriate methodology given the survey objectives.

Census versus sample

Surveys are used to collect information about a population. This information can either be gathered by surveying everyone in the population (i.e. by conducting a census) or by sampling only part of the population (i.e. by conducting a sample survey).

The choice between using a census or sample should be driven by the requirements of the data, as there are advantages and disadvantages associated with either method.

The key advantages of a sample survey include:
- reduced cost;
- reduced time in terms of collection, processing data and release of data; and
- reduced respondent burden.

The key advantages of a census include:
- being more suitable for very fine level data and sub-populations (assuming that satisfactory response rates are achieved);
- producing data for the population that can be used as benchmark in future surveys; and
- producing a measure of the population which is not subject to sampling error.

Survey managers are strongly encouraged to use the method that will result in the least amount of respondent burden, but still achieve data that is fit-for-purpose. In general, unless the target population is very small or very fine level data analysis is needed, a census is not usually required.

Probability versus non-probability sampling methods

When forming a sample of businesses to be approached to participate in a survey, an appropriate sampling method needs to be chosen. There are a number of factors that may impact on the choice of method, such as:
- the objectives of the survey;
- the data collection method chosen to meet the survey objectives;
- will the results need to be representative of the whole population;
- the availability and quality of a frame; and
- the resources available for the survey.

However, at a broad level the choice can be broken into whether the method is probability or non-probability sampling.

Probability sampling

Probability sampling refers to any sampling method where each unit (e.g. a business) has a known non-zero probability of being selected to participate in the survey. Using your frame, a sample of units is selected from the population and the probability of selection is recorded.

A probability sampling method is usually required for surveys where estimates need to be generalised to the target population (as opposed to inferring for only those units that actually respond to the survey). The probabilities of selection are used (sometimes in combination with auxiliary information) to form estimates for the population.

The probabilities of selection and/or sample size can also be used to estimate the sampling error (e.g. standard errors and confidence intervals).

Non-probability sampling

If the probability of selection for each unit is unknown, or cannot be calculated, the sample is called a non-probability sample. Non-probability samples are often less expensive, easier to run and do not require an exhaustive frame. Examples include people responding to a survey they find on a website, or a survey that is sent to a non-random sample (e.g. largest 100 businesses).

Non-probability sampling methods can be really useful in some scenarios, such as pilot testing. However, because there is no control over the representativeness of sample, it may be misleading to make generalisations about the population, and the precision of estimates (e.g. standard errors) cannot be calculated.

Comparison chart of probability versus non-probability sampling methods

	Probability methods	Non-probability methods
Common types for business surveys	- Simple random sampling - Systematic sampling - Stratified sampling	- Quota sampling method - Convenience/haphazard sampling - Purposive sampling - Snowball sampling
Probability of selection	Known	Unknown
Frame	Good quality frame required	Frame not necessarily required
Cost	Higher	Lower
Precision	Measures of sampling error can be estimated	Precision measures cannot be calculated.
Generalisability	Estimates may be made of the population	Estimates of the population may be of poor quality unless the sample design has been carefully tailored for this purpose.
Suitable for:	- When estimates of a population are required - When the quality of results needs to be measured/controlled	- Case studies - Qualitative research - Exploratory/pilot studies - When generalisations about the population are not required

Stratification

Stratification is used to improve the efficiency of the design of the sample. This is achieved by grouping similar units together (into strata), and then sampling from these strata. By grouping similar units into strata, the overall sample size required to meet sampling error constraints should be reduced. It can also be used to control the quality (sampling error) for different sub-groups that will be produced in outputs.

Sample size

The sample size for a survey will impact on both the quality of results and the cost of the survey. Survey managers will need to consider their required responding sample size (i.e. number of completed/submitted surveys) and the sample size to approach (i.e. number of units contacted to complete a survey).

Important issues to take into account when determining an appropriate sample size include:
- the objectives of the survey.
- the size of the population. When the population size is small, it needs to be considered carefully in determining the sample size. But when the population size is large, it has little effect on the sample size.
- the variability of the population. The more varied the population, the larger the required sample size will be. Previous data can sometimes be used to give an indication of the level of variation that can be expected.
- the resources/budget available for data collection and processing.
- the level of detail required. For surveys that require fine level breakdowns of data, then the sample size should be sufficiently large enough to cover this level of detail.
- the type/level of stratification required to meet data output needs.
- the level of precision required. More precise estimates require greater sample sizes. Survey managers should ensure the quality level will achieve fit-for-purpose results, and that sample sizes are not unnecessarily high.
- the level of expected non-response. When the response rate is expected to be low, the sample size can be inflated to increase the number of responses. This approach should be treated with caution. As discussed in **Section 4.3 Response Rates and Non-response Bias**, high levels of non-response carry a number of issues for the interpretation and use of data and should be mitigated where possible through the use of improved data collection practices and follow-up strategies.

4.3 Response Rates and Non-response Bias

The survey response rate refers to the proportion of responses received from the number of eligible respondents that were surveyed. When data is not collected from a respondent who is otherwise eligible to respond, this is referred to as non-response.

The expected response rate can be expressed as a percentage of:
- the number of businesses that you anticipate will agree to complete (or partially complete) the survey, divided by
- the number of businesses that you will attempt to make contact with.

Non-response can occur for an entire survey form or for only some questions in a survey (partial non-response). Respondent burden is a factor affecting response rates. In particular, survey managers should avoid:
- a questionnaire that is too lengthy;
- questions that are poorly worded; and
- questions that ask for information at too fine a level of detail.

A key reason for using a probability sampling method is to maximise the likelihood that the sample will be representative of the target population. This enables the data to be generalised to the population, without having to survey the entire population. However, the representativeness of a sample can be affected by the proportion of non-responding units in the sample. Generally speaking, the lower the response rate, the higher the risk that the responses from the responding units will not be representative of the population.

Non-response is particularly concerning when non-responding units vary in a systematic and meaningful way to responding units. For example, say you were to contact businesses after a transaction to ask them how satisfied or dissatisfied they were. Businesses who were very dissatisfied with the transaction may decide that they no longer want any contact with you and therefore choose not to complete your survey. The resulting data would provide a positively-biased level of satisfaction. Such data does not accurately reflect the population of all the businesses you've had a transaction with. This issue is referred to as non-response bias.

It is very difficult to determine whether data is biased due to non-response. One method to determine whether data may be biased is to conduct a post enumeration survey (PES). A key aim of a PES is to find out why a business did not respond. For example, was it because they did not have time; because they didn't receive the survey in the mail; or was it for a reason that relates to a data item you were trying to collect, such as satisfaction/dissatisfaction? If the PES shows a number of non-responding businesses did not respond due to a reason related to the data item you were trying to collect, then this is indicating the data may be biased.

A PES may also be used to attempt to collect responses for key data items from a sample of previously non-responding units. With such data it may be possible to determine whether there is a significant difference between responding and non-responding units for these key data items, with the potential to apply a correction to the final data set in order to account for any systematic non-response bias.

In practice, a PES can be a time consuming and expensive activity to perform. However, analysing the frame/survey data of respondents/non-respondents may help identify if there are any patterns to the responses. For example, are response rates lower for different industries, sizes, sectors, etc? This information may then be used in the weighting strategy to help reduce the risks of non-response bias (**see Section 5. Estimation and Analysis for more detail**).

The most effective way to deal with non-response is to maximise the response rate, where possible. Response rates can be maximised through good forms design principles and data collection methods, such as using short simple questions and communicating the purpose of the survey effectively to the target audience. Assuring respondents that their individual responses will remain confidential is also critical to allaying respondent fears about privacy.

Survey managers should also consider employing a well-planned targeted follow-up strategy to attempt to gain responses from initially non-responding businesses, in order to maximise the final response rate. However, it's important that a reasonable balance is struck between targeted follow-up in order to maximise the response rate, and the added burden of contacting units more than once to ask for a response.

Section 5. Estimation and Analysis

5.1 Estimation

*This section is only relevant to those surveys that are based on a probability sampling method. If you are using a non-probability method then proceed to **5.2 Analysis and reporting**.*

When analysing survey results, we often estimate the population total and mean (average). In order to do this, we need to scale/adjust our survey results to reflect the population. This "scaling" is done through a process of weighting and estimation. Survey managers should try to ensure that their estimates are both accurate (i.e. minimise bias) and precise (i.e. have appropriate levels of sampling error).

Weighting
In order to avoid drawing incorrect conclusions, great care must be taken to weight and aggregate the data correctly. Weighting is the process whereby each unit in the sample has its response inflated to represent the response from all similar units in the population. A unit's weight indicates how many units in the population it represents. For example, a unit with a weight of 10 represents itself and 9 others. The weight allocated to each sample observation depends on the process used to select the sample. The simplest form of weighting is where a simple random sample (SRS) of size n is selected from a known population of size N.

Number-Raised Estimation
The number-raised estimator is a simple and unbiased estimator that can be used to calculate population estimates. The estimator assumes a simple random sample, with each unit receiving the same weight (N/n). If we observe a sequence of n observations $y_1, \ldots, y_n$ from a population of size N, then the number-raised estimator for the population total is the sample total multiplied by the ratio of population size to sample size (N/n). The number-raised estimate is unbiased as the average of all possible samples is the true population total.

$$\hat{Y} = \left(\frac{N}{n}\right) \sum_{i=1}^{n} y_i$$

Business Example
Suppose we want to estimate total employment (y=employment) for 100 cafe businesses located in the City of Melbourne (N=100). Due to resource constraints we can only approach 10 businesses (n=10). Each business has a one in ten chance of selection and hence each business selected represents itself and 9 other businesses. The weight allocated to each selected business is therefore 10. Let's also say that of our 10 selected businesses, they report the following number of employees: 2, 4, 12, 21, 20, 23, 32, 43, 34 and 22. Using the formula above, we estimate the total employment for the population to be:

Estimate of employment = (100/10) x (2+4+12+21+20+23+32+43+34+22) = 2,130

Advantages
This form of estimation is easy to use and does not require any benchmark information. It is relatively simple to calculate and its variance formula is known.

Disadvantages
Number-raised estimation has problems in that it produces a larger sampling error compared to some other methods and returns results of a poorer quality for unrepresentative samples.

Weighting to Sub-population totals

In the previous example we used the original probabilities of selection to weight our data, with all units receiving the same weight (N/n). However, we can use finer breakdowns of our population to create final weights and use this to help correct for poor samples and help minimise non-response bias. To illustrate, let's consider our cafe example. Say that of the 10 cafes that responded, two were small cafes, while eight were large cafes. Furthermore, let's say that we know from our frame that of the 100 cafes in Melbourne, 50 are small and 50 are large. We can see from our sample that we have an over representation of large cafes. Large cafes may have different characteristics to small cafes (e.g. larger turnover, more employees, etc.). However, under the simple number raised estimator, each unit will get the same weight of N/n = 100/10 = 10, and therefore the estimates for the population may be poor quality given the over representation of large cafes. To address this we can "post-stratify" (or re-weight) our results to make it more representative. This involves comparing the sample we achieved to the frame counts, and calculating estimates for each sub-population separately.

To do this, we split the sample to consider the two sub-populations (small cafes and large cafes). For the small cafes, we have a sample of 2 cafes representing the 50 in the population, so each receives a weight of 50/2 = 25. Let's also say they report employment counts of 2 and 4. For the large cafes, we have a sample of 8 cafes representing the 50 in the population, so they would receive a weight of 50/8=6.25, and let's say they report employment counts of 12, 21, 20, 23, 32, 43, 34 and 22. We then estimate the employment total for small cafes, and the estimate of employment for large cafes, and add them together to get the overall total cafe employment for Melbourne:

Estimate of employment for small cafes = (50/2) x (2+4) = 150
Estimate of employment for large cafes = (50/8) x (12+21+20+23+32+43+34+22) = 1,294

Estimate of employment for all Melbourne cafes = 150 + 1294 = 1,444

You can see that this estimate of 1,444 employees is smaller than the estimate of 2,130 employees that we obtained in the first example. That is because our new estimate has re-weighted the results to ensure smaller cafes are more fairly represented.

It's important to note that although we have post-stratified our sample, the new weights still add to the population total. Specifically, we have the weights from the two small cafes (2 x 25) and the weights from the eight large cafes (8 x 6.25), which together add to 100. So we are still weighting our sample of 10 up to 100, but we have tailored the weighting to help ensure more representative estimates.

5.2 Analysis and reporting

Adjustments

When data is being prepared for output there are some common adjustments that can be made to improve the quality of the final output. Most of these adjustments usually apply to probability-based samples, where the intention is to make the data as representative of the population as possible.

Outliers

Outliers in the data are those units whose responses are unusual and not thought to be representative of other units in the population.

Generally, an observation should only be treated as an outlier if:
- it has a large effect on the estimates of interest <u>and</u>
- it is considered to be atypical based on your knowledge of the population

There are a number of methods for treating outliers. However, the simplest method is to reduce the outlying unit's weight to 1, meaning that the unit only represents itself in the population. The weights of the remaining units in the same stratum need to have their weights adjusted upwards to account for the removal of the outlying unit.

Data Errors

If the survey manager feels that the responding unit has reported incorrect data and this data cannot be corrected through editing or imputation, then it may be appropriate for this unit's response to be completely removed from a data set. For example, say you are conducting an online survey covering financial data items, and the survey should take about one hour to complete. The response for one of your completed surveys looks atypical and is having considerable impact on estimates. After examining the survey metrics you see that the respondent took only five minutes to complete the survey form. You know that it would be impossible to correctly/accurately answer the questions in this amount of time, and you suspect that the respondent has reported false/made-up data. You are unable to correct this data and in order to preserve the quality of your overall results; you remove the responses from this unit from the final estimates and adjust the weights accordingly.

Non-response treatment

Where response rates are still low after all reasonable attempts of follow-up are undertaken, you can reduce bias by using population benchmarks/counts to post-stratify the sample (for more detail see **Section 5.1 Estimation**), intensive follow-up of a subsample of the non-respondents or imputation for item non-response (non-response to a particular question).

The main aim of imputation is to produce consistent data without going back to the respondent for the correct values thus reducing both respondent burden and costs associated with the survey. Broadly speaking the imputation methods fall into three groups:
- the imputed value is derived from other information supplied by the unit;
- values from other units can be used to derive a value for the non-respondent (e.g. average); or
- an exact value of another unit (called donor) is used as a value for the non-respondent (called recipient).

When deciding on the method of imputation it is desirable to know what effect imputation will have on the final estimates. If a large amount of imputation is performed the results can be misleading, particularly if the imputation used distorts the distribution of data.

Analysis

In analysing results, there are some commonly used statistical methods which enable you to summarise your results. The analysis that is to be carried out should be taken into account at the planning stage of the survey. Even at that early stage, the output tables should be specified and the analysis techniques decided upon to ensure that the necessary data is collected.

Data analysis involves interpreting the data, breaking it down and manipulating it to answer the survey objectives. The data is sorted into categories and summarised to obtain information like descriptive statistics, frequencies, percentages, correlations and measures of both location (e.g. mean, mode, median, percentiles) and spread (e.g. range, variance, standard deviation, standard error). This information is used to formulate hypotheses and make inferences about population data by estimating confidence intervals and testing the hypotheses to determine significances and trends in the data.

Reporting

Presenting the results in a clear and logical format to the client is one of the most important tasks for the survey manager. When presenting results, the format of the presentation should be tailored to address the aims and objectives of the survey and to satisfy the potential users of the results. Consideration should be given to the level of statistical understanding of the clients and users, particularly in regard to statistical terminology. The presentation needs to be effective, easy to understand and convey the main features of the data.

Users should also be made aware of the quality of the survey outputs, so they can understand how much confidence to place in the findings. Users should be aware of whether the results are representative of the population or whether they are only indicative of those that responded (e.g. like a focus group). It is also useful to communicate other quality metrics such as response rates and levels of sampling error.