

Primary Data Collection - Surveys

Table of Contents

Introduction.....	2
Step 1: Identify the variables of interest	2
Step 2: Define the population of interest	2
Step 3: Develop an analysis plan.....	2
Step 4: Design your survey questions.....	2
Step 5: Determine sample size and sampling method	4
Step 6: Develop data collection tools	5
Step 7: Test your survey	6
Step 8: Distribute your survey	7
<i>Considerations for encouraging a high response rate</i>	<i>7</i>
Step 9: Enter and store your survey responses	8
Step 10: Analyze and present your results	8

INTRODUCTION

Primary data is data collected for the purpose for which it is being used. In this case, it is data you collect yourself for the purposes of tracking indicators. If data cannot be obtained from external data sources or internal administrative data sources, you will need to collect it. This document focuses on surveys, a common way to collect population level indicator data from your citizens. Note that there are many other data collection methods available to you, such as focus groups and interviews, but these are beyond the scope of this toolkit.

This document introduces the main steps and considerations in developing and performing a survey. A key theme throughout this document is the reduction of **error**. **Error** in this context is the difference between the true value and the value measured by your survey. Different types of error will be discussed in this document.

STEP 1: IDENTIFY THE VARIABLES OF INTEREST

Before you start your survey, you should have a strong understanding of the objectives and key data gaps you are looking to fill. Surveys can be costly and difficult to design. Starting with a clear statement of objectives (see [Chapter 2 - Survey Methods and Practices - Statistics Canada](#)) can help guide decision making at each step of survey design and analysis.

See the [Developing indicators](#) section of the Toolkit website. In order to limit costs and avoid survey fatigue, your survey will likely focus on indicators for which good data is not available from other sources. You may choose to maximize the value of your existing data by asking the same question as another survey your government ran so that you can compare results over time.

STEP 2: DEFINE THE POPULATION OF INTEREST

The population of interest for your survey might be your entire citizen population, or a subset of the population (e.g., citizens living on-lands).

STEP 3: DEVELOP AN ANALYSIS PLAN

Once you know the indicators and population you will be focusing on in your survey, it is time to develop an analysis plan.

Your analysis plan will dictate how you measure each indicator, and therefore, how the survey questions are designed. If you do not develop your analysis plan until after the survey is done, you may find that you are not able to do the analysis you want to do with the data you have collected.

See the [Data Analysis](#) section of the Toolkit for more information about developing an analysis plan. Note: Your analysis plan must be developed before you design your survey!

STEP 4: DESIGN YOUR SURVEY QUESTIONS

After you develop your analysis plan, you can design your survey. Questionnaire design is complex and we recommend that you do further reading. Some good places to start are:

- The [free online course on question design](#) by Marist College.

- The book “Improving survey questions: design and evaluation” by Floyd J. Fowler, Jr.
- The book “Social Research Methods” by Alan Bryman, Edward Bell, and James Teevan.

A few key guidelines for designing a questionnaire include¹:

1. Ensure all instructions and questions are clear and specific. Test the questionnaire out on yourself and several others to check for unclear instructions or questions.
2. Avoid ambiguous terms that respondents could interpret differently. For example, instead of using terms such as ‘often’, ‘not very often’, ‘several’, or ‘a few’, use very clear language such as ‘3-5 times per day’.
3. Avoid long questions, as respondents may skip over them.
4. Avoid double-barrelled questions, such as “How important is spirituality and culture in your life?”. Since spirituality and culture are different things, the respondent will not know how to answer the question. Each question should ask about one thing.
5. Avoid general questions, such as “How satisfied are you with your life?”, since this could be interpreted very differently by different respondents.
6. Avoid questions that include negative terms, such as ‘not’, since the respondent might skip over the ‘not’ and give the opposite response to what they intend.
7. Avoid technical terms that respondents may not understand.
8. Limit the number of open questions (question which can be answered however the respondent wishes), as they are time consuming to analyze and can put off respondents due to the time and effort needed to complete them (i.e., the survey’s response burden).
9. Keep the questionnaire as short as possible to limit respondent fatigue.
10. Consider the impact of question order (i.e., priming effects). The order in which questions are asked can impact the way a respondent answers them. Pew Research has an article on [priming and order effects](#).
11. Ensure that response options align with the phrasing of the question. An example of lack of alignment between a question and response options is:

Open question: A survey question which the respondent can answer however they wish.

Closed question: A survey question for which there are fixed response options.

How important is spirituality in your life?

- Very significant
- Significant
- Neither significant nor insignificant
- Insignificant

¹Bryman, A., Bell, E., Teevan, J. Social Research Methods, 3rd Canadian Edition.

- Very insignificant

The response options should range from “very important” to “very unimportant” in order to align with the phrasing of the question.

This range of answers is an example of a Likert Scale. There are lists of commonly used Likert Scales online, like this one from [Iowa State University](http://www.iastate.edu/~lbrunson/psych501/501lect11.htm).

12. Ensure that response options are mutually exclusive and comprehensive. In other words, response options should not overlap with each other and should encompass all possible situations. In order for a set of response options to be comprehensive, you may have to include options such as “other”, “unknown”, or “don’t know”.
13. Ensure that there are an equal number of positive and negative response options, so that the responses are not weighted towards one direction or the other. An example of a set of response options that is biased towards the positive is:
 - Extremely satisfied
 - Very satisfied
 - Satisfied
 - Neither satisfied nor unsatisfied
 - Unsatisfied

This set of responses has 3 positive options, 1 neutral option, and 1 negative option. In order to balance the responses, there should be 3 negative options that are equivalent to the 3 positive options, as shown here:

- Extremely satisfied
- Very satisfied
- Satisfied
- Neither satisfied nor unsatisfied
- Unsatisfied
- Very unsatisfied
- Extremely unsatisfied

STEP 5: DETERMINE SAMPLE SIZE AND SAMPLING METHOD

If it is not possible to survey all the relevant individuals in the population (e.g., due to budget or time constraints) you may need to survey a subset of the population (i.e., a sample). Sample size and sampling methods are complex and important ideas. This section provides an introduction to these topics.

The **sample** is the subset of the population that is selected to respond to the survey. If you want to project/generalize your survey results onto your entire population of interest, you must limit **sampling error** by having a large enough **sample size** and choosing a **representative sample** (a sample that represents the characteristics of the larger

population. For example, a representative sample should reflect the same age distribution, gender distribution, and income distribution of the larger population).

The technique used to project/generalize your sample results is called **inferential statistics**. A representative sample is the ideal, but in practice it is often impossible to truly achieve a representative sample. There are methods like stratification that can bring a sample closer to being representative. Practical statistics is often about doing your best and making assumptions that you can justify. Further details can be found in the [Inferential Statistics](#) section of the Toolkit website.

Sampling error is the error introduced by not surveying the entire population. It is the difference between your sample population and the entire population of interest. Sampling error cannot be completely eliminated, but it can be controlled through sample size and sampling methodology.

Calculating the necessary sample size is a key part of planning your survey. In order to get meaningful results, your sample size must be big enough.

The sampling method is how survey respondents are selected. The sampling method may be either non-probability sampling or probability sampling:

1. **Non-probability sampling:** Sampling techniques that are not designed to limit bias and error, such as convenience sampling (i.e., surveying individuals who are easily accessible). Results from this sampling cannot be projected/generalized onto the larger population. Non-probability sampling is typically used for exploratory studies.
2. **Probability sampling:** Sampling techniques that are designed to limit bias and error so that results can be projected/generalized on the larger population. Examples of probability sampling include random sampling and systematic sampling. Each member of the population must have some chance of being selected for the sample population.

A **census** is a survey that attempts to collect data from all members of the population. A census does not have sampling error because there is no sample; everyone participates! Performing a census is the best way to limit sampling error but can be very costly depending on the size of your population and how often you conduct a census.

STEP 6: DEVELOP DATA COLLECTION TOOLS

After developing your survey, you will need to develop data collection tools.

Ensure that the survey is accessible to everyone. That may mean creating multiple formats for the survey. Options may include paper, online, and one-on-one interviews.

Ensure that quality control processes are incorporated into the data collection tool. Quality control processes are easiest to incorporate into online surveys. For an online survey, you may build in features like reminders to fill out empty fields, minimum and maximum values for numeric fields, and restrictions on the type of data that may be entered (e.g., only text, only numbers, only dates). For a paper survey, format and labeling are particularly important to ensure the survey is as clear as possible and respondents fill it out properly.

STEP 7: TEST YOUR SURVEY

Before you begin distributing your survey, it is important to test it to identify any potential errors or improvements.

Non-sampling error comes from problems with question design that cause respondents to give answers that are inaccurate. This can come from confusion about the wording of questions (e.g., missing the word “not” in a question) (see Step 5), questions that are technically impossible to answer correctly (e.g., phone numbers of people without phones), or survey fatigue (e.g., respondents giving answers without reading questions because the survey has gone on too long, or submitting an incomplete survey).

Testing your survey helps to ensure you are collecting useful, actionable data (i.e., data that can be used to meet your research objectives). Depending on the time and resources available, you may wish to test your survey by:

- Taking the survey yourself;
- Sending the survey to other members of your governments and to individuals with relevant content knowledge;
- Conducting multiple test-iterations of the survey; or
- Sending the survey to a small pilot group of actual respondents.

If possible, it is ideal to pilot the survey with a small sample of individuals from your population of interest. Using the answers and feedback from this small sample, edits can be made to the survey before it is sent to the rest of the population. Be sure to check the validity and quality of responses for anything unexpected. By piloting the survey, you can also generate dummy data, and use this as a final check that your survey is able to generate the charts, data, and information you need for your intended research question.

Whichever method of testing you choose, encourage testers to:

- Walk in the shoes of potential participants
- Think of unusual cases. Ensure the questions and response options do not exclude any person or situation.
- Think about how different situations could change the meaning of an individual's answer.
- Do participants have enough information to respond truthfully? For example, do you use acronyms they might not be familiar with?

Testing your survey will reduce the risk that participants could be confused or frustrated by questions they do not understand or that do not apply to them, which in return will lower the response burden and produce higher quality data.

Survey testing also provides a good opportunity to evaluate the survey's length. If the survey is too long and requires a large amount of energy, participants are more likely to leave the survey incomplete. To prepare your participants for the survey and to be honest about the amount of effort required, it is helpful to include an estimated *time to completion* in the introduction.

STEP 8: DISTRIBUTE YOUR SURVEY

Distribute your survey according to the sampling method designed in step 5.

Ensure respondents are given essential information about the survey before they start to fill it out. Essential information includes the purpose of the survey, under what authority the information is being collected, how the information will be used, and how their privacy will be protected, as well as information about how respondents can access the results. This information can be included in the introduction of the survey. See the Statistics Canada [“Policy on Informing Survey Respondents”](#) as an example of the type of information that you might consider providing.

The response rate is the percentage of people who complete your survey out of the total group of possible survey participants or desired sample. Surveys with high response rates are less likely to suffer from non-response bias or non-response error. If a substantial number of individuals do not respond to your survey in relation to the total sample size, there is a chance that those who participated have different traits compared to those who chose to opt-out of the survey.

An example of potential non-response bias is satisfaction surveys. Suppose your survey results only capture the opinions of individuals who have had bad experiences, and individuals who have had good experiences do not participate. In that case, it may appear that the service you are offering is not satisfactory. When, in fact, individuals are satisfied, but you were unable to collect that information. If you achieve 100% response rate, there is no bias introduced by non-response.

Considerations for encouraging a high response rate

Consider the following practices to increase your chances of achieving a high response rate:

- Ensure transparency, openness, and choice around data use. Respondents should understand what information is being collected, what it will be used for, what risks there may be in sharing data, and any data retention or privacy policies it will be subject to in clear and plain terms.
- Be honest and upfront about the survey’s length, requirements, and topics.
- Provide respondents with the opportunity to skip a question or (better yet) choose a response that indicates the question does not apply to them, they do not know, or they prefer not to answer.
- Do not surprise your participants. Make an intuitive, smooth survey flow. Ask yourself: what order makes sense? What questions do you really want them to answer?
- Think of any new question as an added cost to an individual's time (Is it necessary? Is it worth it?)
- Make things easy and simple to read and see
 - Use simple and attractive colours and images to improve respondents' experience. Consider the look and feel of the data collection tool you have chosen.
 - Make your survey easy to find (bright posters, easy links to click, or websites to search)
- Use simple language that is appropriate for your audience. Do not assume your participants will know something they might not.

STEP 9: ENTER AND STORE YOUR SURVEY RESPONSES

You can collect survey responses through multiple methods. All responses must be stored together in the same database before analysis can begin.

If some, or all, of the survey responses were collected on paper or through interviews, responses will need to be entered into a database manually. This process is most fallible to human error, so ensure that you develop quality control protocols like double checking that data has been entered properly.

Survey results collected through an online platform will need to be exported and stored securely.

Ensure that you store the data set with a data dictionary containing information about each survey field (e.g., question, data format, any additional instructions included in the questionnaire) so that any future data users properly interpret the data.

Have a plan or policy for how long you will retain your data and the appropriate process for deletion, destruction, or archiving.

STEP 10: ANALYZE AND PRESENT YOUR RESULTS

It is very important to present a measure of the sampling error when you present your survey results. It is also important to present your response rate, sampling method, and consider how non-response error may be impacting your results. See the [Data Analysis](#) section of the Toolkit website for information.