3. Defining and identifying uncertainty

Analytical uncertainty can feed through into analysis and subsequent decision making from many different sources

We encounter uncertainty throughout the decision making process and in the analysis which supports it. In addition to uncertainties around the analytical question, we will also find uncertainty in the context of the decision being made, the data and assumptions feeding into the analysis and in the analysis itself. As analysts we need to understand and describe contextual uncertainties to ensure our analysis has impact; and we need to describe and quantify analytical uncertainties to ensure decision makers are clear about how far analytical results can be used to support their conclusions.

Early identification is important

Try to identify and record all the potential sources of uncertainty in your analysis at an early stage. Early identification of uncertainty is important; if you overlook a potential source of uncertainty this could reduce the usefulness and impact of your subsequent analysis. See the **presenting and communicating uncertainty** section for Tornado diagrams that are a useful way to communicate the size of uncertainty.

This section sets out a range of techniques to help you understand and assess the sources of uncertainty in your analysis.

3.1. Defining uncertainty

There are a number of ways to classify uncertainty. A common classification divides uncertainty into known knowns, known unknowns, and unknown unknowns, as we explain in Table 3.1. **Other classifications** consider, for example, the range of things about the analysis which may be uncertain and whether uncertainty relates directly to these "objects" of uncertainty or to the quality of evidence behind them. We recommend following one of these frameworks when assessing the uncertainties that affect your analysis and the decisions it will inform.

Table 3.1: Classifications of Uncertainty

Classification	Aleatory uncertainty	Known unknowns - Epistemic uncertainty	Unknown unknowns - Ontological uncertainty
Definition	Sometimes referred to as "known knowns", aleatory uncertainty is the **things we know that we know**. This refers to the inherent uncertainty that is always present due to underlying probabilistic variability.	Known unknowns are **things that we know we don't know**. This type of uncertainty comes from a lack of knowledge about the (complex) system we are trying to model. Assumptions are used to plug these gaps in the absence of information.	Unknown unknowns are **things that we don't know we don't know** . It usually comes from factors or situations that we have not previously experienced and therefore cannot consider because we simply don't know where to look in the first instance.
Can it be quantified?	**Yes** it can be quantified. We usually characterise it using a probability distribution function (PDF). A PDF gives all the possible values that a variable can have and assigns a probability of occurrence to each. As analysts, the challenge for us is to derive the PDF. If you find that you can't then you may instead have a known unknown.	**Yes** it can be quantified (but isn't always) – e.g. through sensitivity analysis. These techniques try to quantify the uncertainty by altering assumptions and observing the impact on modelling outputs. They will work if the range of assumptions tested covers the range of unknown variables.	**No** it cannot be quantified. We cannot identify unknowable unknowns, so there are no actions we can take to quantify them. What we can do is be clear about the sources of uncertainty we have included, so that any others subsequently identified would likely add to that uncertainty.
Can it be reduced?	This type of uncertainty cannot be completely removed. We can sometimes reduce it through data smoothing or increasing the size of a sample, but there will always be some random variability.	Known unknowns are reducible by gathering information to lessen the gaps in our knowledge. Using new data sources, expanding our data collection or conducting research can remove the need for assumptions or refine their ranges.	This type of uncertainty is not reducible. However, this type of uncertainty can usually be separated into "unknowable unknowns" and "knowable unknowns". Horizon scanning can help identify knowable unknowns. Once they are identified they become known unknowns.
Example	Tossing a coin is an example of aleatory uncertainty. We can observe the possible outcomes (heads or tails) and the probability of each occurring (50:50), therefore create the PDF. However, prior to the coin being tossed we cannot reduce the uncertainty in outcome.	Taking our coin toss example, we don't know whether the coin is fair in the first instance. We may assume the coin is fair and will give a 50% probability of each outcome. Once we start to toss the coin, we start to gather information on its fairness. The longer we toss the coin the better our information gets and the greater the reduction in the known unknown.	Unknown unknowns are often future events or circumstances that we cannot predict, for example, somebody swaps the coin to a weighted one without our knowing, or steals the coin altogether! Previous analysis is no longer reliable as it didn't account for this change.

3.2. Sources of analytical uncertainty - Data

The data that feeds into your analysis project will have been previously specified, defined, and collected. In some cases, you will do this yourself, but you may also draw on data sources collected by others. Having chosen your data sources for your project you will need to think about how well your data describes the reality of the situation you are modelling or analysing.

To gain a full picture of the impact of data uncertainty on your analysis you should think through what you know about where your data has come from. You should use a data log with quality and impact Red Amber Green (RAG) ratings. Consider the following questions:

How your data source compares with your analysis objective

How well do the definitions and concepts in the data chosen fit with what you are trying to measure? Differences between the data and your target group can mean that a dataset captured for one purpose is inappropriate for another. For example, you might want to analyse London & South East but only have data for the whole of the UK.

Where the data come from and how they have been collected

How rigorous was the data collection process? Was the data owner's quality assurance sufficiently robust? For survey data, would respondents have fully understood the question intent? Some datasets are subject to regulation and compliance with standards or other codes of practice. In such cases, quality should be well documented and assured like in National Statistics.

When considering uncertainty in input data, you should think about whether the data being used was gathered for an alternative purpose and if it has been manipulated and how you can adjust or account for this. Accompanying data descriptions (or a quick exploration of the source data if these don't exist) can be helpful in understanding data limitations of the data and whether any adjustments made could conflict with or bias your analysis.

What period the data covers

More uncertainty will occur if either the data don't match the time period of interest and/or if the data are volatile.

Whether your data has been subjected to any pre-processing

For data obtained in a processed state from others you may need to explore what processing steps were taken to determine how that may affect the data you are using. For example, missing values may have been imputed, survey data may have been weighted to make survey results representative of a wider population, extreme values and outliers may have been removed, data sets may have been combined (possibly resulting in false positive or false negative matches), disclosure controls may have been applied (potentially biasing the data set). Consider how the retention or exclusion of an outlier will affect your results. Truncation or removal of outliers will typically introduce bias but this may be tolerated in exchange for reduced variance.

Check whether there is any bias or uncertainty in the data

Statistical sources often come with supporting information about accuracy and reliability. You can sometimes find information on variance (or standard errors, confidence intervals, coefficients of variation) and you may find indications of likely bias, from special studies comparing or linking sources. These direct measures of quality, together with indirect measures such as response and coverage rates can tell you a lot about the uncertainty. In the absence of direct measure of variance, be aware that small sample sizes will increase the margin of error in your results.

3.3. Sources of analytical uncertainty - Assumptions

Considering the assumptions you're making in your analysis is critical to any uncertainty analysis

Consider where you have used assumptions

Assumptions are used when we have incomplete knowledge. All models will require some assumptions, so you need to ensure that assumptions are robust and consistently understood. You should use an assumptions log with quality and impact RAG ratings and they should be signed off by stakeholders. Where did the assumptions come from? How were they generated and why? What is the impact if they are wrong, and how often are they reviewed?

What assumptions are outside the scope of the model?

There are often parameters outside of the scope of the model that have been implicitly assumed. For example, models may assume no substantial policy changes in related areas and there may be deliberate limits in the coverage or timelines of your analysis – deliberate modelling exclusions that allow timely and effective analysis. These assumptions and limitations provide the context in which the modelling results are appropriate. You need to be aware of the restrictions that these assumptions impose on the interpretation of analytical results and take care to explain where modelling results can (and cannot) be used.

Assess the quality of each assumption

Assumptions should be based on robust evidence. The less evidence to support an assumption the more uncertain it will be. High quality assumptions will be underpinned by robust data, while low quality assumptions may simply be an opinion or may be supported by a poor data source.

Assess the impact of each assumption

The importance of an assumption is measured by its effect on the on the analytical output. The higher the impact of an assumption the more uncertain results will be. Critical assumptions will drastically affect the results, while less importance assumptions may only have a marginal effect on results. More weight should be given to gathering evidence to improve the quality of critical assumptions.

What don't you know?

Some uncertainties can't be captured in an assumption as we don't have perfect insight. However, effort should be made to identify all possible uncertainties and capture these as assumptions. The assumptions log will convey the boundary of what has been included.

3.4. Sources of analytical uncertainty - Analysis

Undertake appropriate AQA

An additional, but important source of analytical uncertainty is in the analysis itself. With verification and validation of models, good Analytical Quality Assurance (AQA) practices can help identify the restricted uses of analytical outputs and help minimise the possibility of errors. However, mistakes can still be made, so being clear with decision makers about the extent to which analysis has been quality assured can help them understand how far they may rely on analytical results in support of their decision making. Please see the **AQuA Book**, **AQuA Book resources** and **BEIS QA tools and guidance** for more information.

Carrying out adequate quality assurance is an important way to ensure sources of uncertainty have been sufficiently mitigated. Ideally, the AQA should be carried out throughout the project - before, during and after the analysis to inform at all stages. The AQA process should involve checking the analyst has done the following, for example:

- Considered why the methodology is appropriate to solve the analytical problem or answer the research question
- Considered why the analysis is appropriate for the type of data collected
- Understood the data that will be used in the analysis, including main sources of potential error and limitations of the data/analysis
- Analysed and interpreted the data in a consistent way
- Made it clear how analytical constructs e.g. categories, classifications, typologies etc. have been developed.
- Ensured all interpretations are well supported by the data, accurately reflecting the meanings assigned by the participants

Note that these steps should also be applied for quality assurance of both quantitative and qualitative research. More information on AQA of qualitative research is included in **mitigating uncertainty in qualitative research**.

3.5. Sources of uncertainty in experimental and quasi-experimental evaluation designs

Experimental and quasi-experimental evaluation designs are used in government to understand and estimate the impacts of policies. They do so through statistical comparison to a group or time period unaffected by the intervention. This unaffected group acts as a proxy for what would have happened to the affected group in the absence of the policy and is commonly called the counterfactual. It is also possible to compare multiple versions of an intervention with a control group, these are known as multi-arm trials.

Commonly used experimental and quasi-experimental methods include randomised control trials, difference in difference and interrupted time series analysis. The **Magenta Book** provides a full guide to evaluation methods. When conducting or commissioning this type of research, analysts have an important role in ensuring that potential sources of uncertainty are understood, adequately addressed and effectively communicated to stakeholders.

The sources of uncertainty in experimental and quasi-experimental research can be broadly categorised into data, study design and statistical analysis. Analysts who employ these methods should consider to what degree each of these three areas create uncertainty in understanding and estimating the impact of a given policy or intervention.

Data

Regarding quantitative data for evaluations, some of the things you need to consider are the sample size, representativeness, choice of indicators and whether there is any missing data.

You need to ensure your sample size is large enough to answer your research question(s)

If the sample size is too small for an experimental or quasi-experimental design, it will not be possible to achieve sufficient statistical power and provide a robust answer to the research question(s). It is also important to consider whether you need to provide estimates for population subgroups. The more you break down the sample into groups, the greater the overall sample size needs to be in order to draw statistically significant conclusions about the subgroups of interest. A statistical power analysis helps to estimate the minimum sample size required for a study, given a desired significance level, effect size and statistical power.

Representativeness: Is your data broadly representative of the target population that is being studied?

The generalisability of the research findings may be limited if certain subgroups of the population are not adequately represented in the sample. This is particularly true when differences between population subgroups are related to the outcomes that are being studied.

Some concepts cannot be measured directly and therefore must be proxied by other observable or measurable phenomena, a process called operationalisation

For example, health status is a concept that cannot be observed directly, and therefore researchers need to operationalise the concept by other indicators (e.g. body mass index or smoking status). Uncertainty is introduced when a concept of interest is not operationalised appropriately, and the proxy indicators do not adequately capture that concept or distinguish it from others. You should ask yourself how well the indicators measure the concept that is being studied. Are the assumptions underlying the operationalisation valid?

Missing data or missing values are a common source of uncertainty and can have a significant effect on what can be inferred from the data

Uncertainty is introduced when data is missing systematically, or in other words, not completely at random. There are two types of systematically missing data: missing not at random (MNAR) and missing at random (MAR).

Missing data is particularly problematic when the reason why data is missing is related to a concept or intervention that is being studied. Data which are MNAR are missing due to reasons related to unobserved outcomes. Data are MAR when the reason the data are missing is related to the observed outcomes (i.e. variables for which we have complete information).

For example, in a study on depression, data would be MNAR if men did not respond to a survey because of their level of depression. This is because the concept of interest (depression in this case) is unobserved.

On the other hand, data on depression would be MAR if men were generally less likely to respond to a survey, irrespective of their level of depression. In this example, sex is a variable that is directly observed.

Reasons for missing data include: social desirability bias in surveys, attrition in longitudinal studies, data entry errors, poor quality data collection instruments, or under-sampling of groups that are difficult to reach. You should ask: is some of the data in your dataset missing? If so, why is it missing?

Study Design

The design of a study and assumptions about how potential effects will be identified can be a significant source of uncertainty.

Uncertainty is increased when the chosen control group is not comparable to the group that is affected by the policy or intervention

In the absence of a valid counterfactual, the estimated impact of a policy or intervention may be significantly biased and the true impact will be uncertain. For example, when the research does not include an independent control group and relies solely on a before-after analysis. The problem of relying on a before-after analysis is that it is not possible to determine whether the observed effects of an intervention would have occurred in the absence of the intervention.

Binary before-after comparisons can be misleading if they are presented without contextual information

Comparisons from one period to another (such as comparing this week or month to the same period last year) are common. However, they present data out of the context of the underlying trend or do not account for the effect of seasonality. For example, drawing comparisons in outcomes between January 2020 and January 2021 would be misleading as external factors, such as the COVID-19 pandemic, would have had significantly impacted trends. Alternatives to before-after comparisons are discussed in chapter 4, **Mitigating uncertainty**.

Where randomisation is compromised or not possible, there is greater uncertainty in attributing observed changes to the intervention being studied

Randomised controlled trials (RCTs) are the gold standard for testing hypotheses. They involve randomly assigning participants to the intervention or control group. This creates a counterfactual to which you can compare the intervention group. When randomisation is compromised or not possible, there will be greater uncertainty in attributing changes in outcomes to a given intervention. A few examples of randomisation being compromised are:

- Endogeneity (i.e. when the allocation of an intervention is influenced by the outcome indicator)
- Non-compliance (i.e. when participants who should receive the intervention do not receive the intervention)
- Breach of protocol
- Treatment contamination or spill over (i.e. when participants in the control group are exposed to the intervention). This can be an issue with studies involving cluster randomisation based on geographic areas, such as local authorities.

However, in many contexts randomisation is not feasible or ethical. In these situations, it is possible to employ a quasi-experimental method to create a counterfactual (for more details see the **Magenta Book**). Analysts must ask themselves if the design of the counterfactual creates potential for uncertainty in the research findings.

Analysis

Statistical analysis can present an important source of analytical uncertainty in evaluation design, which may lead researchers to unwittingly make biased or invalid inferences. Analysts should ask themselves if the chosen statistical models accurately describe the relationships between the variables of interest and take account of potential sources of bias. Are effects conditional on other variables, or are they expected to vary across groups? Have potential alternative explanations or theories been explored in the analysis? Have all relevant variables been included in the analysis?

Interaction effects and heterogeneous treatment effects can introduce uncertainty into the results of analyses if not they are not addressed

Common issues in experimental and quasi-experimental studies include interaction effects. This is when the causal relationships between two variables depend on the state of a moderating variable, or when effects of an intervention vary across different groups (i.e. heterogeneous treatment effects). For example, where the relationship between the dose of a drug and the efficacy of the treatment varies across genders. When policy interventions have heterogeneous effects, focusing only on aggregate effects and failing to account for differences across groups may lead to invalid inferences. Similarly, uncertainty will result from failing to account for relevant moderating variables in the analysis. Another source of uncertainty results when the statistical model falsely attributes the effect of a missing variable to those variables that are included in the model. This is known as omitted variable bias.

The way in which research findings are interpreted can also be a source of uncertainty

An example is the ecological fallacy, in which findings from analysis of aggregate data are erroneously attributed to an individual. For example, research of aggregate data shows that countries where there is a high average fat consumption also have a high breast cancer death rate. If you were to infer from this finding that a woman who has a high fat diet is more likely to die from breast cancer, this would be falling foul of ecological fallacy. The error here lies in the fact that statistical inference is intended to generalise from a sample to a population, and not from a population to an individual. Committing an ecological fallacy can lead researchers to make invalid inferences, thereby creating uncertainty around the true impact of a given policy or intervention.

3.6. Sources of uncertainty in survey research

Survey research can involve qualitative and/or quantitative data, collected through a range of data collection methods such as online surveys and telephone or face to face interviews. Both the type of data collected, and the mode of data collection should be determined by the aims of the research, noting that these factors can influence the degree to which uncertainty is introduced.

The sources of uncertainty in survey research can be broadly categorised into the survey design, sampling strategy, data collection method and analysis.

Survey Design

A poorly designed questionnaire can greatly increase the level of uncertainty in survey data

Good questionnaire design is vital to ensuring the validity and reliability of survey responses. A valid questionnaire is one that measures what it intends to measure. That is, the objectives of the questionnaire and the items within it are clearly understood by the respondent and elicit the information required by the researcher. Reliability refers to the consistency of a survey measurement and the extent to which the measurement is able to elicit the same information from the same person each time it's administered, assuming all else remains unchanged. There are a number of tests and methods for ensuring questionnaire validity and reliability (see section 4 on **mitigating uncertainty in survey research** for more information).

The choice of response scales can introduce uncertainty into survey research, particularly when response scales are not chosen optimally

The most common survey response scales are: dichotomous (e.g. agree vs. disagree) and rating scales (e.g. five-point Likert scale: strongly agree, agree, undecided, disagree, strongly disagree). There is a tendency for responses with rating scales to regress to the middle of the scale in surveys, a phenomenon called error of central tendency. This can be related to the length of the survey, or survey fatigue, and the tendency of respondents to avoid extreme responses. Survey responses can also be unreliable if labels on a five-point Likert scale have no clear meaning. For example, the labels 'somewhat satisfied' or 'extremely satisfied' can be confusing and risk being interpreted differently by respondents, thereby introducing uncertainty into the survey results. A third example of how bad survey design can lead to uncertainty is called non-differentiation in ratings, or survey straightlining. Straightlining occurs when respondents lose their motivation to engage with the survey and consequently rush through it by giving their answer to a series of questions in the same place on a rating scale. When designing surveys, researchers should take great care in choosing survey response scales and consider how their choices may be a source for uncertainty.

The order of questions or response options can also impact how a respondent interprets and responds to survey questions

Respondents' answers to a question can be influenced by previous questions posed and by the answers they gave to those previous questions; a notion called priming. If questions are always presented in the same order, this impact may be difficult to detect. Similarly, the order in which response options are displayed can affect which one is chosen.

Some concepts cannot be measured directly and therefore must be proxied by other observable or measurable phenomena, a process called operationalisation

This is a key stage in the survey design because the survey questions must capture the concepts of interest. For example, the concept of health is not easily measured directly and therefore researchers may want to operationalise the concept by other indicators (e.g. body mass index or smoking status). Uncertainty is introduced when a concept of interest is not well operationalised and the proxy indicators do not adequately capture that concept or distinguish it from other concepts. You should ask yourself how well the variables describe the different dimensions of the concept that is being studied. Are the assumptions underlying the operationalisation valid?

Sampling Strategy

With the exception of a census (which surveys every member of a given population), survey research typically relies on data taken from a sample of a population under study.

Survey data are subject to sampling error, which occurs when the sample being used is not representative of the population. A representative sample is one that accurately represents the population on specific characteristics, in that the sample and population have similar distributions on the variables of interest, e.g., gender, age, socioeconomic status, or education. There are many dimensions on which you might evaluate representativeness - it all depends on the required level of detail, the scope of your study and what information about your population is available. All samples contain some degree of error, and therefore uncertainty, but the more representative a sample is of its population, the less error and uncertainty it will contain.

The following sections summarise some common causes of sampling error that ought to be considered by analysts and researchers when considering the extent to which uncertainty exists within survey data.

Random sampling will minimise uncertainty but is very difficult to achieve in practice

In random sampling every member of the target population has an equal chance of being selected and thus should eliminate sampling bias. Other probability-based techniques such as stratified sampling (where participants are selected in the proportion that their subcategory occurs in the population) or systematic sampling (where every nth person is chosen), are likely to result in a degree of sampling error but the uncertainty caused can be estimated and measured.

However, where a non-probability sampling technique is used, the likelihood of sampling error and response biases occurring is much higher and it's not possible to estimate the extent to which a sample is unrepresentative. Such techniques include volunteer sampling, where individuals have chosen to be part of the study, and opportunity sampling, where participants are simply chosen from those available at the time. With this technique, the degree to which responses are likely to accurately reflect those of the population cannot be calculated, and therefore findings should not really be extrapolated to the wider population - despite this often happening in practice.

If your sample is too small, it may not allow you to draw reliable inferences

There are a number of things to consider with sample size:

- Is the sample large enough to be representative of the population under study? If not, any conclusions you draw should not be generalised to your population under study. Generally speaking, the bigger a sample, the more likely it is to be representative. Note, however, that this is not always the case: sample size is only a useful indicator of sample quality when an appropriate sampling technique has been employed.
- How precise do you need your results to be, or what is the margin of error you are willing to accept?
  - How certain do you need to be that your results are not due to chance? This is your significance level.
  - How certain do you need to be that your results will detect an effect when there is an effect to be detected? This is statistical power.

You can conduct a power analysis to estimate the minimum sample size required for a study, given a desired significance level, effect size and statistical power. If you do not manage to reach this minimum sample size, you increase the likelihood that your results are erroneous. For example, if the sample size and consequently statistical power is low, the probability of concluding there is no effect when, in fact, there is one, goes up. This is increasingly likely if you are looking to detect a small effect, as small samples offer weaker test sensitivity than large samples.

If your sampling frame does not represent the target population, uncertainty is introduced

A sampling frame is a record of the target population containing all participants of interest from which we can extract a sample. Sampling frames can include government registers, postcode lists, records of demographic information provided by those who have signed up to an online survey website. The vast majority of sampling frames will have some defects due to inaccurate information being provided or records not being up to date. The smaller your sample and the greater the number of dimensions on which you want it to represent your target population, the greater the impact of inaccuracies in the sampling frame.

Data Collection Method

Survey mode can introduce selection bias when certain members of a population are more likely to be included in the sample than others

Surveys can be administered using a variety of modes, including face-to-face interviews, telephone interviews and self-completion web-surveys, and these often vary in terms of the demographic they tend to reach. For example, older age groups are generally more difficult to reach through online surveys. The topic of the survey may also influence what mode is more or less appropriate: measuring internet access within the general population using an online survey will produce biased results, as all survey respondents would have internet access, otherwise they wouldn't have been able to participate in the survey.

Some survey modes have little means to ensure inaccurate data is not provided or the survey is not completed with errors

With self-administered surveys, we have to trust that the data being provided is accurate - and if accurate demographic data is key to your analysis then this source of uncertainty should play a larger influence in your choice of survey mode.

There are different degrees of uncertainty intrinsically associated with different types of self-administered surveys – for instance, self-completion paper questionnaires generally show a higher number of unanswered questions than online surveys.

Interviewer bias can introduce uncertainty in data collection

Interviewer bias is when characteristics or behaviours of the interviewer influence how participants respond to questions. It relates to aspects of the interviewers and the way in which they ask questions and respond to answers—it is distinct from bias arising from the content or wording of questions. Such bias may stem from perceptions of the interviewer's identity. The interviewer's sex, ethnicity, age, attractiveness, social class, level of education, perceived life experience, or professional background may affect how participants respond to questions, especially if these characteristics seem to relate to the interview topic.

Linked to this is the interviewer's ability to establish rapport with the interviewee: participants may not feel comfortable to disclose accurate information, especially on personal or sensitive topics, as a result of who is interviewing them.

Interviewer bias may also arise from the actions and behaviours of the interviewer, for example:

- Using certain language, phrases or leading questions
- Using a tone of voice or inflections to imply a presumed answer
- Using non-neutral body language that establishes a mood or projects onto the conversation

These include:

- Businesses or individuals being unreachable
- Businesses or individuals refusing to respond
- Respondents giving inaccurate answers
- Processing or analysis errors

For example, inaccurate answers to a question about money spent on fuel would lead to a difference between the estimate and the population value even if the entire population were surveyed. These errors are usually very difficult to quantify and to do so would require additional and specific research.

Analysis

Missing data or missing values are a common source of uncertainty and can have a significant effect on what can be inferred from the data

There are techniques to deal with missing data - the two primary methods being imputation or removal of data. However, to decide the appropriate technique you must understand the reason why data is missing. Incorrect use of techniques to address missing data can occur through misunderstanding of the reason why data is missing. The different potential reasons are explained in the 'Data' section of **'Sources of uncertainty in experimental and quasi-experimental evaluation designs'**.

When data is missing systematically, or in other words, not completely at random, simply removing observations with missing data is likely to result in bias as the missing information is unknown. For an example of this, see the 'Data' section of **'Sources of uncertainty in experimental and quasi-experimental evaluation designs'**.

If your significance level is too high, you increase the likelihood of concluding that a significant difference exists when in fact it has occurred by chance

We can use statistical significance to decide whether we think a difference between two survey-based estimates reflects a true change in the population rather than being attributable to random variation in our sample selection.

A type I error (also known as a false positive) occurs when you conclude that a significant difference exists when in fact it has occurred by chance. The probability of making a type I error is represented by your chosen significance level. A 5% standard is often used when testing for statistical significance, which means that you accept a 1 in 20 chance of the observed change being calculated by chance if there is actually no underlying change.

If your sample size is too small, you increase the likelihood of concluding there is not a significant effect, when actually there really is

A type II error (also known as a false negative) occurs when you conclude there is not a significant effect, when actually there really is. It is related to the power of a statistical test: the probability that a test will find a statistically significant difference between two samples. A type II error is more likely to occur if your sample size is too small for a significant difference to be detected at your chosen significance level.

There are minimum sample sizes that you need to reach in order to conduct robust statistical comparisons between sub-groups. Even if your overall sample size is large, if some groups of interest are small, it is not appropriate to conduct analyses using disaggregated data at this level. The recommended minimum sample sizes can be determined with a statistical power test, which takes into account your desired effect size and confidence level - the greater the effect size and the higher the confidence level, the greater the sample you'll need.

3.7. Sources of uncertainty in qualitative research

Qualitative research projects are intended to explore and explain a sample of views, perspectives, behaviour, understanding and experiences of particular individuals or groups. The aim of qualitative research is to provide an in-depth understanding of a phenomenon rather than to establish its prevalence, probability or causality.

Qualitative research provides rich and deep insights into a specific phenomenon or experiences of a particular group within society, which wouldn't otherwise be possible with quantitative research methods. Well-designed qualitative research will provide robust, insightful data to understand the big picture and go beyond anecdotal evidence.

There is inherent uncertainty and bias in all research and analysis methods, and qualitative research is no different. Qualitative researchers are trained to be mindful of the pitfalls of conducting and analysing qualitative data. They can account for, mitigate and minimise sources of uncertainty when designing and undertaking qualitative research to ensure it is robust, reliable and findings are presented and used appropriately. The following sections highlight some of the main sources of uncertainty at different stages of a qualitative research project.

Sampling in qualitative research aims to get good coverage of the population of interest, rather than full representation

Due to practical constraints it is not possible to include representatives of all the different sub-groups within the population in qualitative research. Instead priority is placed on depth rather than breadth of coverage - participants are included in the research on the basis that their views or experiences are worth exploring in-depth in their own right, and not because they are expected to be representative of a wider group or population. However, there is inherent uncertainty as the sample participants' experiences might be radically different from others.

The research team is selective and uses their judgement when deciding who to include and exclude from the research sample. Quotas are set to ensure a sufficient number of research participants meet the key criteria or characteristics across the sample (e.g. balance of gender, age, location). As a result, individuals or groups with low prevalence in the population or who may be harder to reach could be excluded or under-represented in the sample. Where understanding the perspective of these groups is a priority for the research they may be purposely over-represented in the sample.

Purposive (selective) sampling in this way helps to ensure that opposing perspectives are taken into account in the study. Nonetheless, some respondents' perspectives may not always be included in qualitative studies while some sub-groups' views may be over-represented. This needs to be considered when drawing conclusions from the research especially when making generalisations about wider attitudes, experiences and behaviour.

Recruitment Methods can introduce uncertainty into qualitative research

Identifying participants to engage in qualitative research can be time-consuming and costly, especially when looking to include hard to reach groups or ability to participate is limited by practical constraints such as the location and proposed timings for fieldwork. Steps are therefore often taken to minimise both and make the process as efficient as possible.

Researchers often rely on specialist recruitment agencies or pre-populated lists of potential respondents to identify suitable participants for the research study. The research team will typically produce a recruitment specification, specifying quotas and inclusion criteria for participants to take part. These typically aim for diversity in terms of demographic information, location, extent of experience/ engagement with a process etc. The aim is to achieve a range of perspectives within the time and budgetary constraints. However, there may be biases in terms of location or diversity, as recruitment agencies aim to meet the recruitment specification at least cost. This might mean, for example, that there may be demographic or locational biases in the sample, for example due to travel constraints or researchers' working hours.

Furthermore, for convenience the recruitment agency might seek to recruit the participants from a pre-existing contact list in the first instance. Participants on those contact lists might differ in significant ways from those who are not – especially if the participants have been involved in previous research recently. These respondents might, for example, be fatigued from the research process, or otherwise simply be less interested or more engaged in the research process than the wider population. For this reason, researchers sometimes build conditions into the recruitment process to exclude those who have recently participated in any qualitative research.

If recruitment agencies are unsuccessful in recruiting from their existing contacts, they may turn to social media or other online sources to recruit participants. Whilst this may enable them to attract new people who haven't participated in research before, the use of online recruitment methods means that they may again mean that certain groups are excluded from participation and there is an element of bias in the sample.

The discursive nature of qualitative research means data is not collected in a uniform manner, and a variety of factors - including the researcher - can influence the process

The conversational and interpretive nature of qualitative research can introduce uncertainty because there can be inconsistency in the way in which data is collected, shared and understood across the project. This includes data collected via discussion (e.g. interviews and focus groups) as well as that collected via observation and self-reporting techniques (e.g. diaries, journals, videos).

Qualitative research methods, such as interviews and focus groups, typically depend on interactive discussions between the researcher and the participant(s). This enables the researcher to probe and clarify the participants' responses and pursue interesting lines of enquiry. This means that even when the researchers are using the same discussion guide, participants might reveal different insights to different researchers – depending on a variety of factors, such as the rapport they have established with the researcher and/or other participants, and the degree of probing from the researcher, and the environment in which the interview is taking place. So, if multiple researchers are involved in the data collection process or different methods are used in the same study (for example a mix of one-to-one interviews and focus groups) there may be differences in the data collected across the fieldwork.

Moreover, the time it takes to discuss a topic can vary, and this can mean some topics are not covered to the same extent in the time allocated across different interviews. This can result in inconsistencies in the format, coverage and content explored across the set of interviews, focus groups or supporting methods (e.g. journals, blogs) even when there is only one researcher involved in all. This tactic can be deployed on purpose to ensure the full range of topics are covered in sufficient depth across the fieldwork when it is not possible to discuss all in detail with each respondent or group of respondents. As such it does not undermine the quality or reliability of the research but must be accounted for when analysing and reporting the data.

By its nature, some qualitative research relies on what respondents report in interviews, focus groups and any written and audio-visual material they share as part of the process. Thus, an additional uncertainty arises in terms of reliability. Some respondents will be less willing to share information compared to others, while some may have difficulty recalling their views or experiences to answer the research team's questions especially if a sensitive subject is being discussed. Similarly, the impact of social desirability bias - i.e. the tendency to answer questions in a manner that will be viewed favourably by others - can be strong in qualitative research settings where there is direct engagement between the researcher and respondent, and between respondents. Whilst these issues are not unique to qualitative research, they tend to be more prevalent when there is direct engagement between the researcher and the respondent.

To address this, researchers might choose to supplement or replace interviews or focus groups with observational methods - such as ethnography, usability research, accompanied activities or video recording. These techniques can help a researcher to achieve a more objective view and can be used to pick up insights that the respondent does not consider salient or interesting. However, there can be an element of bias and subjectivity in these methods too as they are dependent on what the researcher notices or hones in on and how they interpret the data which, in turn, can influence data collection and also analysis.

Ideally, qualitative researchers will stop collecting data when they have reached saturation. Saturation occurs when no new insights emerge that are unaccounted for by theory or by data. It is typically detected when the research team finds repetition of insights across respondents. However, saturation is not always achieved in practice, due to time and budget restraints, or a lack of respondent diversity. Instead it is common for the number and type of research participants to be specified at the outset of the study as part of the research design process. Failure to reach saturation means it is not possible to state that the findings are conclusive and no new themes, insights or perspectives are likely to emerge. As such there will be inherent uncertainty about conclusions inferred from the research which must be accounted for.

The approach to data analysis can influence what and how data is used

Qualitative research produces rich, detailed and often large volumes of data. Sometimes trade-offs need to be made when deciding how to analyse and use the data, that is, between presenting the full range of evidence on the one hand, and focussing on key themes and commonalities across the data. These decisions will also be influenced by the approach taken to the analysis. For example, thematic analysis will seek to draw out overarching themes in the analysis, whilst narrative analysis will focus more on how people make sense of their experience rather than the experience itself, and discourse analysis will hone in on the language used by research participants. The choice of analytic approach can therefore determine what and how data is used in analysis and reporting.

Sometimes researchers choose to synthesise or summarise the experiences of research participants with similar characteristics, behaviours or opinions to create typical or illustrative reference cases. They may also prioritise common or more prevalent themes or insights when reporting findings. In doing so, some of the nuance observed between respondents may be lost when findings are generalised.

Qualitative research is subject also to uncertainty due to differences in interpretation – the same information or quotation might be interpreted differently, for example, between the researcher and the respondent – especially if the respondents have substantial demographic differences from the researchers. Likewise, the research team may differ in how they interpret different participants' responses. This can mean the same data is analysed differently by different researchers – they may identify different salient points, or group participants differently to answer the same research question.