The Ultimate Guide to Effective Data Collection

The Ultimate Guide to Effective Data Collection

Chapters:

The Ultimate Guide to **Effective Data Collection**

Bonus Content

- *1. Designing a Great Survey*
- *2. Piloting a Survey*
- *3. Field Data Collection Plan*
- *4. Census Survey*

Introduction:
Why Data Quality is Crucial

> “In God we trust. All others must bring data.”
>
> - *W. Edwards Deming*
> Statistician, professor, author, lecturer and consultant

Deming describes the importance that organisations, funders, and people in general give to data today. Data has become fundamental in nearly every aspect of life.

Businesses and corporates use data to make better decisions, increase profits, grow revenues and improve efficiency. Organizations such as hedge funds, stock brokers and investment banks — where a split second delay in decision making can lead to huge losses — have been stalwarts in using data to make the smallest decisions.

In addition, the development and policy spaces have seen the use of data to drive decision making and increase impact. Non-profit and government organisations are using data to inform decisions, such as how much money to invest in a particular project or how to improve impact per dollar spent.

With the growing importance placed on data, surveys have become an indispensable tool for every organization, from billion-dollar tech companies to rural nonprofits.

Creating a survey seems simple — just ask a few simple questions, and you'll get back data to solve your every problem. However, designing a survey correctly takes time and knowledge. A poorly-designed survey will lead to useless data, wasting your time and money. As computer scientists say, "Garbage in, garbage out."

Case Study

In 1936, the Literary Digest polled 2.4 million people on the upcoming U.S. presidential election. After conducting one of the largest and most expensive polls in history, the *Literary Digest* predicted that Alfred Landon would win the election 57% to 43% against the incumbent Franklin D. Roosevelt. At the same time, George Gallup polled around 50,000 people and predicted a win for Roosevelt.

The actual results of the election were 62% for Roosevelt against 38% for Landon. The *Literary Digest* poll's prediction had an error of 19%, the largest error in the history of major public opinion polls in the U.S.

The explanation for this error — survey design. Though Gallup surveyed only 2% of the people that the *Literary Digest* did, Gallup's data was far more accurate because he designed the survey and sampled the population more effectively.

Designing a survey involves several considerations:

• What is the purpose of your survey? What data are you looking to collect?
• How can you best collect that data? What sort of survey and research methodology should you use?
• How should you write the questions in your survey?
• Who should you survey?

This ebook is designed to take you through these questions and help you design a survey that will give you high-quality data. Chapter 1 will help you think through the purpose, outcomes and indicators of your survey. Chapters 2 and 3 will help you determine what data collection method you should use, as well as whether you need a qualitative or quantitative survey. Chapters 4-8 address writing the questions in your survey — what you want to ask and best practices around how to ask it. Lastly, Chapter 9 covers all aspects of sampling your population — sampling methods, best practices, and a quick sample size formula.

Chapter 1: Survey Design: Creating Your Research Question, Outcomes and Indicators

The most important part of your survey is determining your purpose – why are you conducting this survey and what do you want to learn?

Setting your research question, outcomes and indicators clearly makes writing the rest of your survey far simpler. Moreover, it ensures that everyone in your organization is on the same page about your survey.

This chapter will show how to build your research question, outcomes and indicators through an exploration of two case studies.

Determining Your Research Question

Before you start collecting data, it is important to figure out your research question. As part of this process, you also should think broadly about who you can survey.

To fully formulate your research question, you should be able to answer three questions:

1. What is my research question?
2. Why am I collecting this information?
3. Who can I collect this data from?

Tool Tip
The people you are collecting data from are called the "target population" or "sample".

Case Study #1

You run an education NGO, which works on a teacher training program across 1,500 schools in Bihar and Uttar Pradesh. As part of your program, your team trains teachers on how to improve your students' reading skills.

1. What is the research question behind collecting data on the impact of the programme?

How has my NGO improved the teaching skills of the teachers we work with, and how has this improved student reading skills?

2. Why am I collecting this information?

To measure the impact of my NGO's programme so I can compare it to my other programmes and communicate its impact to my funders.

3. Who can I collect data from?

I can collect data from the students and teachers where the NGO works.

Case Study #2

Your NGO works with women's SHGs from a district in Jharkhand. Your model for change is two fold – you directly impact the SHGs you work with by helping them fundraise, which in turn empowers women who participate in the SHGs.

1. What is the research question behind collecting data on the impact of the programme?

How has my NGO contributed to increasing the funding of the SHGs we work with?

2. Why am I collecting this information?

To measure the impact of my programme and show the effect of our work.

3. Who can I collect data from?

I can collect data from the SHGs that my NGO works with and the families of the members of the SHGs. I can also collect data on SHG funding from my program officers.

Tool Tip
The second question might seem unnecessary. After all, it doesn't directly go into your research question. However, it is an essential part of the survey design process. If you cannot fully answer why you are conducting your survey, you are not ready to start your survey.

Once you can answer these three questions for your own survey, you have figured out your research question and target population.

Determining Your Outcomes

Once you have determined a research question, you can create the set of outcomes for your survey.

An outcome is something that you can track to measure data on your research question. Outcomes should be feasibly measurable.

Case Study #1

You run an education NGO, which works on a teacher training program across 1,500 schools in Bihar and Uttar Pradesh. As part of your program, your team trains teachers on how to improve your students' reading skills.

Research Question

How has my NGO improved the teaching skills of the teachers we work with, and how has this improved student reading skills?

Outcomes

Change in teachers' teaching skills
Change in students' reading skills

Case Study #2

Your NGO works with women's SHGs from a district in Jharkhand. Your model for change is two fold – you directly impact the SHGs you work with by helping them fundraise, which in turn empowers women who participate in the SHGs.

Research Question

How has my NGO contributed to increasing the funding of the SHGs we work with?

Outcomes

• Change in the funds raised by the SHGs before and after they entered my programme
• Empowerment of women in the district

Determining Your Indicators

Once you have determined your set of outcomes, you can create the indicators for each outcome.

An indicator is a data point (or data points) that measures an outcome. Indicators must be measurable (using either qualitative or quantitative data) within the framework of your survey.

Case Study #1

You run an education NGO, which works on a teacher training program across 1500 schools in Bihar and Uttar Pradesh. As part of your program, your team trains teachers on how to improve your students' reading skills.

Research Question

How has my NGO improved the teaching skills of the teachers we work with, and how has this improved student reading skills?

Outcomes

Change in teachers' teaching skills
Change in students' reading skills

Indicators

Measuring change in teachers' teaching skills:

• Conduct written assessments over time.
• Observe the teachers' classes at intervals.
• Take qualitative feedback from students regarding the teachers' classes.

• Ask the teachers to reflect on their improvement over time.

Measuring change in students' reading skills:

• Track students' homework assignments.
• Track scores on reading exams.
• Conduct reading assessments at intervals.
• Ask students to reflect on their reading improvement over time.

Case Study #2

Your NGO works with women's SHGs from a district in Jharkhand. Your model for change is two fold – you directly impact the SHGs you work with by helping them fundraise, which in turn empowers women who participate in the SHGs.

Research Question

How has my NGO contributed to increasing the funding of the SHGs we work with?

Outcomes

• Change in the funds raised by the SHGs before and after they entered my programme
• Empowerment of women in the district

Indicators

• Measure funds raised by SHGs each month, starting when they enter your programme
• Take qualitative feedback from women on whether they feel empowered, each month
• Measure women's likehood to make household decisions, each month

After creating your research question, outcomes and indicators, you will know exactly what data you need to collect. From there, it is much easier to design the rest of your survey.

Chapter 2:

Data Collection Methods

Once you know what data you want to collect, it is important to figure out which data collection method you will use. Each method has its own advantages, disadvantages and use cases.

Tool Tip
Any research is only as good as the data that drives it, so choosing the right method of data collection can make all the difference.

Observation

Seeing is believing, they say. Making direct observations, when the situation allows for it, is a very quick and effective way of collecting data with minimal intrusion. Establishing the right mechanism for making the observation is all you need.

Advantages

• Non-responsive sample subjects are a non-issue when you are simply making direct observations.
• This mode does not require a very extensive and well-tailored training regime for the survey workforce.
• It is not as time-consuming as the other models that we will discuss below.

• Infrastructure requirement and preparation time are minimal.

Disadvantages

• Heavy reliance on experts who must know what to observe and how to interpret the observations once the data collection is done.
• Possibility of missing out on the complete picture due to the lack of direct interaction with sample subjects.

Use Case

Making direct observations can be a good way of collecting information about mechanical, orderly tasks, like checking the number of manual interventions required in a day to keep an assembly line functioning smoothly.

Questionnaires

Questionnaires, as we consider them here, are stand-alone instruments for data collection that are administered to the sample subjects either through mail, phone or online. They have long been one of the most popular data collection methods.

Advantages

• Questionnaires give the researchers an opportunity to carefully structure and formulate the data collection plan with precision.
• Respondents can take these questionnaires at a convenient time and think about the answers at their own pace.
• The reach is theoretically limitless. The questionnaire can reach every corner of the globe if the medium allows for it.

Disadvantages

• Questionnaires without human intervention (as we have taken them here) can be quite passive and can miss out on some of the finer nuances, leaving the responses open to interpretation. Interviews and Focus Group Sessions, as we will see later, are instrumental in overcoming this shortfall of questionnaires.
• Response rates can be quite low. Choosing the right question types can help to optimize response rates, but very little can be done to encourage the respondents without directly conversing with them.

Use Case

A survey can be carried out through directly-administered questionnaires when the sample subjects are relatively well-versed with the ideas being discussed and comfortable at making the right responses without assistance. A survey about newspaper reading habits, for example, would be perfect for this mode.

Interviews

Conducting interviews can help you overcome most of the shortfalls of the previous two data collection methods that we have discussed here by allowing you to build a deeper understanding of the thinking behind the respondents' answers.

Advantages

• Interviews help the researchers uncover rich, deep insights and learn information that they may have missed otherwise.
• The presence of an interviewer can give the respondents additional comfort while answering the questionnaire and ensure correct interpretation of the questions.
• The physical presence of a persistent, well-trained interviewer can significantly improve the response rate.

Disadvantages

• Reaching out to all respondents to conduct interviews is a massive, time-consuming exercise that leads to a major increase in the

cost of conducting a survey.
• To ensure the effectiveness of the whole exercise, the interviewers must be well-trained in the necessary soft skills and the relevant subject matter.

Use Case

Interviews are the most suitable method for surveys that touch upon complex issues like healthcare and family welfare. The presence of an interviewer to help respondents interpret and understand the questions can be critical to the success of the survey.

Focus Group Discussions

Focus Group Discussions take the interactive benefits of an interview to the next level by bringing a carefully chosen group together for a moderated discussion on the subject of the survey.

Advantages

• The presence of several relevant people together at the same time can encourage them to engage in a healthy discussion and may help researchers uncover information that they may not have envisaged.
• It helps the researchers corroborate the facts instantly; any inaccurate response will most likely be countered by other members of the focus group.
• It gives the researchers a chance to view both sides of the coin and build a balanced perspective on the matter.

Disadvantages

• Finding groups of people who are relevant to the survey and persuading them to come together for the session at the same time can be a difficult task.
• The presence of excessively loud members in the focus group can subdue the opinions of those who are less vocal.
• The members of a focus group can often fall prey to group-think if one of them turns out to be remarkably persuasive and influential. This will bury the diversity of opinion that may have otherwise emerged. The moderator of a Focus Group Discussion must be on guard to prevent this from happening.

Use Case

Focus Group Discussions with the lecturers of a university can be a good way of collecting information on ways in which our education system can be made more research-driven.

Tool Tip
Keeping these factors in mind will go a long way toward helping you choose between the four data collection methods. The recent evolution of technology has given researchers powerful tools and dramatically transformed the ways that surveyors interface with survey respondents.

Chapter 3: Qualitative vs. Quantitative Research

Before you formulate your questionnaire, it is important to consider what type of information you'd like to collect — qualitative or quantitative. Both qualitative and quantitative research have their places in data collection.

Quantitative Research

Quantitative research (derived from the word "quantity") describes research that produces countable or numerical results.

Examples of Quantitative Questions

How long does it take you to travel to work?

- ◻ 0-20 minutes
- ◻ 21-40 minutes
- ◻ 41-60 minutes
- ◻ Over 1 hour

What forms of transportation do you use while traveling to and from work? Please select all that apply.

- ◻ Personal car or taxi
- ◻ Auto
- ◻ Rickshaw
- ◻ Bicycle
- ◻ Metro
- ◻ Other

Would you move to a new location just to decrease your commute time?

- ◻ Yes
- ◻ No
- ◻ Not applicable

Rate each form of transportation on a scale of 1-5.
(1 is strongly dislike, 2 is dislike, 3 is neutral, 4 is like, 5 is strongly like)

- ◻ Personal car or taxi
- ◻ Auto
- ◻ Rickshaw
- ◻ Bicycle
- ◻ Metro

Qualitative Research

Qualitative research describes research that produces non-numerical results. It generally investigates the "why" and "how" of your research question.

Examples of Qualitative Questions

Do you like your commute to and from work? Why?
How do you generally get to and from work?
Why is the metro your favorite form of transportation?
Is there anything else you'd like to tell us about your commute?

When to Use Qualitative and Quantitative Research

Qualitative research is often used as exploratory research. It is helpful to provide insights into the problem you want to research more, or it helps to identify ideas and hypothesis for future quantitative research. Qualitative research also is useful in learning more about the "why" and "how" behind your question.

Quantitative research is a great way to generate numerical data, create usable statistics, and generalize results or uncover patterns from a larger population.

To figure out whether you should use qualitative research, quantitative research, or a mix of the two, look at your research question, outcomes and indicators. (If you don't have these, go back to Chapter 1!)

Examples

Research question: Are the children in my classrooms improving?

Quantitative data:

- Children's test scores over time
- Children's grades over time
- Children's scores on an evaluation created for this research

Qualitative data:

- Parents' opinions on whether they think their children are improving (and why)
- Teachers' thoughts on whether they think their students are improving (and why)
- Students' feedback on whether they think they are learning more (and why)

Research question: Is my women's empowerment program making participants feel more independent?

Quantitative data:

- Ask participants to rank their independence on a quantitative scale before and after the program
- Ask participants if they feel more independent (Yes or No question)
- Ask participants how likely they are to engage in measures of independence (i.e. standing up to their husband, taking more control over household finances) on a quantitative scale before and after the program

Qualitative data:

- Ask participants how they feel after completing the program
- Ask participants about whether they think they are likely to engage in measures of independence (i.e. standing up to their husband, taking more control over household finances) before and after the program
- Ask participants' friends, husbands, and/or children about the participants' behavior before and after the program

As the previous examples show, many research questions can be answered using both quantitative and qualitative research. To decide which is right for you, think about your research question, what questions you need to answer, and the type of data that you are hoping to collect.

Chapter 4:

Choosing Your Survey Questions

Now that you are aware of the different elements of a questionnaire, the next step is to think about the various types of questions that you would want to ask in a given questionnaire. The below questions can help you decide which questions you should ask.

1. What Kind of Information Do You Need?

Different categories of information include: personal background (name, religion, age, caste, gender, etc.), education information, health information, government schemes subscription, etc.

For example, say that we are looking to measure the change in students' learning outcomes. We could decide that we need some personal details of the students (age and gender) as well as learning levels of the students, classroom activities of the teachers, and some school-level information. We would not need details on the religion or caste of the students, personal details on the teachers, or information about the students' families.

Tool Tip
To arrive at the different sets of information, put the outcome in the centre of a paper and write all the things that can possibly impact that outcome. Talk to your program officers and field staff about it.

2. What Information Can Be Easily Collected?

Personal information can be easily collected but BMI, height, weight, etc. might be difficult to collect. It is easy to ask someone their weight, but the accuracy of this data is often low. Measuring people's weight with a scale is far more accurate, but it is also more difficult.

Choose parameters that are useful and can be collected effortlessly.

For example, say that you want to judge a teacher's classroom skills. You might be tempted to capture a lot of information about the classroom — you can probably sit in the classroom and capture classroom activities for an hour. Or you could simply do a 5-minute observation to learn about what happens on a typical day. You need to balance the effort in collecting additional information and the value of that information.

Tool Tip
To arrive at the final data points, think of the following things: how difficult will it be to collect that information, how would respondents react to a particular question, and how quickly can you collect a particular piece of information?

3. What Information Is Actually Useful for the Organization?

It is tempting to collect all information that you can. But it is important to only collect information that is useful for the organization.

For example, say that you want to judge a teacher's classroom skills through a 5-minute observation. It would be easy to simultaneously collect other information on the school or students. However, don't collect information just because you can collect it! Only collect information that will help you with your analysis.

4. Did You Include the 5 Key Questions (Introduction, Identifiers, Consent, Open-Ended Fields and Validations)?

Always have the following questions in your questionnaire.

- Introduction: the right introduction to the survey can set the tone of the survey and is often helpful in making the respondent understand why the survey is crucial and how it will help her/him in return.
- Identifiers: name, age, father's name and location (for example).
- Consent: most surveys in India require organizations to seek the beneficiary's consent. It's a good ethical practice.
- Open-ended fields: ask for any information that might not be captured by the specific question types.
- Validations: information like GPS and time taken to validate whether the questionnaire was filled correctly.

By leveraging smartphone-based tools for data collection, you will be able to automatically capture GPS location and the average time taken for surveys. This will be helpful in creating a check to ensure your field surveyors are collecting accurate data from the ground.

For instance, if you are collecting data from households in a village, then possibly the GPS coordinates of each survey response should be a minimum of 15 metres apart. Similarly, if the average time taken for a survey is 20 minutes and one of your surveyors is submitting responses in under 5 minutes, this could indicate issues with his data quality and validity.

Example

Say that you want to measure the

improvement in student learning outcomes at a given school. Go through the four questions above to create the most important questions in your survey.

Question 1: What kind of information do you need?

- Reading level of the kids, to see improvement in learning levels
- Teacher effort aimed at reading, to see improvement in teaching skills
- Background information on teachers and students, to track improvement and changes
- School information, to explain differences in student learning outcomes in different types of schools

Questions 2-3: What information can be easily collected and is relevant to the organization?

- Reading level of the kids: use the ASER battery and sample a few kids from a classroom. Information on all the kids is not necessary.
- Teacher effort aimed at reading: observe classrooms for 5-10 minutes while the teachers are teaching reading skills.
- Background information: for students, only collect age, gender, class, and name; for teachers, collect only experience, classes and subjects taught.
- School information: collect information on number of students, teachers, teacher-pupil ratio, as well as average fees. Any other information that is irrelevant to the analysis should not be collected.

Question 4: Did you include the 5 key questions?

- Add introduction
- Add relevant identifiers for students (if not already covered in background information)
- Add consent
- Add an open-ended field for surveyor comments
- Add GPS and time stamp to validate the information from the field

Chapter 5: Choosing the Right Survey Question Types

Choosing a question for your survey is not enough. It is essential to choose the correct question type. A good question asked in the wrong way will not give you good data.

There are 8 main question types that you can use for surveys. Below are descriptions of each question type and when to use each.

Text

This is the most open-ended type of question. One can type in anything. It is ideal to use this type of question to collect a person's name or to collect qualitative information such as "Any other feedback".

Example: What is your name?

Dichotomous (Yes/No, True/False)

Dichotomous questions seek a binary response to a question.

Example: Do you love to read?
A. Yes
B. No

Numerical

This is used to capture numbers. Numerical questions should only be used to collect specific numbers that cannot be confined into certain ranges.

Example: How many times do you jog in a month? (numbers only)

Multiple Choice Questions

This is the most frequently-used question type. This can be single-select or multiple-select based on the need of the question. Multiple choice questions (or MCQs) are highly recommended, since they reduce the chances of capturing the wrong information.

Example (single-select multiple choice): What is your current education status?

- Uneducated
- Primary
- Secondary
- Senior Secondary
- Graduate
- Post-graduate
- Doctorate

Example (multi-select multiple choice): Which of the following subjects do you study?

- Maths
- Hindi
- English
- Science
- Social Science
- Environmental Studies
- Other

When creating options in an MCQ, it is very important that all options are mutually exclusive but collectively exhaustive. (See Chapter 7 for more information.)

Tabular/Roster

These are used to capture the same sets of information about multiple entities. For example, personal background about a household can be captured using a table.

	A	B	C	D
1	Name	Relationship with head	Age	Education
2				
3				
4				
5				

Scale

This question type is generally used to record preferences, opinions and ratings.

For example, you can use scale for capturing information on how good a particular class was. This question could use a 5-point scale question: Bad, Fair, Good, Very Good, Excellent.

Example: How was the food?
Poor, Average, Good, Excellent

Media Questions

Sometimes you want to capture information like pictures, audio, or drawings. If you are using mobile-based technology, you can use media questions to verify captured information.

Example: Take a photo of the anganwadi you visited.

Maps and Timestamps

If you are using mobile-based technology, it is also possible to capture geo-coordinates and time-stamps to enrich your data and make it more verifiable.

Example: Take the geo-location of the anganwadi you visited.

Chapter 6: Best Practices Around Writing Survey Questions

Chapters 4 and 5 talked about choosing the correct questions and question types for your survey. Here are a few additional tips to help you frame questions correctly.

1. One Question at a Time

Keep the questions simple, crisp, and to the point. Make sure that you are not asking multiple questions in one question – it might make the answer complex and confuse the respondent.

For example, see what is better:

Q1: Do you love to study? Why?

OR

Q1: Do you love to study?
A. Yes
B. No

Q2: If yes, why do you love to study?
A. To maximize knowledge
B. To learn new things
C. To score good grades
D. Other

Remember, the "length" of your questionnaire is not determined by the number of questions, but by the time taken to answer them. The second format of asking the same question breaks down the questions while simultaneously reducing the time taken to answer them. People who answer "No" in Q1 will not be asked the next question – reducing the time taken by the surveyor in explaining Q2. Breaking down the two questions also allowed us to turn Q2 into a closed-ended question, which reduced the time taken in answering Q2.

2. Beware of Subjective Questions

Use text/subjective questions only when there is no other suitable type of question you can use. Most subjective questions can easily be written as MCQs. The problem with subjective questions is that, if you let people input answers, the same thing can be said in many different ways.

Take the case of a village name: Gandipet. People can have many different ways of writing the same thing: Gandipetta, Gandipetu, Gandipettu, etc. Hence, it is always better to list out all possible options.

For questions like state/district/block/village name, list all possible options in the form of a list. For questions related to age, give a list of suitable ranges.

3. Ask Objective Questions

Do not include the answer in your question, as it will introduce surveying bias.

For example, rather than asking “Do you think India is on a downward trajectory?”, you should ask “What trajectory is India on? A. Downward B. Upward C. Other”.

4. Avoid Negative Questions

One of the best ways to be objective is by avoiding negative questions. Most positive questions are more direct than negative questions.

For example, rather than asking “What are the reasons India is not growing?”, ask “What are the factors affecting India’s growth?” This reduces bias and makes your question more objective.

5. Be Careful while Asking Sensitive Questions

Don’t ask any critical or sensitive information directly. People are often unwilling to share sensitive information with a third party.

For example, asking a question like “What is your income?” as a numerical question might result in dishonest answers by survey respondents. Asking this question as a multiple choice question by bucketing responses into different income brackets might result in more accurate responses.

For any sensitive questions, keep it under cover. For example, if you were to ask students whether they have cheated in an exam, it might be better to ask the question in a multiple-choice question that does not focus on cheating:

I have done the following with my school friends:
A. Played cricket/soccer after school
B. Been punished for coming to class late because I was playing
C. Shared answers in a test
D. Had lunch from another student’s tiffin box

While asking sensitive questions, it also helps to use the right words. In the previous example, “copying” and “cheating” were not used because they will make the question negative. Using more neutral words like “sharing” makes people more likely to answer honestly.

6. Don’t Ask for Too Much Detail

It is important to have the right amount of detail. Don’t dig deeper than needed, and do take only superficial information.

For example, if you want to know about the sources of energy at home, don’t ask about all the appliances used. Only ask whether specific energy sources (electricity, cowdung, gas, etc.) are being used or not. Make sure to cover all the major sources of energy.

Chapter 7:

The MECE Framework: Mutually Exclusive, Collectively Exhaustive Questions

Once you are clear about your research question and the type of data you will collect, the next step is to put together a questionnaire that can help you collect that data. Before putting together the questionnaire, it is important to understand the MECE framework.

MECE Framework

"MECE" stands for "Mutually Exclusive and Collectively Exhaustive".

When designing your questionnaire, it is important to ensure that all the different questions and sections are mutually exclusive and collectively exhaustive. "Mutually exclusive" means that no two questions should be repeated. "Collectively exhaustive" means that questions should be chosen in a way that captures all the required information.

It is important to note that the MECE framework applies to both questions and answer choices.

Examples

Consider the following two questions, and figure out whether they are MECE.

Q1: What is the educational status of all the members of the household?

Q2: Name the highest educated member here.

Answer: This is not MECE because we can capture the information needed in Q2 in Q1 itself. We don't need another question. Thus it is not mutually exclusive.

Consider the following answer choice, and figure out whether it is MECE.

Question: How many children do you have?

A. 1
B. 2
C. 3
D. 4 or more

Answer: This is not MECE. This answer choice covers all the positive values, but it doesn't give an option for 0. Thus it isn't collectively exhaustive.

Exercise

See whether the answer choices in the following questions are MECE or not:

Q1: What is your religion?

A. Hindu
B. Muslim
C. Christian
D. Sikh

Answer: No, these answers are not MECE. We haven't included Jainism, Buddhism, and several other religions. The choices are mutually exclusive but not collectively exhaustive. (See the Tool Tip below.)

Q2: Which bracket does your age lie?

A. 0-10
B. 11-20
C. 21-43
D. 44-80

Answer: No, these choices are mutually exclusive, but they are not collectively exhaustive. They don't cover the option for ages greater than 80

Q3. Which category do you fall in?

A. General
B. OBC
C. SC
D. ST

Answer: Yes, these answers are MECE. The answer choices are mutually exclusive (no overlap) as well as collectively exhaustive (covers all possible options).

Tool Tip
An easy way to ensure that a multiple choice question is collectively exhaustive is to add the option "Other". If the enumerator chooses "Other", you can ask the question "If other, please specify".

Chapter 8:

Closed vs. Open-Ended Questions

Every question on a survey will be either an closed or open-ended question. This means that closed and open-ended questions are at the core of your survey design. It is crucial to know the difference between closed and open-ended questions and when to use each.

The Difference Between Closed and Open-Ended Questions

Closed-ended questions have a defined, closed set of responses. This means that respondents only have a limited number of options for their answer to the question. Closed-ended questions come in a multitude of forms, but they are often in the form of multiple choice (single or multi-select), scale or dichotomous questions.

Here are a few examples of closed-ended questions:

How old are you?

- 0-10 years old
- 11-20 years old
- 21-30 years old
- 31-40 years old
- 41-50 years old
- Over 50 years old

Do you feel better today than yesterday?

- Yes
- No
- I feel the same

Are you pregnant?

- Yes
- No

How do you feel about this health program?

- Strongly dislike
- Dislike
- Neutral
- Like
- Strongly like

In contrast, open-ended questions do not have a defined set of responses. This means that the set of possible responses is infinite, so respondents can provide any answer they like. Open-ended questions are generally in the form of narrative or text questions.

Here are the previous examples, reworded as open-ended questions:

How old are you?

How do you feel today compared to yesterday?

Is it possible that you might be pregnant?

What do you think about this health program?

These are open-ended questions because the answers to these questions are not pre-determined, like they were previously. For example, in the first question, the respondent is not limited to 6 age brackets; they can answer with any number, or even ages like “8 years and 3 months”.

When to Use Open-Ended Questions

In general, open-ended questions are useful for qualitative research, learning more information about a topic, and surveys with small sample sizes. There are four main cases when open-ended questions should be used.

1. Preliminary Research

It is often helpful to conduct preliminary research to learn more about your problem before conducting your final survey. Open-ended questions are a key component of preliminary research, since you generally won’t know the answers that you’ll receive. Instead, you are looking to gain information that you likely don’t know.

For example, imagine that you want to improve your website. Before you can write an effective survey, it would be useful to get people’s general thoughts on your website. This will help you write a more targeted final survey.

Your preliminary research could use open-ended questions like:

• What do you think about this website?
• What are your favorite parts of the website?
• What are your least favorite parts of the website?
• What could be improved on the website?

Using the answers to these questions would help to write the final survey. For example, if many people said during preliminary research that they don’t like the colors on the website, you could include a section on your final survey where respondents rank different color palettes.

In addition, you can use preliminary research to improve the closed-ended questions in your final survey. Most surveys use closed-ended questions, but writing closed-ended questions requires knowing the possible set of answers to your questions. Often, you don’t know this before you start your survey. Preliminary research with open-ended questions is helpful to learning the set of answers for future closed-ended questions.

For example, imagine that you want to learn more about why people are not attending your meetings. It would be easy to analyse the results of a closed-ended question like:

Why did you miss the last meeting?

• It was too early for me to attend
• It was too late for me to attend
• It was too far from my house
• Other

However, preliminary research would be a great way to learn the full set of possible answers. For example, you could use an open-ended question (e.g. “Why did you miss the last meeting?”) to get more information on why people miss meetings. Then, once you understand the most common reasons, you can write a much better closed-ended question in your final survey.

2. Expert Interviews

Experts usually know more about a subject than you will, so it is useful to use open-ended questions to get as much information as

possible. Limiting experts to a pre-determined set of responses with closed-ended questions will be less productive than giving them the freedom to demonstrate their knowledge and talk at length.

3. Surveys with a Small Sample Size

For a large number of respondents, it can be difficult to read and analyse the answers to open-ended questions. Open-ended questions can often lead to responses of several sentences or paragraphs. Comparing these answers across dozens or hundreds of respondents is extremely difficult and time-consuming.

However, this becomes much easier if you're conducting a survey with a small sample size (e.g. under 20 respondents). For small sample sizes, open-ended questions are a great way to solicit more detailed information in a way that is still analyzable.

4. The End of Any Survey

The end of a survey is the perfect place to include an open-ended question. No matter how well designed a survey is, it can never account for all possible opinions and data. Including an open-ended question at the end of a survey — such as "Is there anything else you'd like to tell me?" or "Is there anything that I've missed?" — will allow respondents to share extra information, opinions, or concerns. Giving respondents the freedom to include additional information or comments is also a good way to show respect for the time and effort they took in completing your survey.

When to use Closed-Ended Questions

Closed-ended questions should be used for easier analysis and reporting of the data you are collecting.

For example, imagine that you are polling 1,000 people about their internet usage. If you ask the open-ended question "Tell me about your internet usage?", you will end up with 1000 unique responses that cannot easily be analysed or reported. Instead, if you use a closed-ended questions like the one below, you will be able to better understand and report the results.

On average, how many hours do you use the internet per week?

- 0-5 hours
- 6-10 hours
- 11-15 hours
- 16-20 hours
- Greater than 20 hours

With a closed-ended question, you can easily analyse the data and report a clear result like "63% of respondents use the internet less than 5 hours per week".

Tool Tip
In general, qualitative research will use open-ended questions and quantitative research will use closed-ended questions. See Chapter 4 for more details on qualitative and quantitative research.

Chapter 9:
Sampling Your Population

Ideally, a survey should gather data on every single person in the target population. For example, a survey about learning outcomes

at a small school could track the test scores of every student. Collecting data on everyone in the target population is the best case scenario, since it ensures that everybody who matters to the survey is represented accurately.

However, this is only possible if the population is small enough and the researchers have sufficient resources to reach out to everyone. This often is not the case, so researchers have to identify a subset of the population to survey.

How you choose this subset of the target population is crucial to the quality of your data. The group must be carefully identified and representative of the larger population, else your data will not be useful for drawing inferences.

If done right, survey sampling can save time and money while allowing you to draw interferences about a large group of people.

3 Things to Keep in Mind While Choosing a Sample Population

1. Consistency

It is important that researchers understand the population on a case-by-case basis and test the sample for consistency before going ahead with the survey. This is especially critical for surveys that track changes across time and space. If your sample is consistent, you can be confident that any change in the data reflects real change across the population, rather than change across atypical individuals in the population.

2. Diversity

Ensuring diversity of the sample is a tall order, as reaching some portions of the population and convincing them to participate in the survey can be difficult. However, for a sample to truly represent the population, the sample must be as diverse as the population itself and sensitive to local differences.

3. Transparency

There are several constraints that dictate the size and structure of the population. It is imperative that researchers discuss these limitations and maintain transparency about the procedures followed while selecting the sample, so that the results of the survey are seen with the right perspective.

Choosing Your Sampling Technique

Probability Sampling

For probability sampling techniques, each person in the population has a defined, non-zero probability of being included in the sample. Probability sampling provides the most valid or credible results because it reflects the characteristics of the population from which they are selected. There are three probability sampling methods: random sampling, systematic sampling and stratified sampling.

Random Sampling

When: There is a very large population and it is difficult to identify every member of the population.

How: The entire process of sampling is done in a single step, where each subject is selected independently of the other people in the sample.

Pros: In this technique, each member of the population has an equal chance of being selected for the sample.

Cons: When there is a very large population, it is often difficult to identify every member of the population so the pool of subjects can become biased. For example, dialing numbers

from a phone book may not be entirely random since the numbers would correspond to a localized region.

Use case: Want to study and understand the rice consumption pattern across rural India? While it might not be possible to cover every household, you could draw meaningful insights by building your sample from randomly-selected districts or villages.

Systematic Sampling

When: Your given population is logically homogenous. This means that they all share a characteristic that is important to the survey. For example, suppose a supermarket wants to study the buying habits of their Sunday customers. The customers who enter the supermarket on Sunday are a logically homogeneous population since they share 2 key qualities: “customers of the supermarket”, and “visited the supermarket on Sunday”.

How: Arrange the elements of the population in some order and select terms at regular intervals from the list.

Pros: Systematic sampling is far simpler than random sampling, and it ensures that the population will be evenly sampled. In random sampling, there is a chance that the sample might include a clustered selection of subjects. This can be avoided through systematic sampling.

Cons: The possible weakness is an inherent periodicity of the list (i.e. if the people you are surveying are already ordered in a certain non-random way). This can be avoided by randomizing the list of your population entities, as you would randomize a deck of cards for instance, before you proceed with systematic sampling.

Use Case: Continuing with the earlier example, the supermarket can use systematic sampling to study the buying habits of their Sunday customers. They can choose every 10th customer entering the supermarket and conduct the study on this sample.

Stratified Sampling

When: You can divide your population into characteristics of importance for the research.

How: A stratified sample, in essence, tries to recreate the statistical features of the population on a smaller scale. Before sampling, the population is divided into characteristics of importance for the research. For example, by gender, social class, education level, religion, etc. Then the population is randomly sampled within each category or stratum. If 38% of the population is college-educated, then 38% of the sample is randomly selected from the college-educated subset of the population.

Pros: This method attempts to overcome the shortcomings of random sampling by splitting the population into various distinct segments and selecting entities from each of them. This ensures that every category of the population is represented in the sample. Stratified sampling is often used when one or more of the sections in the population have a low incidence relative to the other sections.

Cons: Stratified sampling is the most complex method of sampling. It lays down criteria that may be difficult to fulfill. This can place a heavy strain on available resources.

Use Case: If 38% of the population is college-educated and 72% of the population have not been to college, then 38% of the sample is randomly selected from the college-educated subset of the population and 72% of the sample is randomly selected from the rest of the population. Maintaining the ratios while selecting a randomized sample is key to stratified sampling.

Non-Probability Sampling

For non-probability sampling, the sample is constructed with no probability structure. The selection is not randomized, so the resulting sample is not fully representative of the target population. There are three non-probability sampling methods: convenience sampling, snowball sampling and quota sampling.

Convenience Sampling

When: During preliminary research efforts.

How: As the name suggests, the elements of such a sample are picked only on the basis of convenience in terms of availability, reach and accessibility.

Pros: The sample is created quickly without adding any additional burden on the available resources.

Cons: The likelihood of this approach leading to a sample that is truly representative of the population is very poor.

Use Case: This method is often used during preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample. For example, interviewing 10 of your friends over the phone would count as convenience sampling.

Snowball Sampling

When: When you can rely on your initial respondents to refer you to the next respondents.

How: Just as the snowball rolls and gathers mass, the sample constructed in this way will grow in size as you conduct the survey. At the end of the survey, you ask your initial respondents to refer you to other people to survey.

Pros: Though the costs associated with this method are significantly lower, you will still end up with a sample that is very relevant to your study.

Cons: You restrict yourself to only a small, homogenous section of the population.

Use Case: Snowball sampling can be useful when you need the sample to reflect certain features that are difficult to find. For example, to conduct a survey of people who go jogging in a certain park in the mornings, snowball sampling would be a quick, accurate way to create the sample. You can find someone who jogs in the park in the morning, then ask them to refer you to their friends who also jog in that park in the morning.

Quota Sampling

When: When you can characterize the population based on certain desired features.

How: Quota sampling is the non-probability equivalent of stratified sampling. It starts with characterizing the population based on certain desired features and assigns a quota to each subset of the population.

Pros: This process can be extended to cover several characteristics and varying degrees of complexity.

Cons: Though the method is superior to convenience and snowball sampling, it does not offer the statistical insights of any of the probability methods.

Use Case: If a survey requires a sample of fifty men and fifty women, a quota sample will survey respondents until the right number of each type has been surveyed. Unlike stratified sampling, the sample isn't necessarily randomized.

Tool Tip
Probability methods are clearly more accurate but the costs can be prohibitive. For the initial stages of a study, non-probability methods might be sufficient to give you a sense of what you're dealing with. For detailed insights and results that you can rely on, move on to the more sophisticated probabalistic methods as the study gathers pace and takes a more concrete structure.

Minimizing Sampling Error

There is one easy way to minimize sampling error – increase the sample population size. The more respondents you have, the more accurate your survey will be. However, it isn't always possible to increase the sample population because of financial restrictions.

Avoiding three common errors will help to minimize sampling error without increasing your sample size.

1. Avoid Population Specification Errors

A population specification error occurs when a critical segment of the population is not included in the sample. This is the result of a knowledge problem or gap. The results of a survey with a population specification error will shed some light into your issue, but they cannot provide the full picture.

For example, imagine that you want to learn more about household decision making, so you survey men about the decisions in their household. This would be correct if only men make household decisions, but often women and children also have influence over decisions. By only surveying men, you will miss out on part of the picture.

An easy way to avoid population specification errors is to learn how similar surveys sampled their target population. By checking on other surveys, you can be sure that you are not forgetting a critical segment of your sample population.

2. Avoid Sample Frame Error

Sample frame error occurs when a survey samples the wrong segment of the total population, usually because the surveyor has missed a new trend or change in their target population.

For example, imagine that you want to learn how attendees feel about your program, and you use the attendance sheets from two weeks ago to create a target population. However, unknown to you, a new group of people started attending your program one week ago. The results of your survey will be misleading, since they do not include one of the key segments of your target population.

An easy way to avoid sample frame error is to take plenty of time to study your target population. Be sure that nothing has changed about it recently, and be sure that you have accounted for all types of people in your sample.

3. Avoid Non-Response Error

It is normal for some targeted respondents to not respond to a survey. However, this can become a problem if the non-respondents generally hold a view that is different from the respondents. As a result, the final data will be skewed toward the opinions of those who responded.

For example, imagine that you want to survey housewives about their free time, and you do this by calling them on the phone during the day. The women who do not respond are the ones who have less free time (since they don't have enough time to pick up the phone). Meanwhile, the women who respond are the ones who have more free time. The results

of the data will report higher free time among housewives than is actually true.

An easy way to avoid non-response error is to conduct follow-up surveys or contact those who did respond through alternative means. In addition, keeping the questionnaire short will help encourage people to respond the first time.

Calculating Sample Size

Determining the size of your sample population is one of the most difficult decisions to make in your survey. A larger sample can yield more accurate results — but the more responses you collect, the more expensive it gets.

Statistical Variables

To calculate the best sample size for your needs, you need to make 2 decisions about how accurate you want your data to be.

1. Margin of error
(also known as confidence interval):

No sample will be perfect, so you need to decide how much error you are willing to allow. The margin of error determines how much variance you want to allow in your data.

For example, if you set a margin of error of 5% and find that 68% of women take iron pills, then that means that the real percentage of women who take iron pills is between 63% (68-5=63) and 73% (68+5=73).

2. Confidence level:

How confident do you want to be that the actual number falls within your margin of error? The most common confidence intervals are 90%, 95%, and 99% confident.

For example, imagine that you set a margin of error of 2% and a confidence level of 95%, then you find that 68% of women take iron pills. You can be 95% confident that the true number of women who take iron pills is between 66% and 70%.

The confidence level corresponds to a given z-score:

	Confidence level	Z-score
1	90%	1.645
2	95%	1.96
3	99%	2.576

Sample Size Formula for a Population of Unknown Size

Often, a population is too large to easily measure. When this is the case, you can use a sample size formula that does not account for the size of the large population that you are surveying.

The sample size formula when you don't know the size of your large population is

$$n = \frac{z^2}{4m^2}$$

where *n* is the sample size you should use, *m* is the margin of error, and *z* is the z-score. (Note that if you have a margin of error of 5%, *m* = 0.05)

It is usually safe to use a margin of error of 5% and a confidence level of 95%. If you plug these numbers into the formula, you get a sample size of 384. This means that 384 is a safe sample size for a large population of unknown size.

Sample Size Formula for a Population of Known Size

You can calculate a more accurate sample

size if you know the size of the population that you are surveying.

For example, if you are studying the learning outcomes for a school with 200 students, your population size is 200. If you are studying women in Gujarat, the population size is the total number of women in Gujarat.

This does not have to be exact. Even an estimate of the population size will result in a better sample size than using the formula above.

The sample size formula when you know the size of your population is

$$n = \frac{(pz^2)}{4m^2(p-1)+z^2}$$

where *n* is the sample size you should use, *p* is the size of the population being surveyed, *m* is the margin of error, and *z* is the z-score. (Note that if you have a margin of error of 5%, *m* = 0.05, z=1.96)

It is usually safe to use a margin of error of 5% and a confidence level of 95%. The sample size formula with those figures (margin of error of 5% and confidence level of 95%) is

$$n = \frac{384.16p}{p + 383.16}$$

where *n* is the sample size you should use, and *p* is the size of the population being surveyed.

Tool Tip
Don't know which sample size formula to use? The last formula (for a sample of known size with margin of error of 5% and confidence interval of 95%) is your safest bet.

References

1. Introduction - Case Study
Details on polling during the 1936 U.S. Presidential Election from
Dennis DeTurck, "Case Study 1: The 1936 Literary Digest Poll", University of Pennsylvania
(https://www.math.upenn.edu/~deturck/m170/wk4/lecture/case1.html)

2. Chapter 9 - "Calculating sample size"
Sample size formulas derived from Glenn D. Israel, "Determining Sample Size",
Fact Sheet PEOD-6, University of Florida, 1992
(http://sociology.soc.uoc.gr/socmedia/papageo/metaptyxiakoi/sample_size/samplesize1.pdf)

Bonus Content

1. Designing a Great Survey

Surveys are the bedrock of data-driven research. The quality and reliability of your data and, by extension, your entire project hinges on the effectiveness of your survey. Get it wrong and you'll be left wading through the murky marshes of meaningless data that make little or no sense, with an immense investment of time and effort wasted. Going ahead with an undercooked survey also reflects poorly on your organization. The people who have been through your earlier survey will be less than accommodating when you reach out to them next time even if you return with a new and improved version.

It is important then to get your survey right the first time, and doing that requires a concerted and conscious focus on thought and planning. Don't let this be a daunting proposition though; think of it as an opportunity. Good data means good results, and if your survey has been designed well, half your work is done right there. It's easy, really. Let's walk through the eight steps to designing an ideal survey. For ease of understanding, we'll consider these steps with respect to an impact assessment for a self-help group.

1. Set your goals

Defining the purpose of your survey in clear, unambiguous terms is absolutely vital. It sets the direction for everything you do. Coming back to the drawing board and reminding yourself of the purpose of the survey can be a good way to get back on track when the team feels like it's stuck in a rut and is unable to draw inspiration to continue driving forward. The goals of the survey also dictate all the other aspects of survey design to a large extent.

Example

An example of a clear, constructive goal could be "assess the impact of microfinance on people living in a certain district". As we move through the rest of the steps, we'll see how this goal lends itself to every aspect of the survey.

2. Narrow down on your target population

In order to collect data that is relevant to the purpose of your study, it is important that you reach out to the right people. Identifying the right sample for your survey is another critical aspect of survey and bears heavily on the structure and mode of survey that you choose to employ. The nature and language of the questions used while formulating the questionnaire also depend heavily on the target population.

Example
Selection of the target population follows largely from the goal that you have selected for your survey. The target in the case of the example we've taken here would be the working population of the district between 18 and 60 years of age.

3. Structure the survey

Dividing a questionnaire into categories results in an intuitive structure that is easy for participants to navigate. The researcher can also improve the survey experience further by providing additional explanation at the beginning of each section. This gives the respondent an idea of what to expect. Categorizing the questions while designing the survey, even before you get down to writing the questions, helps maintain the focus on the different research objectives and ensures a balanced output.

Example
Dividing the questionnaire, for instance, into sections on Particulars (personal background information) and Impact Assessment (Social, Educational and Cultural) will lend a logical flow to the survey that makes it easier to grasp. This will also make it easier for you to assign categories to the data that you collect and simplify the analysis process.

4. Select the mode of your survey

The mode that you employ to administer the survey depends on the sample type and size. Use the mediums that are the most effective in reaching your target population. The time span of the research is also an important factor that impacts this decision. Leveraging technology to maximize the extent and depth of your reach might come in very handy in such situations. SocialCops has had great success in using mobile application and low-cost smartphones to collect data at the grassroots level.

Example
Selecting the wrong mode for your survey can cripple your research. Launching a web-based survey that needs people to visit a website in order to answer the questionnaire, for instance, will be nearly useless while trying to reach people in villages that have limited or no internet access. The mode of the survey should use tools and infrastructure that can easily reach your target population and account for the comfort level that your respondents have towards the technology that you employ.

5. Choose the right question type

Using the right tool for the right job is essential in any endeavor. Questions are the tools of your survey and picking the wrong question type can be as awkward as using a screwdriver to knit a pullover. Throw in a good mix of closed-ended questions – dichotomous (yes/no), multiple choice, and ordinal scale (rank, preference) – after considering the purpose that each question type will serve. Top it up with open-ended questions where necessary. Read more about how to optimize your survey quality by choosing the right question types.

6. Formulate the questions

Words mean different things to different people and taking care of some of the finer nuances involved in formulating an effective question can go a long way. Don't leave any scope for ambiguity. Be clear about what you need and get your questions proofread by somebody who is not familiar with your study before sending out the survey. Brevity is essential. Respondents are more likely to respond positively to questions that are concise and able to hold their attention. Avoid unnecessary jargon; the language should be as simple and generic as possible. And finally, the answer choices must be well defined.

Example
If you ask respondents to rank their level of satisfaction on a scale from 1 to 5 but fail to explain whether 1 is very satisfied or 5 is very satisfied, their responses will be of little value.

7. Introduce the survey

Introduction is often considered to be the most critical part of the survey, as a majority of the respondents make up their mind about whether they would like to answer the survey after going through the introduction. The introduction, thus, needs to make a good, strong first impression. It sets the tone for the rest of the survey and lays down the context in a simple, understandable way. Begin with a statement thanking the respondents for their time and explain the subject of the study along with a confidentiality statement to address privacy concerns. Mention the expected time required to complete the survey and display the percentage completed as the respondent moves ahead.

8. Take the field

Once the survey is ready, execute your plans through a robust collection mechanism. If the survey is being conducted in person, make sure the people who administer the survey completely understand the purpose of the survey and are comfortable conversing with the population they're supposed to interact with. Don't forget to train them on the use of open-ended and unstructured questions and brief them about the technicalities of the survey. Before you head out to launch your survey in the field or online, test the survey with a small control group to see if everything functions the way you expect it to.
Eight simple steps and you're all set to go! If you have designed your survey well and executed your plans to perfection, good, clean data filled with tremendous potential for gathering insights will start surging in.

2. Piloting a Survey

Collecting primary data at any scale is challenging. Though data quality can be difficult to measure, it is crucial to ensure that you are not wasting time on poor quality data. Creating a good survey is one of the best ways to ensure data quality. A bad survey will only lead to bad data, and thus bad inferences.

How can you ensure that your survey will collect relevant, accurate, useful data?

Pilot your survey. In a pilot, you can test all aspects of your survey — question flow, order, language, etc. — before you use the survey to collect real data. Piloting helps you identify and fix issues that would have led to poor quality data. A pilot is like putting your survey through a simulator to understand what is right and wrong.

Piloting can be a time and energy-intensive process. However, it can also be fun, since piloting leads to unparalleled levels of learning!

The goal of piloting

Piloting should help you answer the following questions:

- Am I catering to my audience?
- Will my data collectors be able to seamlessly collect data using this survey?
- Does my survey format cater to my needs? Is it capturing too much or too little?
- Is my survey collecting the data in the correct format?

In addition, piloting can help you understand implementation hassles. By testing out your survey in advance, you can predict any problems that might arise during the actual roll-out. For example, a pilot can help you learn how surveyors should be traveling from place to place, whether you should inform respondents in advance, what your surveyors should carry with them, and more.

Piloting can be full of sleepless nights and stressful days! Here are 10 tenets of piloting to help the process go smoother. We developed these tenets from our experiences rolling out all kinds of survey — large and small, long and short, quantitative and qualitative.

1. Ensure that you have done enough secondary research

You don't have to learn everything about your survey in the field. A pilot will be more successful if you first check how others have conducted similar surveys. A good best practice is to read at least 10 similar surveys. These can be found in research papers, company websites, and ebooks. Learning from existing data collection materials will give you a great head start.

2. Take feedback from your organization

Present your draft survey to lots of people for feedback. Some of the basic issues or changes can be identified by people within your organization. You can also speak to people who have done similar exercises before or reach out to a few experts in other offices.

3. Choose a representative sample for piloting

During your pilot, you should survey a sample of your final audience. Don't choose this sample randomly. Choose each person in your pilot deliberately.

Make sure to consider the following factors for each of the participants in your pilot:

- Age
- Gender

- Education status
- Income status
- Caste (in the Indian context)
- Geography

You can also read this resource to learn more about how to build a rigorous sample.

4. Pilot the survey in the correct medium

Ensure that you pilot the survey in the medium it will roll out in. Do not do any of the following:
Pilot a paper survey if you will actually collect data on a mobile app.
Pilot an English survey if you will actually collect data with a Hindi survey.
Pilot a small survey of 10 questions if you will collect data with a long survey of 100 questions.

5. Integrate the pilot with your training

Usually, organizations conduct surveyor training before data collection starts. Use the pilot as part of this training. The pilot will give the surveyors a chance to test out the survey in the field, ask questions about the survey, and figure out if they have any issues.

6. Question the survey questions

The most important part of the format is to understand what questions are you asking and what information you are getting.
One of the problems to look for during your pilot is compound questions — a question that asks two or more questions at the same time. For example, “How many children do you have and which one is the youngest?” is a compound question. It actually is asking two separate questions: “How many children do you have?” and “Name the child which is the youngest.” Be sure that each question on the survey only asks for one piece of information.
In addition, the pilot will show which phrases in your survey are widely understood and which ones are vague terms or jargon. Be sure to eliminate any vague words or jargon from your survey, no matter how obvious they might seem to you.

7. Examine the data being collected

Don’t just focus on the questions themselves; be sure to also look at the data that is being collected. Sometimes, issues in the collected data can show problems in your survey. Use your data to make a data-driven decision on what’s working and what’s not working.

8. Pilot all aspects of the survey

Pilot the question flow, the order of the questions, question types, and even the help text to ensure that the survey is clear to both your surveyors and respondents. To pilot question flow and order, you should observe how comfortable surveyors and respondents are as they go through the survey. Check whether the surveyors or respondents get confused or give the same information twice. If this happens, it might be helpful to change the order of the questions or add help text.

9. Collect feedback from everyone

Since piloting is an iterative process, it is important to ensure that you include viewpoints from all stakeholders. Get feedback from supervisors, surveyors, observers, respondents, and any other people in the pilot, and make sure everyone contributes during feedback sessions. In addition, you can observe different people to compile your own feedback.
Consider and balance all these viewpoints carefully. It is important to make sure that you don’t get biased by any single opinion.

10. Never re-pilot the same version of a survey

At the end of each day, incorporate all the

feedback into your survey. Then you can use this updated survey for the next day of piloting. By immediately updating your survey, you won't waste time on the same feedback.

11. Be thoughtful about what changes to make

You don't need to include every piloting insight. Sometimes, you will get bad feedback if a surveyor misunderstands something or the sample respondents are not chosen correctly. It is important to remain critical and make sure you are only incorporating the right feedback — something that will drastically improve the survey for everyone (rather than just one or two surveyors) and will not make the survey more difficult. A good practice is to make sure that a significant number of people (more than 60%) agree to the survey changes.

A typical pilot day

Morning session: Make sure that your surveyors know the purpose of the survey and what you are testing before they collect data. For example, tell them to think about which questions to keep and which to remove, which questions should be asked first, how the questions can be worded better, whether help text should be added, and how the respondents reacted, how interesting the survey was, etc. If you tell your auditors what you are looking for, you will get better responses during the evening session.
Day monitoring: While the data collectors are collecting data, you shouldn't sit back and relax. Shadow them to understand and observe the entire exercise. Don't correct them when they go wrong; just record the mistake.Be sure to follow diverse people and watch several surveys end to end (if the survey is a reasonable length).
Evening session: Once everyone is back from the field, gather everyone's viewpoints on the survey. It's ideal to go over the entire survey, question by question.
Night session: Assemble the day's feedback, analyze what feedback to include, and update the survey accordingly.

That's how you pilot like a boss!
Happy data collection.

3. Field Data Collection Plan

Once you have created a great survey, you need to deploy it in the field. Field data collection is a complex process that often requires lots of time, money, and people. Deploying your survey in the field effectively will not only help save these resources, but it will also make your data more reliable.
This plan covers the 5 steps required to deploy a survey in the field. (The process of deploying a survey in the field is known as a field plan.) Though this content was written with mobile-based data collection in mind, all types of surveys and organizations can benefit from the steps and best practices below.

Step 1: Identify your resources

The first step of creating a field plan is to estimate the resources you have for your project. The main resources to consider are budget and time.
Budget and time will determine how many responses you should collect, how many surveyors you should recruit, and how much you should spend on resources like devices, documents, projectors, refreshments, training logistics, salaries, and more.
Your Scope of Work (SOW) will help determine the budget and time available for your project. The SOW is a formal document that describes the work activities, deliverable, timelines and milestones, pricing, quality requirements, governance terms and conditions, etc. The SOW should also outline all the parties involved in the project, the budget, and the timelines.

Step 2: Recruit your field staff

Your field staff are crucial — they are the eyes and ears on field, as well as the people collecting every piece of data. That is why it is vital to find reliable and trustworthy field staff.

Before recruiting field staff, it is important to consider their responsibilities.

What should they know in advance? How will they be trained? How will they collect data? Who will be responsible for monitoring the data collection process? Once you answer these questions, you will be one step closer to hiring a great team of field staff.

Next you must create a fixed hierarchy.

Creating a fixed hierarchy ensures that a clear reporting and management system is in place, project timelines are being met, and field staff are supported at every step of data collection. The most common hierarchy uses three types of people: Field Managers, Monitors, and Surveyors.
Field managers lead the entire team working in the field. They supervise and oversee tasks of field employees, run training programs, and ensure that everyone works as effectively as possible.
Monitors directly manage and support the surveyors. It is usually best to recruit 1 monitor for every 10 surveyors.
Surveyors collect data from the field by directly interacting with the sample population and recording survey responses.

The last step is to recruit field staff at all levels.

When recruiting field staff, consider the following four factors:
1. Education qualifications
2. Ability to handle technology
3. Previous experience
4. Willingness to learn new things

Step 3: Create a plan

An implementation plan is the complete plan for training and data collection. Having a plan in place ensures that everyone is on the same page, which makes data collection go more smoothly.

First, estimate the number of days required for training field staff.

Ideally you should not train more than 50-60 people in one room at one time. Based on how complex and long your survey is, you could do a one-day training, two-day training, or trainings spread out over multiple days. By multiplying the duration of training (in days) and number of batches (with each batch 50-60 staff members), you can calculate the number of total days needed for training.
Tip: If you are short on time, conduct parallel trainings of multiple batches.

Second, outline the training.

Your training will have two main components:
1. Understanding the technology
2. Understanding the questionnaire
Assess how complex each of these components will be, then accordingly estimate what materials and people will be required for each training.

Third, estimate the number of days for data collection.

You can do this by calculating the following:
1. Total number of responses required (also known as the sample size)
2. Total number of surveyors
3. Average time taken to fill each survey
Once you know these three numbers, you can calculate the number of surveys each surveyor can complete in one day. It is equal to the length of the day (most field surveys run for 8 hours per day) divided by the average time per survey.

Multiply the number of surveyors by the number of surveys per surveyor per day, then divide the sample size by this number. Add the days added for re-surveying and you will have the total number of days for data collection. The number of surveyors and total number of days of data collection are a trade off. Hiring more surveyors will decrease the time needed for data collection. However, training time and total field staff will need to increase, which requires more resources. In contrast, hiring fewer surveyors will increase the time needed for data collection.

Fourth, create a daily schedule for your field staff.

Here is an example schedule:
- **10 am:** Morning session
- **11 am:** Survey starts
- **2 pm:** Lunch time
- **3 pm:** Survey restarts
- **6 pm:** Evening Session

It is a good practice to include morning and evening sessions as a part of your team's daily schedule. Morning sessions allow for a quick check-in with the team on how they are feeling, their goals for the day, and the progress of their project. Evening sessions allow for feedback and review. A good practice is to ask the team how their day went, if they faced any challenges, and if they have any feedback for the questionnaire or data collection app during the evening session.

How to calculate your sample size

Determining the size of your sample population is one of the most difficult decisions to make in your survey. A larger sample can yield more accurate results — but the more responses you collect, the more expensive it gets.
The best way to calculate a sample size is to use the size of the population you are surveying. For example, if you are studying

the learning outcomes for a school with 200 students, your population size is 200. If you are studying women in Gujarat, the population size is the total number of women in Gujarat. This does not have to be exact.

Step 4: Train field staff

After creating an implementation plan, you need to prepare for and conduct training for your staff.
Important preparations include the following:

- Phones/tablets should be procured and given to the field staff.
- Make field staff aware of the fact that training will happen on tablets.
- Create a plan for setting up the training material.
- Field staff should be aware of the purpose of the training.

The best way to conduct training for mobile-based data collection is to divide the training into three parts: technical training, survey training, and data collection training.

Technical training focuses on learning the mobile-based data collection technology. Before conducting the technical training, it is helpful to go through the data collection app yourself. Be sure to try the full functionalities of the app – make a questionnaire from scratch and fill out an entire questionnaire to understand how the application functions.

Survey training should happen only once your field staff is comfortable with the device and the app. Then the next step is to train the staff about the survey itself. Survey training should focus on explaining the logic behind the survey questionnaire and how to ask the different types of questions on the survey.

Make sure you cover the following:
1. If you have multiple questionnaires, explain the purpose and types of different questionnaires. Suppose you have two questionnaires – baseline and monitoring. Tell field staff why baseline data is important and how it can help make the program better. Then tell them what monitoring is and how it will be helpful.
2. Explain survey details (such as the average time taken, flow of the questions, potential pitfalls, etc.) to the field staff so that they can best collect data.
3. Run the field staff through each question to explain how they should be asking it. Some questions might be tricky, especially if the question is difficult to understand for the respondent.

Survey training is also a good time to get feedback on the questionnaires, so be sure to ask field staff if there are better ways of asking each question.

Data collection training covers the best practices of data collection. This helps to improve the quality of data collected by surveyors.

Best practices include the following:
1. Do not prompt answers to respondents. Only state the possible options.
2. Do not repeat questions beyond a certain point.
3. Let the respondent take his/her own time.
4. Follow the sequence of the questions. Don't skip questions based on your judgment.
5. Read out all the options correctly to the respondents.
6. While moving from one section of the questionnaire to another, give the respondent a heads-up on the upcoming questions.
7. Data collection needs to be unbiased (recorded without personal bias), time-bound (recorded in estimated time), and as per the format of survey.

In addition, explaining the entire project to surveyors can help encourage them to collect good data. Some points to cover include:

- The purpose of the project.
- Surveyors' role in the larger picture of informing program's decisions.
- The fact that the organization trusts

surveyors with the important task of data collection.

- The project uses mobile-based tools to help make data collection easier, quicker, and more reliable for surveyors.

Story from the field

Pramila was a volunteer with Swades Foundation and was helping with the data collection process. She dropped out of school after clearing grade 7. When she entered the training workshop, she was very apprehensive of using a tablet to collect data in different forms — typing up text, capturing audios and videos, and mapping households.

After four hours of training, Pramila was able to not only use the tablet well, but she was also able to use Collect (our data collection app) well.

During this training, there were many Pramilas in the room who learned to use the app in just four hours.

This happened because the training was well-planned, interactive, inspiring and well-executed. You can also inspire your field staff and kickstart data collection by planning your trainings well and communicating project rationale and learning outcomes for both the project and the team. In addition, sharing success stories like that of Pramila can help field surveyors feel confident in their ability to contribute to the project.

Step 5: Monitor your team

The final step is monitoring the team while they collect data.

Observe surveyors on the ground. While your surveyors are piloting or collecting data, observe them to learn common mistakes, doubts, and best practices. You can share these learnings in the morning or evening sessions so everyone can benefit from them.

Monitor the data that is being collected. Regularly monitor the data that is coming in to check for any inconsistencies or problems.

Take feedback from your surveyors. Take feedback on their experience, if they are facing any challenges, or if they require any additional resources.

Give feedback to surveyors. Give constructive criticism to surveyors as you observe them on the ground. Don’t be mean, be specific, and be sure to explain the rationale behind your feedback and how each person can improve.

4. Census Survey

SocialCops recently worked with Tata Trusts as the data intelligence partner to conduct two census surveys in Chandrapur (Maharashtra) and Noamundi (Jharkhand). Chandrapur and Noamundi are some of the most backward areas in their states. The areas have difficult terrain with little or no access to phone connectivity and electricity.
The large-scale census surveys collected data for over 60,000 households within 2 months using tablets. The data was then used to help in micro-planning activities in the area. SocialCops' challenge was to train local people to collect this data while overcoming operational challenges that arose throughout the survey.

These are some of our major learnings from these two surveys.

1. Internet — that thing of tomorrow

When surveying in tribal areas of Maharashtra and Jharkhand, we often couldn't find cellular connection, let alone internet, for hours. Sometimes we had to travel up to 8 hours to reach areas with cell or internet connectivity. We used our Collect app to collect data offline throughout Chandrapur and Noamundi. Collected data was synced to our servers at the end of the week.

2. Gender neutrality — the magic wand that eases social norms

It's interesting how we had to anticipate social dynamics since they could intervene with our work. In one village, our field manager caused problems because he spoke to a woman. This caused problems since men were not supposed to speak to women in that village. It is always a good idea to have a gender-balanced team of surveyors and monitors. This will help to navigate most social norms.

3. Villages — you can't see me

We used Census village lists to decide which villages that we should survey. However, we were sometimes left searching for villages that didn't exist.
Though it is standard practice to trust Census records, try asking local coordinators for a list of villages.

4. Piloting — a never-ending story

Each geography is unique with its own set of demographic, geographic, social, and economic problems. It's important to tweak each survey for its specific geography. Even after consulting experts and doing two rounds of piloting, surveyors were still calling our team for changes in the survey.
With Collect, we can edit surveys in the field and sync those changes on all other devices. This helped us tweak our surveys on the go.

5. Training — a leaky, indirect system

Any large-scale census survey usually involves a train-the-trainers model. This means that we train people who then train the surveyors, rather than us training the surveyors directly. This indirect method can lead to knowledge leaks as information is transferred from person to person.
Closely monitor surveyors' training, since that training can make or break a survey. A good training can reduce survey bias, while a bad training leads to incorrect data.

6. Google Play Services — the jugaad method

It can be difficult to get enough bandwidth to

download a phone app while working in rural India. However, all the tablets used in our census surveys needed Google Play Services (used for updating the phone's location, scanning barcodes, and more), which required 46 MBs to download. How could we do this for 350 tablets?
Instead of trying to directly download the file to the tablets, we used our own jugaad. We first downloaded Google Play Services to our computer, then we transferred the file to all 350 tablets over USB. If we later found a tablet without Google Play Services, we used ShareIt to share Google Play Services over Bluetooth.
Even basic functions like downloading apps might not work in certain areas. Be ready to do some extra work.

7. Survey tablets — collateral in disguise

If you think the fact that you are paying surveyors gives you leverage, think again! Our surveyors didn't let us collect our tablets until their checks were processed — something we didn't factor in as we planned our timelines. Make sure to schedule a time for payment and pay accordingly. Keeping these operational processes running smoothly will increase people's confidence in your organization. Make sure to increase your timelines by 25% for any unexpected issues.

8. Final — there's no such thing

During our large-scale census surveys, issues constantly come up from the field — feature requests on the application, optimization requests, etc. These are worth considering, since they help the survey application become more robust and optimized for people's needs. With frequent updates to our APK, our survey team had to follow a rigorous schedule:

- **7 – 8 am:** Travel to the villages
- **8 – 10 am:** Distribute tablets so the survey can start on schedule
- **10 am – 6 pm:** Monitor surveyors and collected data
- **6 – 9 pm:** Collect all tablets from the field so that the daily surveying schedule is not disturbed
- **9 – 10 pm:** Travel to the district headquarters
- **10 pm – 1 am:** Install updated APKs on 350 tablets
- **1 – 7 am:** Rotate charging tablets and sleep

Make sure that your field team is dedicated and hard working. They might have to spend sleepless nights to run the survey smoothly.