The Strategy Unit.

Person-centred intelligence: Designing and selecting a sample

The Strategy Unit and Ipsos MORI

Ipsos

Ipsos MORI

About this guide

- This guide forms part of The Strategy Unit and Ipsos MORI's series about person-centred intelligence.
- This guide focuses on the considerations that should feed into the creation of a sampling plan.
- The sampling plan identifies which people should be invited to give feedback (also known as the sample).
- Selecting a sample that reflects the population of people you are interested in, with as little bias as possible, is crucial to ensuring the results are useful and reliable.

Contents

How to define the target population

Identifying a sampling frame

How to decide on the sample size

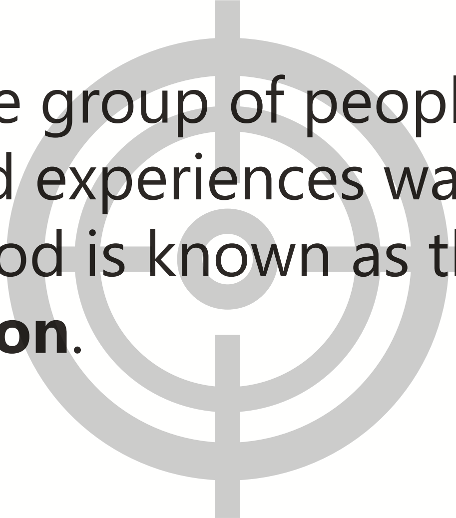
Selecting the sample

Other considerations – data protection & statistical reliability

Note that these steps apply regardless of the approach selected for data collection – whether survey-based or staff-administered.

How to define the target population

What is a target population?



The entire group of people whose views and experiences want to be understood is known as the **target population**.

Defining the target population ensures:

- The sample is selected accurately
- It is clear who the data relates to when making decisions about data collection.

Defining the target population depends on the type of feedback the research is expected to provide and the way the data is intended to be used. These considerations, which are linked to the purpose of your data collection, are discussed in our other guides.

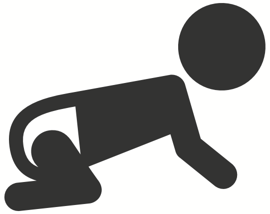
For example, if research is commissioned to understand patient experience of a service, feedback should include patients who have used that service recently enough to clearly remember their experience and have relevant feedback. Therefore, the target population will be patients who have used the service within a specified timeframe (e.g. in the six months prior to data collection).

Alternatively, if the research is about the health status of patients after using the service, feedback should be sought from patients who have been treated within a specific timeframe (e.g. patients who received a hip replacement 12-18 months ago).

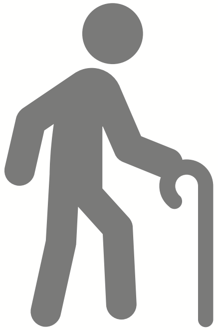
Deciding whether to exclude groups

While it is important that data collection is as inclusive as possible, consider whether data should **not** be collected from certain groups.

For example, it may not be appropriate to collect data from a certain group of patients using a quantitative methodology. Some patients may either be unable to provide feedback quantitatively, or may require a more tailored tool that enables them to give feedback relevant to them.



This includes those:
- Below the age of 16
- With late stage dementia
- Severe mental health difficulties
- At the end-of-life



Where the decision is taken to exclude a group of people (and where it is possible to remove them given the available data), alternative methods of collecting feedback from these patients should be considered to ensure that their views are being captured (for example, qualitative methods such as in-depth interviews conducted face-to-face).

Identifying a sample frame

Creating a sampling frame (Part I)

After defining the target population, a **sampling frame** must be identified.

The sampling frame is the source material, or list, from which the sample to be asked to complete the survey will be selected from.

- First, identify the potential available sampling frames e.g. a database of patients or carers, HR records for staff, or a postcode address file (this has all the relevant postcodes in a given area for a postal survey – which is useful for a survey of the whole population)

- After this, the quality of each sampling frame must be assessed. An ideal sampling frame will have:
  - → Consistent contact information for everyone (whether mailing address, email address or phone number)
  - → The additional information about patients required to identify eligibility (e.g. dates of when each patient used the service, clinical data, number of hours they provide care for, those working in a specific setting)
  - → Up-to-date data
  - → All people within the target population are present in the sampling frame
  - → Each person within the target population is present only once in the sampling frame (or otherwise that there is an easy means of 'de-duplicating' the sample, such as a unique identifier)

Creating a sampling frame (Part II)

It may be necessary to combine multiple sources in order to create a sampling frame that meets these needs.

In reality, it's unlikely that a perfect sampling frame will be available. It may be necessary to take a pragmatic approach and use the best available sampling frame to access the target population. For example, it may be necessary to use a sampling frame in which the contact details are not as up-to-date as desired, but that provides comprehensive coverage of the target group of interest.

In any case, having a thorough understanding of the strengths and weaknesses of the chosen sampling frame is crucial to ensuring appropriate interpretation of the resultant data.

Deciding on a sample frame

Deciding on the sample size (Part I)

Having identified an appropriate sampling frame, the next step is to ascertain the required sample size. The **sample size** is the number of completed responses that your survey receives. It is called a sample, because it only represents part of the target population.

The decision about the required sample size is not a simple one and there is no definitive answer. The decision will be influenced by a number of considerations, some of which are outlined below:

Required precision: It is important to consider the degree to which the findings need to be accurate. If all else is equal, larger sample sizes are more representative of the target population and will therefore provide more accurate insight into that population's experiences and opinions. As such, the chosen sample size needs to be large enough to provide the required degree of confidence in the data. If we know the required degree of confidence, it is possible to work out the sample size required to achieve this. Further detail on statistical reliability is provided at the end of this document.

Deciding on the sample size (Part II)

Required analysis: It is also important to consider the way in which the findings will be analysed. The final sample size must be large enough to allow the required analysis to be undertaken, including any analysis comparing the responses of different groups of people.

For example, to compare the attitudes or experiences of different age groups, the sample size of respondents within each age group must be large enough to provide estimates to a sufficient degree of confidence.

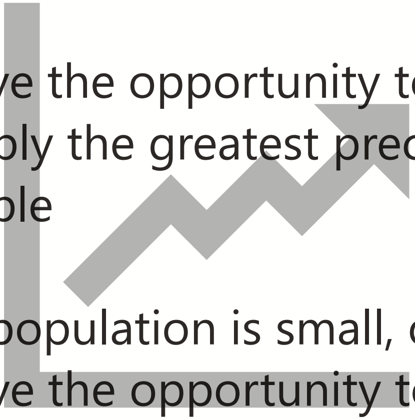
As a rule of thumb, the sample size for each subgroup should be a minimum of 100 to allow analysis to be undertaken (although, as detailed earlier, a larger sample size in each subgroup will provide greater precision).

Available resources: A third consideration is the additional resource requirements that may be associated with larger sample sizes. For example, larger sample sizes may incur higher printing costs or postage costs, or greater amounts of staff time associated with administering questionnaires or analysing free text responses. A trade-off should be made between the required sample size and the availability of such resources.

Census or Sample?

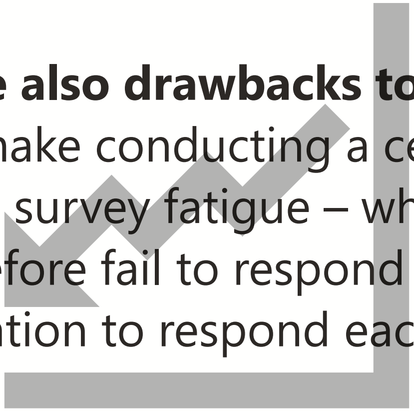
The most straightforward and appealing approach may be to invite all patients within the target population to provide feedback – **effectively to conduct a census.**

There are a number of benefits to taking a census approach:



- Ensures all people within the target population have the opportunity to provide feedback
- Likely to lead to the largest sample size (and possibly the greatest precision)
- Reduces the opportunity for bias to enter the sample

The census approach is appropriate where the target population is small, or where it is prudent to ensure everyone in a target population have the opportunity to feedback, or there is no/limited additional cost of a census.



However, there are also drawbacks to this approach. In some cases, the size of the target population would make conducting a census expensive and/or time consuming. There are also concerns relating to survey fatigue – where people tire of completing a survey on a regular basis and may therefore fail to respond – and so it may be prudent to invite a different subset of the target population to respond each time.

Considering the response rate

As some people will not complete the data collection (called **non-response**), it will be necessary to invite more people to complete the data collection, than the required sample size.

There are a number of potential reasons for non-response:

- Contact details may be incorrect meaning that the data collection invitation never reaches the intended recipient
- Staff collecting the data may not be willing or remember to ask everyone
- The length of the questionnaire or time it takes to administer
- How interested participants are in the topic
- The number of reminders sent to complete the data collection

Tips on how to maximise response rate can be found in our other guides.

Estimating the response rate

Estimating the percentage of people in the sample who data will be collected from will help ensure that an adequate number of people are invited in the data collection to achieve the required sample size.

The higher the expected response rate, the fewer people that you need to invite to take part in your data collection, shown below:

Required number of responses	Expected response rate	Calculation	Number of people to invite
200	20%	200/0.2	1000
200	50%	200/0.5	400
500	20%	500/0.2	2500
500	50%	500/0.5	1000

Selecting a sample

Selecting the sample

This is only relevant where collecting data from a sample rather than undertaking a census of all participants.

In general, the optimal way is to randomly select participants.

Worked examples are shown in the following slides.

Selecting a Sample – Worked Example #1

A commissioner wants to implement outcomes-based commissioning for inpatient services, with one element being based on good patient experience and a requirement for each ward to reach a minimum standard. To implement this, the Trust is required to collect data from patients.

Step 1: Defining the target population

The contract focuses on patients being admitted for planned care. Patient experience will be measured annually, with results needed in March to feed into payment mechanisms. A window of September to November is identified to allow the data to be processed in time for payment. The target population is therefore all those who were admitted for inpatient care from 1 September to 30 November, and excludes those who were admitted as an emergency.

Step 2: Identifying a sampling frame

The database of patients includes patients' contact details, with dates of when they were admitted, and how they were admitted (i.e. emergency vs. planned). The trust decide to collect data using a postal survey. All the patients on the database have a recorded postal address, however, some were admitted more than once in the time period, so the files are de-duplicated.

Selecting a Sample – Worked Example #1

Step 3: Deciding on a sample size

In the sample frame, there are 3,500 patients who were admitted in this three-month period. As there is a minimum requirement for each ward, sufficient data must be collected for all 10 wards. It is therefore agreed to aim for 100 responses per ward, so 1,000 in total (100 response x 10 ward). This also create a good level of confidence at the overall level. They estimate a 40% response rate so need to select 2,500 of the 3,500 eligible patients (1,000 / 0.4).

Step 4: Selecting the sample

The trust downloads a list of eligible patients into Excel. The list is sorted by department. A random number is generated for each patient, which are then copied and pasted as ‘values’. Within ward, the patients are sorted by random number. A selection of 1 in *n* is made for each ward to ensure that 250 are selected per ward and 2,500 overall. A check is made afterwards on proportions for each ward, and by age and gender, to ensure the sample is representative of the target population.

Selecting a Sample – Worked Example #2

An ICS has been looking at data from a range of different sources. Both anecdotally and more formally, it shows that carers of frail older people are not feeling very supported. There is a view that this might impact on patient outcomes. A community health trust decides to pilot a new approach to supporting carers in one of its services.

Step 1: Defining the target population
The new approach is aimed at informal carers; the ICS and trust agree that it is most relevant for those providing significant levels of care. The target population is defined as informal carers who provide 15 or more hours of care for friends or relatives per week.

Step 2: Identifying a sampling frame
There is no sample frame holding names and contact details for the target population of carers.

Selecting a Sample – Worked Example #2

Step 3: Deciding on a sample size

As the new approach is being piloted in one service only, the ICS and trust decide to try and achieve a sample size of 250 carers. This gives a reasonable level of confidence in the findings (+/- 6 percentage points for a question where c.50% give a specific answer).

Step 4: Selecting the sample

The service comes into contact with around 150 patients a week and estimates that one-third have an informal carer providing a significant amount of care (i.e. they have access to 50 carers in the target population per week). They therefore decide to approach all informal carers coming into contact with the service over a specified five-week period, estimating that this will generate around 250 carers.

Selecting a Sample – Worked Example #3

An ICS is embarking on a system-wide transformation that ultimately aims to improve patient outcomes as well as impacting of the day-to-day work of staff. The ICS wants to monitor job satisfaction as the transformation is rolled out and the extent to which it enables them to provide better care for patients.

Step 1: Defining the target population

The transformation will affect staff working in a range of services across a range of organisations. The target population is defined as all permanent members of staff who work in a patient-facing role in a specified set of services.

Step 2: Identifying a sampling frame

A database of staff members can be used to identify which service they work in and whether they are permanent members of staff. From this, staff that work in a patient-facing role can be identified.

Selecting a Sample – Worked Example #3

Step 3: Deciding on a sample size

The intention is to collect the data annually to monitor change, and to collect it via an online survey, which means there is limited additional cost from increasing the sample size. As the transformation is a priority for the ICS and there is a wish to demonstrate the importance of staff in delivering it, a census method is selected. All of the target population are invited to take part, this is a large enough sample size to look at different groups of staff, for example to compare different services, job roles, seniority and so on.

Step 4: Selecting the sample

The sampling frame is compiled along with email addresses, for a paper survey to be distributed to all staff.

Data protection and confidentiality

Data protection and confidentiality

Data collections must be carried out in line with the **General Data Protection Regulations**.

This has a number of different implications, and we recommend liaising with your Information Governance team about this.

More details can be found at: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/information-governance-alliance-iga/general-data-protection-regulation-gdpr-guidance

Statistical reliability

Statistical reliability (Part I)

- When participants in a data collection exercise represent only a sample of the total population of interest, we cannot be certain that the results of a question are exactly the same as if everybody within that population had taken part ("true values").
- However, we can predict the variation between the results of a question and the true value by using the size of the sample on which results are based and the number of times a particular answer is given. The confidence with which we make this prediction is usually chosen to be 95% – that is, the chances are 95 in 100 that the true value will fall within a specified range (the "95% confidence interval").
- The table below gives examples of the confidence intervals for different numbers of responses.
- Strictly speaking, the tolerances applied here apply only to random samples.

Number of responses	**Approximate confidence intervals for percentages at or near these levels (expressed in percentage points)**		
	Level 1: *10% or 90%*	*Level 2:* *30% or 70%*	*Level 3:* *50%*
	+/-	+/-	+/-
100	6	9	10
500	3	4	4
1,000	2	3	3

- For example, taking a sample where 500 responded and where 30% gave a specific answer, there is a 95% likelihood that the true value (which would have been obtained if the whole population had been interviewed) will fall within the range of +/-4 percentage points from that question's result (i.e. between 26% and 34%).

Statistical reliability (Part II)

- Different groups within a sample (e.g. different staff groups, or patients on different wards) may have different results for the same question. A difference has to be of a certain size in order to be statistically significant.
- To test if a difference in results between two sub-groups within a sample is statistically significant, at the 95% confidence level, the differences between the two results must be greater than the values provided in the table below.
- Similarly, if tracking changes between two different years of a survey the difference needs to be of a certain size to be statistically significant.
- Again, strictly speaking the sampling tolerances shown here apply only to random samples

Number of responses	**Approximate confidence intervals for percentages at or near these levels (expressed in percentage points)**		
	Level 1: *10% or 90%*	*Level 2:* *30% or 70%*	*Level 3:* *50%*
	+/-	+/-	+/-
100	8	13	14
500	4	6	6
1,000	3	4	4

- This means, for example, that if surveying 100 patients in 2019 and 100 patients GPs in 2020, the difference in results between the two surveys for a question where c.50% give a particular answer must be 14 percentage points to be statistically significant.

The Strategy Unit.

Ready to move on?

Ipsos MORI

When designing your sample:

- ☑ Identify the entire group of people whose views or experiences you wish to understand and make informed decisions about who should be excluded
- ☑ Choose the most representative list of people within your target population to select a sample from
- ☑ Think about how accurate your data needs to be, what analysis you want to do and the responses you need, along with the resources available, so you can select a sample size
- ☑ Factor this into the response rate to work out how many people you need to invite
- ☑ Select the right number of participants from the sampling frame
- ☑ Comply with GDPR