Uncertainty Toolkit for Analysts in Government

A cross government group has worked together to create an assessing and communicating uncertainty toolkit.

This toolkit sets out good, not best, practice, as analysis and communication must always be tailored to the audience and decision being made.

The toolkit is presented in the first instance as a suggested set of guidelines and we will be consulting with experts from Government, the academic community and other external bodies to develop thinking in line with the latest evidence on communicating uncertainty.

Feedback

We welcome feedback on the content You can get in touch by emailing: **AnalystsUncertaintyToolkit@homeoffice.gov.uk**

Is this toolkit for you?

This toolkit is aimed at all analysts, whether you are new to the Government or an experienced analyst looking to develop the way you communicate uncertainty to stakeholders.

This document sits alongside the Aqua Book guidance on quality analysis. The Aqua Book should be read first as it describes the principles of uncertainty analysis and the processes that should be used in considering it. This is a supplementary document providing additional information for identifying, estimating and communicating uncertainty in analysis to support decisions and decision-makers.

We have also provided a **one-page summary**.

Useful links:

Click here to see the accessibility statement

Introduction

Uncertainty analysis: understanding what you don't know

Uncertainty is unavoidable when making predictions about future events or impacts of decisions. Our knowledge of the present is typically incomplete; and we can't be sure what will happen in the future. We can reduce uncertainty due to lack of knowledge by obtaining more information, or asking experts for advice. However, we can't eliminate uncertainty entirely, so we must find ways to describe and communicate it.

There is a difference between uncertainty and risk; risk typically refers to the likelihood of a future unplanned event or unintended consequence that can be assigned a numeric probability. Forecasts of risk, like other forecasts, are themselves subject to uncertainty. For example, UK Met Office predictions of the risk of rain are uncertain as to whether it will rain at all, as well as uncertainty about how much it will rain. This uncertainty is described using a range from running several forecasts from equally likely initial conditions.

Why does it matter?

Taking account of uncertainty – and being seen to do so – is important for public trust. We must not pretend that the consequences of a policy or decision are certain: they are always uncertain to some degree. For each option, a range of outcomes are possible. Implying certainty about one particular outcome can damage public trust when things turn out differently.

Example

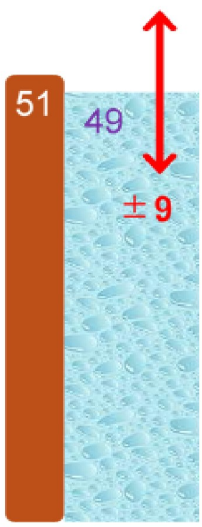
Red River Flood, Grand Forks USA, 1997

The National Weather Service (NWS) predicted, 2 months in advance, the Red River to crest 49 feet.

In response, the levees were built to handle a flood of 51 feet

The actual flood level was 54 feet.

Had the NWS communicated their uncertainty (+/- 9 feet) the several $billion damages as well as the huge personal impact could have been avoided.



This toolkit contains the following chapters:

2. Jointly agreeing how uncertainty should be used

Before conducting any uncertainty analysis, it is important to check that you understand how the analysis will be used. What type of decision is being made? Are you informing a debate or a final decision? This will frame how you approach the analysis and how you communicate it to your customers.

3. Defining and identifying uncertainty

Considering the whole system that influences your analysis helps identify all possible areas where uncertainty can arise. Ensure you understand what is causing the uncertainty in your inputs and outputs, and whether this can be quantified.

4. Mitigating uncertainty

Once uncertainty has been identified it should be reduced where possible. We outline techniques for mitigating uncertainty in a variety of analytical contexts.

5. Understanding and measuring uncertainty

Before conducting uncertainty analysis, you should consider the range of possible techniques. We provide a list of the most common techniques with some examples for reference.

6. Presenting and communicating uncertainty

There is little point conducting any analysis if it has no impact, so this is perhaps the most important section. It is important to engage with your decision makers so they take account of uncertainty in any decisions they make. You should think about how you interact with the audience, tailoring your communication to different groups. This will help them to understand the consequences of uncertainty and why they should be interested in it.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

2. Jointly agreeing how uncertainty should be used

Decision-makers need information about uncertainty in the impacts of their decisions. This section looks at the steps to ensure that analysts and decision-makers agree on the question being asked, how analytical outputs will be used in the decision-making process, and how to appropriately incorporate uncertainty into the analysis.

2.1. Understanding the problem

Clarify what the real question is

It is important to ensure that the question is correctly framed to address the problem. For example, are we really interested in 'how much money is this new policy likely to save?', or should we be asking 'how certain can we that this policy would save more than £x?'.

As well as clearly defining the question, we should also ensure that any important detailed questions to support the analysis are appropriately thought through. For example, are we interested in uncertainty on a calendar year or financial year basis?

Identify the important details

You may also want to jointly identify sources of uncertainty that are so deep that little or nothing can be said quantitatively about their impact on the outcome. An example might be the impact of a natural disaster on a particular policy. Deep uncertainties can have major implications for decision making, and may call for a descriptive approach to the analysis and presentation of options.

Once the key questions have been agreed, analysts and decision-makers should discuss how the outputs will be used. Understanding the role of the analysis in the decision-making process will help to inform which uncertainty techniques are appropriate and proportionate.

2.2. How will outputs be used?

Discuss the role for the analysis

Will the decision be based purely on the results of the analysis, or is it just one of a wider set of considerations informing the decision? The more influential the analysis is, the more important it is to build in robust uncertainty analysis.

Are there dependent models drawing on the analysis?

If the output is to be fed into 'downstream' models, then it is important to understand the requirements of those models. For example, if *scenarios* are used to illustrate uncertainty in your model, then these may not be suitable inputs for a Monte Carlo simulation in a dependent model.

Now that we know what the question is and the context in which the analysis will be used, we should agree how to build appropriate uncertainty into the analysis. It is a good idea to ask the decision maker what it is conceptually that they *would like* to be able to say. This is important as it frames how the uncertainty analysis will be conducted.

2.3. Work with the decision maker to inform their decision at the right time

Explain how the uncertainty can be used to better inform decisions

We can help the decision-maker to understand how information about the uncertainty present in the analysis supports a better informed and more robust decision. It's important to make it clear that not all sources of uncertainty will be quantifiable and/or supported by robust evidence, and we should jointly prioritise the key sources of uncertainty that need to be included in the model.

Avoid misleading results or spurious accuracy by choosing the appropriate presentation

We can assess the impact of uncertainty using many different techniques. Not all will be appropriate for a given piece of analysis. A poor choice of technique may even give misleading results. For example, if there are many sources of uncertainty, the use of 95% confidence intervals to describe just one limited part of this uncertainty may be misleading, and modelling a range of described scenarios may be more appropriate. Appropriate presentation, such as rounding, is equally important to avoid spurious accuracy and impling better precision than is actually the case. Communication is covered further in the **presenting and communicating uncertainty section**.

Discuss how the uncertainty will inform the decision-maker's judgement

Discuss with the decision-maker what level of uncertainty is acceptable. Do they want to know how wrong the forecast would need to be in order to change or rethink the policy? Or are they simply interested in an output "range"? If so, what does that "range" actually mean?

Examples can be very helpful when discussing the analysis with the decision maker. An answer to the question of the net benefits of a policy may be £3m, with uncertainty analysis giving a broad range of £0.5-5.5m. You could discuss with the decision maker how they want to be able to frame the analysis, for example:
"A range of £0.5-5.5m";
"The estimated benefits are £3m, with analysis showing a 90% likelihood that benefits will be between £1-5m";
"Analysis shows that there is an 80% likelihood that the net benefits will be greater than £2m";
"The policy needs to have x amount of takeup in order to break even"

Operational decision makers may not want to see a range of results, but instead want to plan to a certain level of confidence, such as 65% or 95% rather than 50%. For example, when planning the number of schools, prison places or GPs we'll need over the next 5 years, it may be more appropriate to plan to a higher level of confidence than 50%.

Financial decision makers may be interested in understanding the likelihood of receiving a certain level of income, or that risks and opportunities will materialise. For example, HMRC might want to know the likelihood of receiving a certain level from tax receipts. Here, the analysis would need to go hand in hand with financial risk management to mitigate the risks materialising or crystallise the opportunities.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

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.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

4. Mitigating uncertainty

4.1. Mitigating uncertainty in quantitative data analysis

To decide what method is appropriate to use to deal with missing data, you need to understand why the data is missing.

If the data is missing across all observations, for example due to test design, failure in the observations or failure in recording observations, the data can be classified as missing completely at random (MCAR). This is because the reasons for its absence are external and not related to the value of the observation. It is typically safe to remove MCAR data because the results will be unbiased. The statistical test(s) you are performing may not be as powerful, but the results will be reliable.

Listwise deletion involves deleting all data for an observation that has one or more missing values. The analysis is run only on observations that have a complete set of data. If the number of observations with missing data is small, it may be the most efficient method to handle missing data. However, if listwise deletion would result in throwing away a lot of your data, pairwise deletion may be more appropriate. Here, cases with missing data are used when analysing variables where they don't have missing values. For example, if you are missing data on a participant's gender then this participant would be excluded from any analysis using the gender variable but would be included in any analyses for which the participant has complete data. However, the resulting statistics may vary because they are based on different data sets.

If the data is missing systematically, imputation can be more appropriate. The imputation method develops reasonable guesses for missing data. A common technique used when the number of missing cases is small is to impute the mean or median. For datasets with a large amount of missing data, multiple imputation is often used. Here, instead of substituting a single value for each missing data point, multiple imputed data sets are created where the missing values are exchanged for values that encompass the natural variability and uncertainty of the observed values. Each set is then analysed using the standard analytical procedures, and the multiple analysis results are combined to produce an overall result.

The website **How to Deal with Missing Data** provides a useful overview of methods to deal with missing data.

If you need to be very confident that your results are not going to be a false positive (concluding there is a significant difference when it has actually occurred by chance), you can decrease your significance level

The probability of making a type I error (or a false positive) is represented by your significance level (or alpha level): a 0.05 significance level indicates that you are willing to accept a 5% probability that your results occurred by chance.

If you need to be very confident that your results are not going to be a false positive, you can decrease your significance level, for example to 0.01. Because this change increases the amount of evidence required to conclude a difference is significant, it makes your test less sensitive to detecting differences, but decreases the chance of committing a Type I error (or a false positive) occurring from 5% to 1%.

However, using a lower value for alpha means that you will be less likely to detect a true difference if one really exists, which increases the risk of a type II error occurring.

You can decrease your risk of committing a type II error by ensuring your test has enough statistical power

A type II error is when you conclude there is no significant difference when one does in fact exist. You can reduce the risk of a type II error by ensuring your sample size is large enough to detect a difference when one truly exists.

Combining findings from a range of studies, known as synthesis, allows you to draw conclusions from a body of evidence

Synthesis methods include systematic review and meta-analyses. These methods include reviewing study characteristics and quality and, where relevant, combine data in a statistical synthesis of study findings. Evidence synthesis (i.e. systematic reviews and meta-analyses) sit at the top of the 'evidence pyramid'. This means they are considered the evidence source with the least uncertainty and highest rigour as their design minimises bias and maximises your ability to ascertain causality. However, the types of evidence at the top of the pyramid may not be available or feasible for your research topic of interest, in which case you will need to move down the evidence pyramid (note that this is not a comprehensive list of potential evidence sources):

4.2. Mitigating uncertainty in experimental and quasi-experimental evaluation methods

The counterfactual and intervention group must be comparable

This depends on the collection of comparable data from both the intervention and control groups to ensure you can measure the intervention effect. In addition, the intervention effect must be large enough to be distinguished from noise in the data. Details of how to design a counterfactual for quasi-experimental and experimental studies can be found in the **Magenta Book**.

Where it is not possible to run an experimental study to measure impact, you should select a quasi-experimental method based on the availability of a comparison group

Details of which methods are appropriate for different situations are explained in section 3.5 of the **Magenta Book**.

Where no comparison group is available, as in the case of analysing the impact of a global event like the COVID-19 pandemic or 2008 economic crisis, you can use an interrupted time-series approach. However, this is only appropriate where changes are sudden. This approach option is preferred to binary before-after comparisons which can be misleading as they present data out of the context of underlying trends.

Ensuring you have a solid grounding in the theory underpinning your intervention is essential for an effective evaluation design.

Developing a comprehensive theory of change to explain how inputs of the intervention lead to outputs, outcomes and impact and testing it against existing evidence will help identify what data is needed for an impact evaluation. Having a robust, high quality theory of change will also ensure that you do not miss any key indicators and alert you to any potential unintended consequences that may need to be measured.

There is a useful **logic models guide on gov.uk**. This goes through the process of creating a logic model and provides examples and templates. There is also some good guidance on developing and using logic models from the **Nuffield Trust** and a **step by step guide from the US CDC**.

A baseline, or starting point, can also serve as a benchmark against which future progress and effects of an intervention can be assessed.

Comparing estimated effect sizes to effect sizes from other studies in different contexts can be useful for setting realistic expectations of effects and thereby guide the research design (e.g. required sample size, whether effects are homogeneous or heterogeneous). Benchmarking can also help to understand similarities and differences between the effects of interventions in different contexts.

Where data is limited or the intervention will produce small or unpredictable effects, theory-based evaluation could be an alternative.

Theory-based impact evaluations make conclusions about the effect of an intervention through testing the causal pathways through which the change is brought about. This method assesses whether the evidence is sufficient to support these causal pathways and that alternative explanations can be ruled out. The Magenta Book provides further guidance on theory-based evaluation.

4.3. Mitigating uncertainty in survey research

There are numerous different ways to reduce the risk of uncertainty posed by your data collection methods. These include:

Questionnaire testing

Piloting: testing a questionnaire with a small group of experts is a useful method for ensuring questions are appropriately worded and likely to be understood by the target audience. It is a useful test of face validity. That is, it tests whether the question, on the surface, taps into the concepts it's intended to measure.

Cognitive testing: this is a form of qualitative research with prospective survey respondents to understand how they interpret and answer specific questions and what they think about when completing the survey. This helps to ensure results of the survey are properly understood by the researcher and that the questions make sense to the respondent and elicit the kind of information expected.

Question testing: Item non-response occurs when a respondent completes a survey but fails to provide an answer on specific items. The amount of item non-response is considered a useful indicator of data quality. If particular questions are commonly left unanswered, it may indicate a problem (e.g. respondents may not understand what's being asked, the question may be sensitive, the response categories are insufficient and so the respondent cannot adequately answer, etc.). Item non-response is a particular concern if it occurs systematically (i.e. particular types of respondents are less likely to answer a particular question). When conducting a survey, testing for item non-response helps to identify issues with particular questions and any systematic bias that might introduce.

Question sequencing and randomisation

To minimise the priming effects arising from the order of questions, researchers should carefully decide the sequence of questions, as well as considering whether rotating the order of certain questions is appropriate. Furthermore, it's common practice to randomise the order of response options where these are codes or statements, and invert the order of the scale for half the respondents for questions with a rating scale as the response options.

Using scripting and logic in online surveys

To increase data accuracy, for online surveys you can use methods such as only showing particular questions to respondents as a result of their responses to previous questions and ensuring that several answer options cannot be chosen on questions where only one answer option should be chosen.

The size of your sample

Larger samples will tend to be more representative (assuming you are conducting random sampling). Keep in mind that it's unlikely that every sample will be perfectly similar to the population of interest. There will always be a little sampling error associated with any study, unless you sample every single member of your population.

The likely response rates of your survey

You may need to contact a lot more people than you need to achieve a sample that is representative of your population under study and allows you to draw conclusions within the margin of error you are willing to accept.

Your sampling frame (your possible participants) and recruitment procedures

Avoid only recruiting members of a certain subset of your population. Ideally you would randomly sample from your sample frame. Through this, you minimize any selection biases that might occur, such as volunteer bias. Selection bias occurs when the subjects studied are not representative of the target population about which conclusions are to be drawn. However, selection bias is only problematic if the response probability is correlated with the variable you are measuring, i.e. the people who respond to the survey answer differently to your questions to the way in which those who did not respond would have done. Thus, addressing selection bias requires breaking the connection between non-response probability and outcome variables: adaptive response design.

Implementing a stratification protocol, such as proportionate stratified sampling

Let's say you do your research and find out your population under study are 80% women. You could then make sure that 80% of your sample consists of women, such as by quota sampling.

Alternatively, you might want to consider over-sampling: this is the selection of a large number of additional respondents that match certain criteria, to allow researchers to measure more precisely the changes in key populations.While this might not result in your overall sample representing the overall population under study (e.g. the general population), this may provide the most useful approach to allow you to draw conclusions from a small sub-group, such as a particular ethnic group.

Using sample weights to correct for the over-representation or under-representation of key groups

You can weight down the responses from the over-represented group (which may have been purposefully oversampled) to make sure their views do not have a disproportionately large effect when conclusions are drawn based on responses from the whole sample, or weight up the responses from under-represented group to make sure they don't have a disproportionately small effect.

A sample weight is a statistical measurement which is linked to the record of every survey respondent. The sample weight is calculated based on the probability of being selected for the survey for the respondent and can also account for other imbalances which arise in the sampling process, such as non-response adjustment. The value of a weight can be interpreted as a measure of the number of population members represented by that respondent.

For example, if 51% of a population are female, but a sample is only 40% female, then weighting is used to correct for this imbalance. There are a number of different types of weights:

Design weights

We use design weights to account for the different probabilities of being sampled that different respondent types have. Let's say we're collecting data based on a list of addresses. People who live in a place where many families share the same address will have a lower chance of being surveyed than people who live at single-family addresses. Weighting our survey results ensures our results won't be skewed by this discrepancy.

Non-response weights

You can use non-response weights to correct for the fact that some types of people are less likely to be willing to participate in your survey than others. To illustrate, let's imagine that young people in our district are less inclined to answer our survey questions. Weighting our results ensures that we account for this fact by placing more load on the responses from young people who do participate.

Calibration/post-stratification weights

You can use calibration weights to make the characteristics of your sample closely match the characteristics of your population. This is commonly done using demographic data (like gender, age, income level) that is publicly available (from a Census, for example) as the target and adjusting the sample demographics to match that target.

Weights almost always increase the standard errors of estimates and introduce instability into your data. Some researchers like to "trim" the weights to not allow extremely high weights that can increase instability of estimates, but trimming the weights can often result in reducing the representativeness of the weighted data – it's a trade-off between less instability or more accurate representativeness.

4.4. Mitigating uncertainty in qualitative research

There are three main ways to mitigate uncertainty when conducting and using qualitative research: standardisation; validation; and transparency.

Standardisation of tools and processes can help to minimise uncertainty in qualitative research

To minimise the uncertainty that researchers can introduce to the research, steps should be taken to facilitate consistency in approach throughout the research process - from participant recruitment, through to fieldwork (interviews and focus groups), analysis and reporting. The exceptions to this is where the researcher is intended to be part of the research process, and their perspective is a key output (e.g. where they play the role of participant observers or immerse themselves in a culture), or where there are ethical considerations that need to be considered to ensure participant welfare.

Research tools

It is common practice to produce and use standardised research and analysis tools, and to provide briefings to team members about how they should be applied. For example, it is usual to produce a script and questionnaire for recruiters to use to screen participants. This ensures that all potential respondents are asked the same questions in the same format and in the same order to determine their eligibility to participate in the research.

It is also usual for all researchers undertaking fieldwork to follow a standard topic or discussion guide. These can vary in style and level of detail depending on the nature of the study and researcher preferences, but provide instructions for how interviews should be completed including the order in which topics should be discussed, the approximate length of time to spend of each topic and key questions or probes that should be asked to all participants. The level of flexibility that researchers have in using the guide should be agreed up front, and the aims and objectives of the research emphasised so all involved are aligned and clear about how the project should be delivered.

Where possible, researchers other than the lead researcher should aim to observe at least one interview or focus group before undertaking fieldwork so they are familiar with how the topic guide is used in practice. Likewise the lead researcher could observe the first session facilitated by another member of the team to ensure consistency in approaches. A short debrief should be provided after the first interview to discuss any issues or proposed changes required to the discussion guide before further fieldwork takes place. Additional checkpoints during fieldwork should be built into the project plan to review progress and consider if any further tweaks to the discussion guide should be made. All of these measures will help to improve consistency across fieldwork and minimise researcher bias.

Outputs

The main sources of data produced via qualitative research are written outputs. These can range from brief notes taken by the researcher during or immediately after fieldwork - including references to key themes discussed, short quotes, and observations and reflections of the researcher - to full verbatim transcripts which include every pause, 'umm' and 'ah'. Where possible audio or video recordings of interviews and focus groups should be made to facilitate recall. These should be transcribed and used as a supplement to the researcher's notes. Where this is done, a consistent format should be used. To further reduce uncertainty where audio or video recordings have been made, a second transcriber or observer can be tasked with quality assuring transcripts and notes to ensure all relevant information is captured and recorded accurately. Doing so reduces the risk of bias in interpretation or selection of data and provides all researchers in the team with access to data produced via each interview or focus group.

Analysis

Where a team of researchers are involved in conducting the fieldwork and analysing and reporting the findings it is important to establish a common approach to analysis up front. It is common practice to use analysis software such as NVivo or MAXQDA, or even a simple spreadsheet. Regardless of the preferred analytical tool being used, the process can be expedited and facilitated by setting out expectations of how data should be organised and managed before commencing analysis, establishing a common set of labels or codes that can be applied to the data during analysis, and agreeing points to review and refine the approach.

A range of techniques can be used to validate qualitative, contextualise and test out qualitative findings

A range of techniques can be employed to sense-check interpretation, appropriateness and prevalence of qualitative findings. These include: 1) analytic induction; 2) data triangulation; 3) respondent validation; and 4) analytical quality assurance.

Analytic induction

One of the most common theoretical methods used in qualitative data analysis in Grounded Theory. This is based on the idea that hypotheses about our data should be generated by or 'grounded in' the data itself rather than developed prior to data collection and analysis. The process of generating and testing these hypotheses is analytic induction. This involves systematic interrogation of the data. Hypotheses are generated during analysis and tested on a small number of cases in the sample and then refined or reformulated until all cases fit and all data has been used. Whilst this process is time and labour intensive, and can be complex, it can help to reduce uncertainty by focusing on those things that are consistent or universal within the data and ensuring that no evidence is overlooked or set aside.

Data triangulation

One way to sense-check or validate qualitative research findings and potentially reduce uncertainty is to consider the extent to which they align with findings from other sources of evidence, a process known as triangulation. This could include findings from other similar qualitative studies, or a review of the wider evidence base.

It is also common to adopt a multi-method approach to data collection within one study to be able to look at the same phenomenon from different perspectives. For example, a study could employ multiple qualitative methods - interviews and observation. By doing so the researchers may seek to sense-check what they observe with participants' motives or perceptions of their own behaviour. Another approach commonly used is mixing qualitative and quantitative methods in the same study, using a survey of a wider sample of respondents to test the validity of qualitative findings.

The key assumption behind data triangulation is that if you look at the same phenomenon in different ways and generate similar findings or conclusions then there is greater reliability and less uncertainty. Whilst there is value in this approach and it can provide greater confidence in research findings, it should be treated as a way of providing greater breadth, richness and rigour to the research rather than a means to provide a complete or 'true' picture as qualitative research by nature involves some level of interpretation.

Respondent validation

One way to minimise the risk that the researcher has misunderstood or misrepresented the research participant's perspective is respondent validation. This is particularly useful when undertaking ethnographic research or generating case studies. The researcher will seek confirmation that their findings and perceptions match with the views of those who participated in the research by sharing their work with them and inviting feedback. This technique is particularly important in instances where research participants will be identifiable in the findings (e.g. where respondents have agreed to be used in a case study or organisations have agreed to be named in a report) to ensure they are content with how they are being represented.

However, it is worth being mindful that one of the key purposes and strengths of the research process is to observe social phenomena - to look upon it from an external perspective. The researcher or research team plays a key role in this. By asking respondents to check and correct the researcher's interpretation of the social world, it privileges the respondent's view of their actions which does not necessarily increase the validity of findings. There are, of course, times where research is designed to be respondent-led - whereby the respondent acts as the researcher and may even interview and observe their peers as well as provide insight into their own lives. Where this is not the case, the decision to seek respondent validation of research findings should be considered carefully and alongside the role of the researcher in the research process.

Analytical quality assurance

As outlined in **mitigating uncertainty in quantitative data analysis**, steps applied to quality assuring quantitative data are also relevant to quality assuring qualitative research.

Although the AQUA book does not specifically outline suggestions for AQA of qualitative research, it should be subject to the same attention to best practice and quality assurance in its design and analysis as model building and quantitative analysis. Check whether your department has created its own guidance for conducting qualitative AQA (for example, this exists in the Home Office).

To facilitate rigorous AQA, a complete audit trail should be kept recording decision points made and outputs produced at all key stages in the research process. This includes details about how the research problem or questions were defined at the outset; the approach to sampling and recruitment selection; research tools; copies of any notes, transcripts, audio-visual materials or other materials generated through the research (e.g. respondent diaries, journals or scrapbooks); and decisions regarding data analysis (e.g. theoretical approach, coding framework). AQA can take place during the research process as well as its conclusion. AQA reviewers act like auditors, checking that appropriate procedures have been followed and considering the extent to which theoretical inferences are credible and defensible. Consequently, AQA of qualitative research can be quite a demanding and resource intensive exercise. Whilst it can help to reduce uncertainty, analysts should keep proportionality in mind when deciding how and when it can best support the research process.

Being transparent about how the research and analysis was conducted builds confidence in the findings

As noted above, maintaining a clear audit trail can help to mitigate uncertainty by facilitating peer review and scrutiny of the research process and its findings. When reporting findings from qualitative research it is important to be transparent about both the limitations and bias in the research design and the implications for how they should be interpreted and used. See **communicating uncertainty in qualitative research** for further details. When doing so it is important to reiterate the role and purpose of qualitative research and that findings are not intended to be representative of or generalisable to the wider population.

5. Understanding and measuring uncertainty

At this stage you should have considered where the uncertainties lie. You will now need to consider whether it is possible and appropriate to quantify this uncertainty. In a highly uncertain context, detailed assessment of data-driven uncertainty may not be worthwhile. This is an analytical judgement to be made in light of the full consideration of sources of uncertainty.

This section presents some approaches at a high level then in more detail for quantifying and measuring the uncertainty in your analysis.

5.1. Quantifying uncertainty in inputs

We have highlighted ways to think about the uncertainty coming from specific sources. We now bring this together into approaches that can be applied to understand the size and distribution of these uncertainties and how you might include this information in your analysis. In most cases, the approach to uncertainty quantification is limited by the data and time available to you, though you should also consider where the major sources of uncertainty are in your analysis and pay most attention to these.

In addition, the following approaches can be useful in understanding the size and distribution of the uncertainties in your analytical inputs.

Consider discrete and continuous distributions…

A probability distribution describes the probability of occurrences of different outcomes. Generally, there are two types of probability distribution; discrete distributions and continuous distributions.

data sources may include confidence intervals…

Consider whether you have information about the underlying distribution of the parameter. Often data from other sources will be provided with confidence intervals (or standard errors, etc) that can be used to quantify uncertainty. Where such information is not provided, you may be able to approximate these with knowledge of the sample size and design.

or analysis of past model performance to create distributions…

Distributions can also be created using what you know about error from previous models. Consider the performance of previous forecasts against outturn results. The distribution of previous errors can provide the uncertainty distribution for the current forecast. Additionally, you might be able to use one source to validate another or to help estimate the uncertainty in the original source.

Without quantitative data, use expert judgement

If no quantitative data on the underlying population is available, you may be able to elicit this information from experts . For example, using the Delphi Method, a structured facilitation method, which relies on a panel of (ideally external) experts to build a combined judgement, for example, about a particular number. This method can be used to ask a panel of experts to estimate the range of uncertainty and use the aggregated responses to produce a distribution. Consider tools to overcome biases, e.g. Brier score. These measure the mean squared difference between the predicted and actual outcomes. The lower the Brier score is for a set of predictions, the better the predictions are calibrated, which has been shown to improve judgement

Ranges can work in absence of distributions…

A range is similar to a probability distribution, in that it considers the possible outcomes but does not consider the probability of each outcome occurring. If there are data or resource limitations a range can be a simple way to illustrate the uncertainty in a parameter.

Historical data can inform the range,

Historical data can be used to quantify a range. Consider how the parameter has changed over a suitable time period. The maximum and minimum values could provide a sensible range. When using historical data be aware that you will only be able to assess 'business as usual' uncertainty. If there are future shocks to the system this may fall outside your historic range.

You can also use a literature review…

For parameters that have been the subject of academic studies a literature review can be used to create a range. Consider why different studies may result in different outcomes, and which studies are the most suitable for what you are trying to measure.

Expert steers can inform the range

If no quantitative data is available, consider whether there are relevant policy/operational constraints that will limit your range. Judgement from experts can be also be used to create sensible ranges.

In some situations, it is not possible to create a probability distribution or a range. In such cases, make a qualitative assessment of uncertainty. This is still useful to analysts and customers to consider the magnitude of uncertainty.

RAG rate your parameters

You can make qualitative assessments yourself, and by using expert judgement. A simple approach is to RAG rate the likelihood and impact of uncertainty in your parameters. This qualitative assessment should be considered when thinking about the analytical results. If data is categorised as highly uncertain and having a large impact on results, then final outputs will be subject to large uncertainty. Care in presentation is needed when using these more qualitative methods, however, as it's potentially easier for decision makers to misunderstand the relevance of qualitative assessments of uncertainty for their decision compared to quantitative assessments. It is particularly important to explain the impact of uncertainty on the analytical result and the decision to be made when using these types of qualitative assessment.

Break-even analysis can help re-frame the question

A further option to support decision makers is to use break-even analysis. This is useful to understand at which point a saving becomes a cost or possibly at which point you would take a different decision.

Some decision makers will be used to seeing a range around a central estimate. Break-even analysis works backwards – if we were to break-even what would the input be? This could help bring the policy alive and help assumption owners to really consider how realistic the assumption is.

Break-even analysis helps people understand how much the input has to change before you reach a break-even point so they can consider the probability of this occurring.

5.2. Common techniques for assessing uncertainty in analytical outputs

Now that we have explored how to assess uncertainties in individual analytical inputs, we turn to commonly used techniques for analysing overall uncertainty in analytical outputs, moving from approaches to use when we have a good numerical understanding of input uncertainties, though to approaches which can be used when we know less about how input uncertainties are quantified.

Monte Carlo Techniques

If all significant sources of uncertainty can be quantified, along with the correlations between them, then probabilistic methods can provide a picture of the range of possible outcomes and the likelihood of each.

Outline:

The basic process for a Monte Carlo simulation is to:
1. Define a distribution for each input showing the uncertainty in each. These can be simple distributions based on estimation (e.g. uniform, triangular) or more complex distributions based on data (e.g. normal, beta).
2. Define the correlations between these inputs
3. Randomly generate a value from each input distribution (accounting for correlations)
4. Calculate the outputs of the model deterministically
5. Repeat steps 3) and 4) many (i.e. thousands of) times
6. Analyse the distribution of the resulting outputs

This should be repeated until the key outputs are stable and reproducible at the level of rounding that will be used when communicating the results

Advantages:

- Produces a full probability profile of the range of possible outcomes and the likelihood of each – the gold standard in uncertainty analysis
- Enables analysis of complex interactions between uncertainties within a model
- Produces a visual representation of the range of possible outcomes, that may aid customer understanding
- Can be used to assess the impact of removing or reducing a source of uncertainty

Disadvantages:

- Highly dependent on the accuracy of the distributions used
- May require more resource than other techniques to build
- Correlations can be difficult to define mathematically, and can give misleading results if not properly accounted for
- Can be computationally expensive
- Outputs may not be reproducible if insufficient iterations are used (effectively introducing further uncertainty to the model)

Example:

- Can help assess overall uncertainty when you have uncertainty around many aspects of your model
- Assess uncertainty around a fund forecast
- Estimate the uncertainty around assumptions used in policy costings
- An example of how MoJ use Monte Carlo analysis to assess uncertainty is provided **here**

Convolution

If there are a small number of uncertainty sources, then it may be possible to combine them mathematically, which can be quicker than using probabilistic methods

Outline:

In principle uncertainty distributions can be combined to give an exact distribution for the resulting overall uncertainty (known as convolution).

Consider rolling two dice - the value of each follows a discrete uniform distribution from 1 to 6. If we wanted to know the sum of the two dice, we can combine these uniform distributions to show that the resulting probability profile follows a triangular distribution from 2 to 12.

This avoids the randomness of probabilistic approaches. However this method tends to get very complicated once more than a few uncertainties are involved, and probabilistic methods may be easier.

Advantages:

- Deterministic, so outputs are reproducible and quick to generate
- Works with any input probability distributions (don't need to assume normality)

Disadvantages:

- Can be difficult to do with more than 2 or 3 sources of uncertainty
- Generally only feasible if the distributions are combined using simple mathematical operations (addition, subtraction, multiplication)
- Assumes that each uncertainty is independent (or correlations must be accounted for)

Example:

Any situation where you're combining two probability distributions using simple mathematical operations (addition, subtraction, multiplication)

Summing Uncorrelated Uncertainties in Quadrature

If multiple components need to be summed, and the uncertainty in each can be assumed to be normal and uncorrelated, then summing in quadrature is a useful shortcut for estimating the total uncertainty

Outline:

If the model is simply summing a number of output components (e.g. a number of different funding forecasts), and the uncertainty is known in each, then the uncertainty in the sum of those values can be given by the square root of the sum of the individual uncertainties squared (assuming that all errors are independent and normally distributed).

Advantages:

- Very simple to calculate
- Can include uncertainties from all sources

Disadvantages:

- Only applicable for models where uncertainties can simply be summed - more complex interactions will need to use other techniques
- Only valid if the individual uncertainty distributions are independent and normally distributed

Example:

Useful for financial forecasting, where individual budgets and their uncertainty has been assessed, and these must be summed to give the overall budget. In this situation assuming independence between budget over/underspends is often reasonable.

Also useful for quickly estimating the uncertainty when changing the timeframe of a forecast e.g. if you have a forecast by month (with quantified uncertainty), then this is quick way of estimating the uncertainty in the annual totals – assuming there is no correlation between the uncertainty from one month to the next.

Using Past Variance to Estimate Future Uncertainty

If a forecast has been produced repeatedly over many years, it may be possible to use the accuracy of all the previous forecasts to estimate the uncertainty in the new forecast

Outline:

Start by looking at the variance of all of the previous forecasts after one time period, and create a distribution of these variances. This distribution can be used to estimate the uncertainty of the new forecast after one time period.

Repeat for the variance at each subsequent time period to produce e.g. a fan forecast (smoothing if needed).

Advantages:

- Captures all sources of uncertainty
- Avoids the need to identify, quantify and combine individual sources of uncertainty
- Based on real-world performance, so avoids optimism/pessimism bias (the estimated uncertainty may be asymmetric as a result)

Disadvantages:

- Needs the system to be stable over time (i.e. there's no reason to think that the system is more/less predictable than in past years).
- Needs the approach to forecasting to have been consistent over the years (i.e. there's no reason to think that our forecasts will be more/less accurate than in previous years).
- Would not work for large and nfrequent events.
- The forecasts must have been produced for a sufficiently long time so that we have enough historical data on the variance to capture the full range of plausible variances

Example:

MoJ prison population forecasting

Focussing on the Dominant Uncertainty

If one source of uncertainty has a much greater impact than all the others, then the uncertainty due to this one factor might be a reasonable proxy for the overall uncertainty

Outline:

Consider all the sources of uncertainty affecting the analysis and quantify the impact that each has on the overall uncertainty (even if this is approximate or subjective – see section on 'Defining and Identifying Uncertainty'). If one source has substantially more impact that the others, then simply looking at the uncertainty in this one factor might be a pragmatic approximation for the overall uncertainty measure, ignoring other sources.

This is not a robust way of estimating overall uncertainty, as it ignores all but one source. It should only be used if previous techniques are inappropriate, and there is genuinely one source of uncertainty that has an impact that is an order of magnitude larger than the others. However, when time is tight this may be a pragmatic way of estimating uncertainty.

Advantages:

- Relatively simple and quick
- No need to combine uncertainties
- Focusses attention on the key driver of uncertainty, avoiding unnecessary discussion of relatively minor elements

Disadvantages:

- Will always underestimate the overall uncertainty as most sources of uncertainty are excluded
- Can dramatically underestimate uncertainty if the source of uncertainty assumed to be dominant isn't as dominant as believed

Example:

The Accuracy Tracking Tool **(link to DfE Accuracy tracking tool to come)** can be used to estimate the residual uncertainty once the dominant uncertainty has been modelled. This tool assesses the accuracy of different forecast elements and allows you to see the percentage which both the dominant and residual uncertainty contribute to the total error.

Scenario analysis

If there are many sources of uncertainty with complex correlations between them and the likelihood of them occurring is unknown, then it may be more sensible to use a set of scenarios to illustrate the range of plausible outcomes.

Outline:
- Identify the individual sources of uncertainty, and the range of possible values for each.
- Create a set of coherent scenarios (jointly agreed with the decision maker), setting each source of uncertainty to a value that could realistically occur in parallel with the others (e.g. a crimewave is unlikely to occur if the economy is booming).
- Try to ensure that the chosen scenarios cover the full range of possible outcomes, from highly optimistic to highly pessimistic, with equal attention given to each to avoid bias in presenting the results.
- Then quantify the overall uncertainty in each scenario, combining uncertainties using other methods in this section (the benefit is that this combining can be done manually for a small number of specific cases, rather than having to formulate the combinations mathematically for all possible inputs).

Advantages:

- Don't need to mathematically define the correlations between sources of uncertainty.
- Gives 'real-world' explanations to the range of possible outcomes, which may increase buy-in from the customers.
- Can include low-probability, high-impact events (i.e. system shocks) without needing to define their probability.

Disadvantages:

- The choice of scenarios is subjective, and may not cover the full range of plausible outcomes.
- Risk of optimism/pessimism bias, with scenarios not evenly distributed around the most likely outcome.
- Provides no information about the likelihood of each scenario occurring
- Needs input from a range of knowledgeable people

Example:

Forecasting where a range of policy options are being considered, particularly where the likelihood of an event occurring is unknown, for example early analysis on Exiting the EU scenarios

Judgement

If the previous methods are unfeasible then you could make a subjective estimate of the overall uncertainty using expert elicitation techniques.

Outline:

If the previous methods are unfeasible, then adding a subjective estimate of the overall of uncertainty is an option. Where there is too little information or time to do a quantified analysis, it may be better to provide a judgement on the uncertainty than nothing at all.

Ideally this should be a group decision, using formal expert elicitation methods (e.g. Delphi) to avoid group-think and arrive at a consensus, agreement, or an average (depending on the technique used). If formal expert elicitation isn't possible, then the judgement of an individual is better than nothing, though less rigorous.

However the uncertainty is estimated, make sure that it is clear that it is a subjective opinion rather than results of analysis to prevent it being misused.

Advantages:

- May be able to be produced quickly, though can take time if using formal elicitation methods with large groups
- Requires little to no data
- Based on real-world performance, so avoids optimism/pessimism bias (the estimated uncertainty may be asymmetric as a result)
- Do not need to consider all sources of uncertainty individually
- Do not need to mathematically combine uncertainties

Disadvantages:

- Highly subjective
- Needs a group of knowledgeable experts, who can reasonably be expected to have a grasp of the range of possible outcomes

Example:

Providing context around a high priority figure that needs to be submitted quickly.

Analysis based on a data source of unknown reliability

Analysis where the expected range of results would lead to the same outcome

5.3. Quality assuring uncertainty analysis and accounting for wider uncertainties

It is best practice to test the outputs of the uncertainty analysis before using/presenting

After modelling uncertainty, you should always test the outputs of the analysis before sharing the results. This minimises the risk of errors in your analysis and helps you to understand the detailed outputs fully, including the level of the extreme or the most likely values.

Uncertainty analysis may produce 'extreme outcomes', so that implausible results or scenarios are given. These can be identified through having a sensible checklist of validation rules (e.g. no negative values allowed) and aided with visualisation and filtering. These implausible outcomes could indicate an issue with the setup conditions of your analysis.

Unusual results may indicate a weakness in the use of the technique

Unusual results in uncertainty analysis may also indicate a weakness in how you have used your chosen technique. For example, if using the Monte Carlo technique, there may be unknown correlation which hasn't been accounted for, or you might have used an inappropriate distribution for a parameter.

Accounting for uncertainties that analysis cannot capture

One element you may not be able to test in your analysis is the effect of unknown system shocks, such as a recession. It is not always useful or practical to investigate and analyse these types of epistemic or ontological uncertainty, but they should be accounted for in presenting your analysis, for example by highlighting the risks associated with the analysis and the decision it will inform.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

6. Presenting and communicating uncertainty

We now consider how to choose *what* messages about uncertainty we should communicate to decision makers and give advice on *how* to communicate those messages effectively. A wealth of additional information can be found on the **GSS website (pdf)**.

When reporting findings from any research and analysis, it is important to be transparent about the strengths, limitations and any bias in the research process as well as the findings

Research or analytical reports should include sufficient and appropriate information about the research process, any sources of uncertainty and the steps that have or could be taken to mitigate this. Where possible it is also good practice to make other materials and outputs from the research available either through publication or upon request to enable others to better understand the context and process by which the research was conducted.

The implications of sources of uncertainty on how results should be interpreted and used in decision-making and policy development should also be communicated.

6.1. Deciding what to communicate

Decision makers may be more comfortable with certain messages, which could make communicating uncertainty a difficult task. However, **latest research** suggests that being upfront about uncertainty doesn't undermine trust in the analysis or the professionalism of the analyst, so we should be unapologetic about uncertainty as part of the analytical advice needed to inform better decisions.

You should also consider the onward communication of your message to ensure that when your work is passed on to others its core message and integrity are maintained. Consider what decision makers *must*, *should*, or *could* know.

Don't let uncertainty undermine confidence in the analysis

Decision makers are usually more comfortable with certain messages. This can make communicating uncertainty a difficult task. Decision makers can react adversely when faced with uncertainty, leading them to lose confidence in analysis. When uncertainty is large, decision makers may feel that the analysis is of no use.

A handy phrase to remember is "all models are wrong, but some are useful". All analysis will be uncertain because every model is a simplification of reality. However, that does not stop them from being useful. Models allow us to better understand a system and uncertain results can help identify important features of that system.

If uncertainty means that a straightforward conclusion is not possible, focus your communication on what the uncertainty analysis *can* tell you. This may be the key factor that drives overall uncertainty or a greater understanding of how features of a system interact.

6.2. Presenting caveats

If most uncertainty is quantified, then present this prominently

If most of the overall uncertainty has been quantified, and you are confident that there are no unknowns which are likely to have a major impact on the results then this should be the most prominent message. Discussion of the unquantified uncertainties and risks can be included but should be positioned so that it doesn't reduce confidence in the main results when this is not appropriate.

If little is quantified, then it may be better to present no quantification at all

If there are substantial unquantified uncertainties, then presenting the uncertainty that has been quantified without this important context will give a misleading impression of precision and underestimate the uncertainty. Do not present a range with incomplete coverage as final analysis if you know that there are substantial uncertainties that are not accounted for in that range.

If the largest source of uncertainty is the potential for a risk outside of the analysis to be realised, then this should be the most prominently displayed message.

Front load the important caveats and explain why they matter

Think about how caveats are presented. A long list is unhelpful, but two or three upfront that have the most impact on the results are likely to be more helpful and easily understood. Consider which caveats have the greater impact on the final decision that is being made. You should explain what the caveats mean for decision makers who want to use the analysis, rather than simply setting out what they are.

In evaluation, the caveats and degree to which you can attribute observed changes to the intervention will vary depending on the methodology

It is not possible to eliminate uncertainty around the causal effects in evaluation. However, study design can be used to minimise the level of uncertainty (see **mitigating uncertainty in experimental and quasi-experimental evaluation methods**). The level of uncertainty depends on components such as uncertainty in data used to assess outcomes, the control group and participant identification strategy. For more details on the sources of uncertainty in evaluation designs, see **Sources of uncertainty in experimental and quasi-experimental evaluation designs**).

Analysts must clearly communicate the extent to which findings can and cannot be generalised

Generalisabiltiy is the extent to which findings can be applied to another setting. Communicating the limitations in the generalisability of your study findings is crucial to ensure that results are not used inappropriately and unsupported claims are made. When using the term 'representative', you need to clearly communicate the context and variables for which the findings are representative. Decision makers need to be made aware of issues which limit generalisability. For example, where studies have small sample size or focus on a specific subpopulation or period of time.

Analysts should also report estimates of sampling error to communicate the uncertainty of results

Sampling error survey data is typically estimated and reported using the **standard error**, **coefficient of variation**, and **confidence interval**.

A confidence interval is a statistical estimate used to communicate the uncertainty around a parameter estimate. A confidence interval is the range of values that is likely to include an unknown parameter, such as the population mean, and the interval has an associated confidence level that gives the probability with which an estimated interval will contain the true value of the parameter. For example, if 80% of survey respondents give a certain response, a 95% confidence interval of [75, 85] indicates that the proportion of the population that would give that response would be between 75% and 85%. The confidence interval becomes smaller as the sample size increases. Effect sizes should be reported with confidence intervals, and what level of uncertainty that represents should be explained in accessible terms.

The ONS's page on **"Uncertainty and how we measure it for our surveys"** provides more detail on these concepts.

6.3. Understanding the audience

Understand your audience when choosing appropriate communication methods

Before communicating your analysis you must consider who your audience is, why they should be interested in your work, what they know already, and what key message you need to convey to them to make their decision. People respond differently to different communication methods. It is good practice to always use Plain English, avoid analytical jargon with non analysts, and frame your results in terms of the decision being made and how it is useful for them.

If it is a group or individual you have worked with before tailor the method to what has worked well in the past (or ask them or see how they respond to different formats).

Build relationships with decision makers

A good relationship with your decision maker will help you to choose the right communication approach for them. It will also help you to understand their motivations and the implications of uncertainty for their decisions.

Be conscious of how your analysis may be used in future communication

No matter how carefully you communicate the uncertainty to your immediate client, there is a risk that uncertainty will not be communicated upwards and that only the central numbers will persist, and important caveats will be excluded. This can also occur when writing part of a larger document – you need to ensure the key messages about uncertainty survive the drafting process.

A good relationship with colleagues will minimise this risk. Work collaboratively to ensure that colleagues understand why central numbers could be misleading and in which situations it would be inappropriate for them to be communicated on their own. Always ensure that wherever possible you have sight of further communication before it is sent.

In evaluations, analysts have a responsibility to take all reasonable steps to ensure that findings from quantitative analysis are used appropriately and not misconstrued and must be clear about the level of confidence in findings, taking account of the level of uncertainty and generalisability. For some methodologies it is particularly important to ensure that audiences understand limitations and the implications for use in policymaking; for example, where analysis relies solely on limited datasets or where there is ecological fallacy (discussed in **Defining and identifying uncertainty**).

Consider how to present measures of uncertainty to a non-analytical audience

If presenting results to a non-analytical audience, bear in mind that they may find statistical measures of uncertainty difficult to interpret. Consider other visual methods such as a BRAG (blue/black, red, amber, green) rating scale that communicates your confidence in the results.

6.4. Deciding how to communicate uncertainty

Now you have determined who your audience is and what the message is you need to consider how to communicate with your decision makers. Firstly, think about how you will be communicating your results. Will this be via a written report, presentation, or an oral briefing with or without visuals? Any constraints on format may steer you to an approach.

This section considers a range of approaches based on your understanding of the audience and the type of message you need to deliver.

6.5. Communication basics

Identify who your audience is

As mentioned earlier, knowing who your audience is, what their interest is in this analysis, and how they interpret evidence will determine how you communicate. Test your communication with the audience, do they understand the key message from your prose or visualisations?

Assess whether to describe uncertainty in words or attach figures

How something is written has a major impact on how people perceive the uncertainty. **Research shows** that analysis is seen as less reliable if the outputs are conveyed only in words. Numbers should be presented within a sentence if possible as this helps give context, making them easier to read and understand.
To avoid any possible confusion between analytical conclusions and the uncertainty surrounding them, it is **good advice** to "keep your expressions of the magnitude of uncertainty clearly separate from the magnitude of any evidence you are trying to communicate", while ensuring that caveats to the overall conclusions cannot be disregarded.

Be careful when using descriptive terms and adhere to an established system if possible

Descriptive terms such as 'low risk' or 'very likely' can be interpreted very differently by **different people**. Where possible, it is sensible to attach a numerical probability and use these descriptive terms in support. **Emerging research** suggests that numerically defined uncertainty is more trusted than worded statements, and, in fact, people find worded statements of uncertainty difficult to interpret and may entirely ignore the implication of worded statements in demonstrating that numbers are uncertain.

Nevertheless, these scales can be useful provided the scale being used is well-understood by the intended audience. Setting out scales used alongside analytical presentations is strongly recommended to avoid confusion .

An exception to this is where there is an established system that your audience is used to for attaching terminology to probabilities and mapping between verbal, pictorial and numerical articulations of uncertainty. There are a variety of scales like this and they differ considerably in how they align verbal expressions to numerical articulations:

- **According to the IPCC (Intergovernmental Panel on Climate Change) (pdf)**, "very likely" means 90-100% probability.
- **According to NICE (National Institute for Health and Care Excellence)**, probabilities of between 1 in 100 and 1 in 10 are referred to as "common".
- **• According to GLD (Government Legal Department)**, "high likelihoods" are those greater than 70%.
- **According to PHIA (Professional Head of Intelligence Assessment)**, >=95% is "almost certain".

Use positive and negative framing

Presenting the likelihood of success **may be perceived differently (pdf)** to presenting the corresponding likelihood of failure. Present the information both ways to avoid bias (e.g. "there is an 80% chance of success and a 20% chance of failure"). By adding in the chance of failure, you remind the reader that it exists and how large it is, which may otherwise be overlooked. Visual part-to-whole comparisons can help with this. See **infographics section** for more details.

Decide how to present your numbers

Research shows that there is no clear preference for choosing between probabilities and fractions (e.g. 10% probability, or 1 out of 10). Given this, **the usual preferences of the audience or the 'norm' within the organisation should be followed (pdf)** and you should stick to the same format (i.e. do not mix probabilities, fractions, ratios in the same report). . Some people are more familiar with percentages than fractions or ratios and articulations such as "1 in 100" in **research** contexts can risk the misinterpretation that research has only been done on 100 people.

If using fractions, **keep the denominator constant** (e.g. "1 in 100 vs. 2 in 100", rather than "1 in 100 vs. 1 in 50") and as small as possible while keeping to integers (e.g. "1 in 100" rather than "10 in 1,000"), rounding if appropriate.

Be clear on the specifics and applicability

Saying a '10% chance of rain' is meaningless unless you also state the time period – e.g. in the next hour, or at some point tomorrow – and location.

If the outputs are only intended for use within a specific frame (for example, a specific area or to a particular group of the population), then make sure this is clearly stated alongside the outputs.

Use an appropriate level of precision

Consider the overall uncertainty in the numbers you have calculated. Round them appropriately to avoid spurious accuracy (e.g. perhaps 40% rather than 38.7% if the overall uncertainty is greater than one percentage point).

Use ranges wherever possible

Presenting a single figure is best avoided as it can give a misleading impression of precision (e.g. "between 1,200 and 1,800", rather than "1,500"). Further, emerging research (see **here** and **here**) suggests that ranges are better than point numbers for ensuring that decision makers understand that a number is uncertain at all.

Consider whether to include a 'best estimate' within the range

Commissioners may request a 'best estimate' for ease of onward use, but you must consider the risks in providing this. Try to understand how they intend to use the analysis, so you can provide something that meets their needs while also acknowledging the uncertainty.

Stating a range may be perceived as a uniform distribution across the range. Conversely, stating a range around a best estimate may be perceived as a triangular distribution (or Normal with analytical audiences). The output distribution could of course also be asymmetric or bi-modal. Consider which of these best reflects the actual uncertainty when deciding what to present.

Choose appropriate confidence/prediction intervals and be clear

Don't simply use 95% confidence intervals by default. Think about what the outputs are going to be used for (see **Agreeing how uncertainty should be used to help answer the question** section), and discuss the level of risk and uncertainty that the decision maker wants to plan for – this might not be 5%.

Be clear what confidence level you are using and ensure your audience understands what this means (avoiding precise statistical definitions if it will increase comprehension).

Always label and pull out the key message from a visualisation

People interpret visualisations differently. Always have a title with the key message, otherwise people may not have the time to interpret the visualisation or misinterpret the key message.

6.6. Graphs and visualisation

Graphs can be an excellent way of communicating the quantified elements of uncertainty

Graphs and visualisations are an excellent way of communicating the outputs of analysis, and many graph types allow you to communicate uncertainty within the graphic (provided the uncertainty has been quantified).

Unquantified uncertainties cannot generally be included in graphs, so will need to be communicated through other means (e.g. a risk log and/or assumptions log).

Some graph types cannot be used to show uncertainty clearly

Some types of graphs are not particularly well suited to displaying quantified uncertainty, though this does not preclude their use if they are the most useful way to communicate a core message. You would need to find other ways to communicate uncertainty if using these chart types.

Some simple graphs e.g.
Pie charts, donut charts, stacked charts

May not always be useful when presenting uncertain analysis as they only show a single value for each data point. People can also find it difficult to compare angles or sizes of different sections.

Avoid more complex graphs e.g.
Heat and Choropleth maps, Treemaps, Sankey diagrams

as it can be difficult to communicate uncertainty in very information dense visualisations. There is not space to include the necessary extra information among the colours of heatmaps, the densely packed rectangles of Treemaps or the detailed flow lines of Sankey diagram.

Decide what level of detail to include on uncertainty

You may have the full understanding of the underlying probability distribution, or just a range within which we expect the result to fall. You may choose to only include the uncertainty due to a single dominant uncertainty, or the outputs from a range of scenarios.

For representing numerical uncertainty, scatter plots with 2-dimensional error bars, line graphs with a range, bar or line graphs with error bars and single point graphs with error bars can be useful when representing an uncertain range. Box plots or box plot series can be useful when representing summary statistics. Probability (or Cumulative) Distribution functions are helpful if you have full knowledge of the distribution around a result.

For representing uncertainty due to alternative scenarios, multiple line graphs and descriptions in prose may be most helpful.

6.7. Error bars

Error bars are a simple way to illustrate a range around a data point

Error bars can be added to bar graphs, line graphs and scatter graphs to illustrate a range around a central estimate, within which we expect the value to lie with a given probability. Be aware that a non analytical audience may be less familiar with error bars.

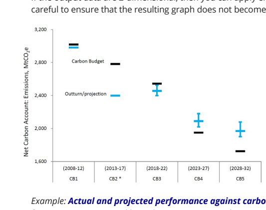
As referred to previously, consider the situation and decide on an appropriate level to display. E.g., don't apply 95% confidence/prediction intervals by default.

State what probability the error bars represent, and describe in prose how the viewer should 'read' the error bar.

Error bars can be applied to series of data points

Error bars can be added easily to a data series or time series. However, if the data are continuous (e.g. a time series) then consider whether showing multiple line graphs would be clearer than a single line graph with error bars.

If the output data are 2-dimensional, then you can apply error bars in 2 dimensions. Be careful to ensure that the resulting graph does not become illegible due to clutter.



*Example: **Actual and projected performance against carbon budgets, BEIS (pdf)** For future emissions, vertical bars show uncertainty in the projections and indicate 95% confidence intervals for the uncertainties that have been modelled.*

6.8. Box plots

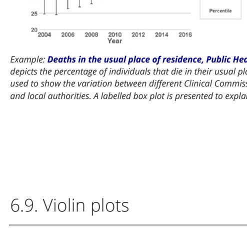
Box plots can convey more information about possible outcomes than a range alone

Box plots can help the audience understand the underlying distribution of possible outcomes in more detail than just a range. Typically they show the median, interquartile range, maximum and minimum values for the range of possible outcomes. This can be particularly useful when the underlying distribution is skewed or non-normal.

Box plots can be arranged in parallel to show the distributions for a range of measures, and can help compare the different shapes.

Think about whether the audience will be familiar with the format

Box plots may not be widely understood by non-analysts, so think carefully about whether the added information will be effective, or whether a simple range would be sufficient. Remember that most non analysts do not use the terms mean, median, inter quartile range frequently and may not know what they are. A labelled example can be used to help the audience interpret the format.



*Example: **Deaths in the usual place of residence, Public Health England** The graph depicts the percentage of individuals that die in their usual place of residence. Box plots are used to show the variation between different Clinical Commissioning Groups, and district and local authorities. A labelled box plot is presented to explain what the ranges mean.*

6.9. Violin plots

Violin plots give a sense of a distribution's shape by showing probability densities

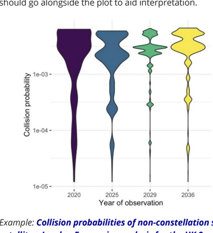
Violin plots are similar to box plots except they also show the probability density of possible outcomes, where a greater width means higher probability. This reinforces awareness of an underlying distribution and gives an intuitive sense of the distribution's shape, although it can be difficult to infer specific probability values.

Violin plots can help avoid bias

Violin plots can help to avoid bias which can occur when interpreting intervals (for example when the uncertainty region is ignored in favour of the central tendency) or when reading error bars (for example within-the-bar-bias, common in bar charts). Within-the-bar bias occurs when viewers report that values are more likely to lie within the bar of a bar chart despite error bars indicating they could equally lie outside.

Consider whether the audience will be familiar with this type of plot

The audience might find interpretation of violin plots hard if they are unfamiliar with them. Weigh up the value added from the additional detail on uncertainty against the increased potential for audience confusion and consider what additional commentary should go alongside the plot to aid interpretation.



*Example: **Collision probabilities of non-constellation spacecraft and constellation satellites, London Economics analysis for the UK Space Agency (pdf)** The graph shows how the collision probability for a subset of 216 non-constellation spacecraft and over 16,000 constellation satellites changes across the years 2020, 2025, 2029 and 2036.*

6.10. Probability density functions (PDFs)

PDFs show complete information on the quantified uncertainty

A probability density function can be used to give complete information on the range of possible outcomes, and the likelihood of each for a given estimate.

Think about whether the audience needs this much information

While presenting complete information may seem ideal, it may be more information than the audience actually needs. Would a prose description of the mean and range be sufficient? If the PDF is approximately normal, then there may be little value in displaying it, as the essential features can be described in a few words.

PDFs can be useful when the distribution of outcomes is multimodal, or otherwise complex

The distribution could be multimodal – for example the marks for students passing a university module may have a peak for a number who don't study very hard, and a peak for those who worked really hard. In this case it could be misleading to present the mean, so a graphical illustration of the distribution may be more effective.

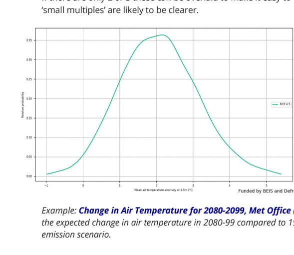
Labelling can be used to highlight the key features

It may aid clarity to draw the reader's attention to important features, such as the mode.

Multiple PDFs can be used to show uncertainty across different measures

If we need to communicate a series of PDFs, then multiple functions can be shown to compare the range of possible outcomes across the series.

If there are only 2 or 3 these can be overlaid to make it easy to compare. With more, 'small multiples' are likely to be clearer.



*Example: **Change in Air Temperature for 2080-2099, Met Office (pdf)** The graph shows the expected change in air temperature in 2080-99 compared to 1981-2000 for a medium emission scenario.*

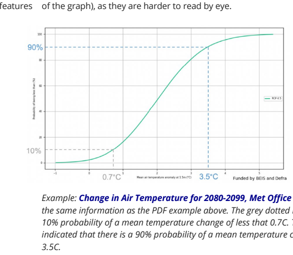
6.11. Cumulative density functions (CDFs)

A CDF may be more helpful than a PDF if there is a specific threshold of interest to the customer

A cumulative density function shows similar information as a probability density function, but cumulatively. A CDF may be more helpful when the audience is primarily concerned with how likely it is that the value will be below (or above) a particular point (rather than the range within which we expect the value to fall). For example, how likely is it that our costs exceed our budget? Rather than, what are our costs going to be?

Labelling can be used to highlight the key features

However, features such as the mode are less clear on a CDF (shown by the steepest part of the graph), as they are harder to read by eye.



*Example: **Change in Air Temperature for 2080-2099, Met Office (pdf)** The graph shows the same information as the PDF example above. The grey dotted lines indicated there is a 10% probability of a mean temperature change of less that 0.7C. The blue dotted line indicated that there is a 90% probability of a mean temperature change being less than 3.5C.*

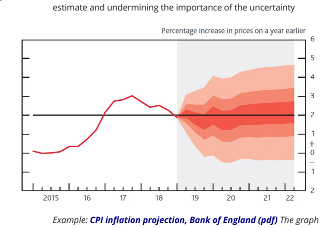
6.12. Fan charts

Fan charts can show how uncertainty changes over time

Fan charts can be used to show a series of different prediction intervals for time-series projections (e.g. 30%, 60% and 90% at the same time). This is essentially plotting selected points from a time-dependent PDF.

Avoid including the mode

Often a central 'best estimate' is not included, to avoid the viewer focussing on a single estimate and undermining the importance of the uncertainty



*Example: **CPI inflation projection, Bank of England (pdf)** The graph depicts probability of various outcomes for CPI inflation in the future. The fan charts are constructed so that outturns of inflation are also expected to lie within the darkest central band and each pair of the lighter red areas on 30 occasions. In any particular quarter of the forecast period, inflation is therefore expected to lie somewhere within the fans on 90 out of 100 occasions. And on the remaining 10 out of 100 occasions inflation can fall anywhere outside the red area of the fan chart (grey area).*

6.13. Multiple line charts

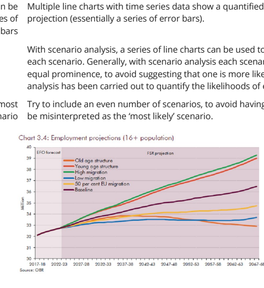
Multiple line charts can be clearer than a series of error bars

Multiple line charts with time series data show a quantified range around a 'most likely' projection (essentially a series of error bars).

With scenario analysis, a series of line charts can be used to show the projections from each scenario. Generally, with scenario analysis each scenario should be presented with equal prominence, to avoid suggesting that one is more likely than another (unless analysis has been carried out to quantify the likelihoods of each).

Avoid a middle "most likely" scenario

Try to include an even number of scenarios, to avoid having a middle option that may be misinterpreted as the 'most likely' scenario.



*Example: **Employment Projections, OBR (pdf)** The graph presents future employment levels for different demographic scenarios. Each scenario is represented by a separate line plot. The graph does not attempt to show the probability of each scenario occurring.*

6.14. Tornado diagrams

Tornado diagrams can be used to show the *sources* of uncertainty

Tornado diagrams are different to most other graphs discussed here. They are not used to show the outputs of the analysis, but to show how different sources of uncertainty contribute to the overall uncertainty.

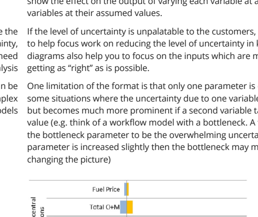
Tornado diagrams depict sensitivity of a result to changes in selected variables. They show the effect on the output of varying each variable at a time, keeping other input variables at their assumed values.

Can help communicate the reasons for uncertainty, and identify further need for analysis

If the level of uncertainty is unpalatable to the customers, then this format can be useful to help focus work on reducing the level of uncertainty in key parameters. Tornado diagrams also help you to focus on the inputs which are most important to focus on getting as "right" as is possible.

Tornado diagrams can be misleading in complex models

One limitation of the format is that only one parameter is changed at a time. There are some situations where the uncertainty due to one variable may appear small initially but becomes much more prominent if a second variable takes on a slightly different value (e.g. think of a workflow model with a bottleneck. A tornado diagram might show the bottleneck parameter to be the overwhelming uncertainty. However, if this parameter is increased slightly then the bottleneck may move elsewhere, completely changing the picture)



*Example: **Nuclear Levelized Cost of Electricity, BEIS (pdf)** . The graph presents the change in the levelized cost of nuclear electricity (£/MWh) that would result from changes in input parameters. Impacts are shown from both 10% upward or downward movement in central estimates and from predetermined parameters ranges (high/low data range). Yellow bars represent an increase in the parameter, blue bar represent a decrease in the parameter.*

6.15. Infographics

Infographics can be useful for public facing communications

Infographics are graphic visual representations of information, data or knowledge intended to present information quickly and clearly. **They can improve people's understanding (pdf)** by using graphics to enhance peoples' ability to see patterns and trends.

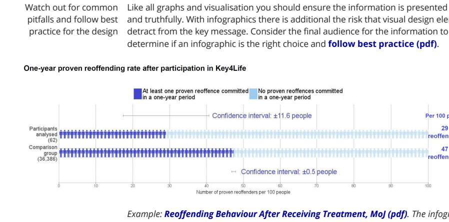
Graphics can grab attention and make messages more accessible

When done well they will grab the reader's attention and become a very powerful way of communicating key messages. Designing a good infographic may be worthwhile if your audience is less confident with data and analysis.

Not all infographics communicate uncertainty, but uncertainty information can be included within the infographic. In the example below, confidence intervals are included as part of the infographic .

Watch out for common pitfalls and follow best practice for the design

Like all graphs and visualisation you should ensure the information is presented clearly and truthfully. With infographics there is additional the risk that visual design elements detract from the key message. Consider the final audience for the information to determine if an infographic is the right choice and **follow best practice (pdf)**.



*Example: **Reoffending Behaviour After Receiving Treatment, MoJ (pdf)**. The infographic uses people icons rather than a bar chart to show the number of reoffenders in a sample. An uncertainty in the form of confidence intervals is included in the infographic*

6.16. Interactive tools

Interactive tools can be used to immerse your reader on complex matters

An interactive tool can help make analysis more accessible to non-specialists, assuming the non-specialists have the time to interact with the data. They can create an immersive experience that is easier for them to understand and is highly memorable. Note, that they often take longer to create and sometimes the key message may be lost.

Focus on specific messages

Consider the overall message and where the uncertainties lie. Which aspects will the audience be interested in and what do they need to know? Remember with interactivity the user chooses what to look at. Will they find all your key messages in the tool or should these be highlighted somewhere? Use this understanding to bring focus to which interactive elements to create.

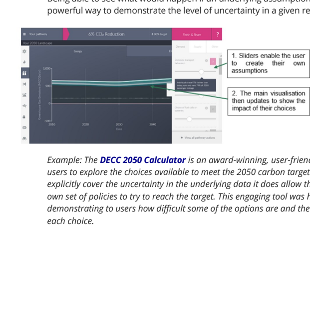
The interactivity will enable your users to manipulate and get a deeper understanding of the message.

Allow reader to adjust a key variable

If a key source of uncertainty is a single variable, then it may be possible to construct a display that can be changed as the user adjusts the value of this variable by moving a slider.

Or, if there are several key assumptions that impact the result a chart may be created that will change depending on the inputs that the user inserts.

Being able to see what would happen if an underlying assumption was to change is a powerful way to demonstrate the level of uncertainty in a given result.



*Example: The **DECC 2050 Calculator** is an award-winning, user-friendly tool that helps users to explore the choices available to meet the 2050 carbon target. Whilst it doesn't explicitly cover the uncertainty in the underlying data it does allow the user to create their own set of policies to try to reach the target. This engaging tool was helpful in demonstrating to users how difficult some of the options are and the relative impact of each choice.*

6.17. Communicating uncertainty in qualitative research

The key principles of presenting and communicating uncertainty outlined in this section of the toolkit are also applicable when working with outputs from qualitative research. The only exception to this is the use of numeric data when presenting qualitative research findings. This is generally not considered good practice due to the small sample sizes and diverse make-up of the population.

Recap of key principles

- Be transparent about the strengths, limitations and bias in the research process and their implications for the findings.
- Decide what to communicate: start by deciding what to communicate, and consider what decision-makers must, should, or could know.
- Present caveats: focus on those which will have the biggest impact on and implications for the decision.
- Understand the audience: when writing for or presenting to non-analysts it is important to use plain English and avoid jargon. When analysis is being used to inform a decision, results should be framed with the decision in mind.
- Consider the reporting format: the level of detail and options for how to present findings and uncertainty will depend on the required communications format – e.g. submission, report, presentation, oral briefing.

The remainder of this subsection includes guidance on how to communicate uncertainty when presenting findings from qualitative research.

Communicate uncertainty in the research design

Section 3 of the toolkit details the main sources of uncertainty for qualitative research. These include: a lack of representativeness of the population and sub-groups within it, potential biases due to researchers' subjective judgements in the process of recruitment and analysis, and differences in how research participants engage with the research process.

When reporting findings from qualitative research it is important to be transparent about the limitations, bias and uncertainty in the research design and the implications for how findings should be interpreted and used in decision-making.

Particular attention should be given to the rationale for the sample design and the judgements made when setting quotas and deciding on the number of participants recruited. The latter is especially important when it has not been possible, due to resource constraints, to undertake fieldwork until thematic saturation has occurred. Failure to do so means it is not possible to state that the findings are conclusive, and no new themes, insights or perspectives are likely to emerge. Including the rationale for and implications of the selected approach to data collection and the analysis is also beneficial.

It is also good practice to include the research tools (e.g. screening questionnaire and topic guide) in an annex. Although it is not common practice to share full details of the analysis process and tools used, including high level information about this could be beneficial to demonstrate the level of rigour applied.

Avoid presenting numbers

As sample sizes in qualitative research are usually small and the research predominantly uses open and exploratory questions, it is not appropriate to report fractions, percentages or proportions when reporting findings. Instead use descriptive terms such as 'few', 'many', 'majority' to denote prevalence of an issue or finding. This will prevent misinterpretation or inappropriate use of data. There is software that can support management and analysis of qualitative data. This can be especially helpful when dealing with large volumes of narrative data and when analysis is being undertaken by a team of researchers. This software can be used to add codes to qualitative data and to count the number of times particular words or themes are expressed. However, this analysis can take findings out of context and be misleading. Whilst it may be useful to use these functions to support interrogation of qualitative data as part of the analytic process, it is not appropriate to use the numeric outputs when communicating and presenting findings. Finally, it is unlikely that uncertainty in qualitative analysis can be quantified. Whilst some endeavours have been made to do this within natural scientific fields, application has not been extended to the social sciences where it is usual to present uncertainty in narrative terms.

Avoid making generalisations

Qualitative research findings will only ever be reflective of the views, behaviours and perceptions of the individuals participating in the research. Whilst these findings may be indicative of the attitudes and behaviour of the wider population they cannot be presented as representative. Whilst it may be possible to make generalisations about the wider population from the data and to create typologies or personas to exemplify the research findings, it is not possible to extrapolate findings to or make inferences about the population at large. When reporting findings, be up front about whether and how far statistical inference is valid and avoid inferring from one group to another.

Contextualise illustrative examples

You can use quotations, vignettes and case studies to bring qualitative findings to life and increase resonance and depth of understanding of a particular issue or viewpoint. When doing so it is important to ensure that they are set in context within the report, so it is clear what they are trying to illustrate. This mitigates the risk of misattribution, misinterpretation and misuse of these tools by the audience. It is especially important if it is expected that research findings will be used to support other communication activities.

Triangulate and validate data

Grounding qualitative findings in existing evidence can demonstrate the extent to which the new findings support or contradict previous theories and the likely reasons for this. This can be an effective way of reinforcing key messages to the audience, as well as mitigating uncertainty. Other data can include past qualitative and quantitative studies, a review of wider literature, as well as new quantitative research to test the prevalence of qualitative findings within the wider population.

If further research is required to validate new qualitative findings and determine if they can be generalised to the wider population, it is worth highlighting this to decision-makers. This will mean they understand the limitations of the current research and it can be used as well as how it can be strengthened.

6.18. Uncertainty messaging in onward communication

Be wary of message distortion

A key challenge that you will face when trying to communicate uncertainty messages is ensuring that the integrity of these messages is maintained when they are communicated onwards, beyond your immediate client. What we are trying to avoid here are situations in which your uncertainty message becomes "distorted" as it passes from client to client. This often manifests itself as important uncertainty information or other caveats being removed from the message, leaving only central numbers persisting in information exchanges. We can define the problem of message distortion using the following terms:

- Immediate Client - Any individual/group that receives information directly from you whether this is verbally or from reading your publication. This group will receive your uncertainty message in its purest form.
- Secondary Client - Any individual/group that receives your uncertainty message through an immediate client and not verbally from you or from reading your publication. This group is not guaranteed to receive your uncertainty message in its purest form and instead may receive a distorted version depending how effectively it is communicated to them by your immediate client.
- Message distortion - A situation where a secondary client receives an incorrect version of your uncertainty message from an immediate client. Note that this framework largely deals with cases of misinformation only in which the sharing of inaccurate information is inadvertent. Detailed guidance on disinformation, the deliberate creation/sharing of false information, can be found in the **RESIST** toolkit

The challenge is to limit the extent of message distortion in the information exchange between your immediate and secondary clients.

Understanding your immediate clients

When passing uncertainty messages to be communicated onwards, it is important that you consider who your immediate clients are, how they are likely to convey your message and whether you are likely to communicate with them again. We can condense the attributes of your immediate client as falling into two core characteristics:

- Ambiguity – This is the extent to which you are unsure that your analytical message will be passed on accurately to secondary clients.
- Accessibility – This is the degree of contact that you have with your immediate client i.e. Is this someone who you sit in meetings with on a daily basis or an anonymous social media user that you are never going to meet in person?

Table A: The Ambiguity/Accessibility Matrix

Ambiguity/Accessibility	High Accessibility	Low Accessibility
Low Ambiguity	Case A: Low Ambiguity, High Accessibility	Case B: Low Ambiguity, Low Accessibility
High Ambiguity	Case C: High Ambiguity, High Accessibility	Case D: High Ambiguity, Low Accessibility

Examples of immediate clients:

- Case A: Low Ambiguity/High Accessibility – A trusted policy colleague who you have worked with for many years. In this situation it is likely that simply maintaining a good relationship is all that is required for them to pass on your message accurately.
- Case B: Low Ambiguity/Low Accessibility – A well-respected figure in academia that works in a similar field to your work, likely to appreciate uncertainty messages in releases but has a lower level of accessibility due to being outside of the organisation.
- Case C: High Ambiguity/High Accessibility – A policy colleague who has little analytical/statistical background and has shown little desire to embrace uncertainty messages in the past and instead prefers to present a "clearer" message without any of the necessary caveats.
- Case D: High Ambiguity/Low Accessibility – A anonymous user on twitter who reads a tweet that features your publication, you have no knowledge of who the immediate client is in this situation and so there is a large amount of ambiguity combined with minimal accessibility.

Tailor your actions depending on the nature of your immediate client

The set of possible responses when communicating uncertainty in analytical messages can be thought of as falling into three categories:

- Communication: Actions related to the way in which you communicate with your immediate client.
- Content: Actions concerning what is communicated to the immediate client.
- Continuity: Any actions to support and follow-up on any initial communication to the immediate client.

Table B: Action Matrix

Ambiguity/Accessibility	Communication	Content	Continuity
A. Low Ambiguity / High Accessibility	• No change needed from an established method of communication that is known to work • If possible, increasing the amount of onward communications that go through this person would be an effective way to make use of a relationship with a reliable contact.	• No change is needed in terms of what content is communicated, given that there is confidence that uncertainty in analysis will be sufficiently communicated forwards. • More detailed and nuanced uncertainty messages can be given to this client, due to a low level of ambiguity combined with high accessibility.	• It is important that this good relationship (an appreciation of each other's perspectives and needs) is maintained. A high level of accessibility should make this easy to do so.
B. Low Ambiguity / Low Accessibility	• Due to infrequent contact, there should be some verbal component when passing information as this gives the opportunity to clear up any misunderstandings immediately rather than in the future by which point a distorted message may have already been passed to several secondary clients. • Asking clarifying questions at this stage is a good way to ensure that uncertainty messages have been interpreted as intended.	• Due to a low level of ambiguity, content is likely to be communicated onwards accurately as long as it is initially interpreted correctly i.e., the risk of disinformation is low. • Therefore, while more focus should be placed on effective communication, the complexity of any uncertainty messages should be made appropriate to the level of expertise of your immediate client in order to ensure they are interpreted in the desired manner. • There is evidence (**Ribeiro et al. 2019, pdf**) to suggest that summarising key information is an extractive way (by sub selecting key phrases), as opposed to abstractive summarisation (using paraphrasing), is a more effective way of preventing distortion when information is cascaded.	• A clear progression for this type of action is to take steps to increase accessibility. This is largely done through developing a relationship with this party. • This could take the form of more regular meetings to discuss common interests or establishing a feedback mechanism on the outcomes of previous communications.
C. High Ambiguity / High Accessibility	• High accessibility means that securing face-to-face/verbal contact isn't likely to be an issue • Instead of simply viewing the individual/group as being untrustworthy, it is important to understand their motivations for the way in which they view uncertainty messages. This will allow you to anticipate any message distortion.	• Focusing on cases of misinformation, a lack of regard for uncertainty messaging often stems from a lack of knowledge on why these caveats are important. • As a result, more emphasis should be placed on why it is important these uncertainty messages are communicated onwards, not just what messages need be communicated.	• Importance of communicating uncertainty can be stressed by pointing colleagues towards relevant materials, such as the GSS guidance on **Communicating Quality, Uncertainty and Change** • Take opportunities to ensure that your message is communicated appropriately at critical moments e.g. Some departments have run trials in which analysts sit with Press Officers on the day of a publication release to help field any questions.
D. High Ambiguity / Low Accessibility	• Social media still very important in broadcasting to a wider audience however, it is clearly very difficult to control how an anonymised recipient uses this information. As such, the focus here should be on ensuring that the correct analytical message is presented on whatever platform the information is circulated on. • Maintaining a "sign-off" on any information released relating to analysis is helpful to ensure that the right information is communicated. • In all cases, it is again a useful exercise to anticipate potential ways in which uncertainty messages may be distorted in onward communication. There are several examples of publications including graphics outlining what the output can and can't be used for.	• "Tightness" of message is crucial in this scenario. Clear and unambiguous messaging leaves as little room for distortion as possible. • There is evidence from the field of health that exaggeration in news is strongly correlated with exaggeration in press releases (**Sumner et al. 2014, pdf**). Therefore, maintaining accuracy in any form of press release is important. • Accompanying releases on these platforms with links to "explainers" written in plain-English will mitigate misinformation spread to some extent. • The GSS **Communicating Quality, Uncertainty and Change** guidance offers advice on how to successfully communicate information to users when under character limits imposed by platforms such as Twitter.	• Responses to your release can be monitored to flag any common misconceptions that emerge. Subsequent FAQ communications can then be released to tackle these. • Keeping track of the volume and type of responses is also useful in determining what dissemination platforms are effective in communicating uncertainty messages.

A number of themes sit behind the set of actions outlined in the Action Matrix. These are additional factors that apply to all of the scenarios outlined by the Ambiguity/Accessibility Matrix:

- **Understanding who your clients are and how they use your analysis** – In all contexts, it is important to try to understand how your client plans to use this analytical information and what their motivations are. Studies show that people selectively seek out information that is consistent with their prior beliefs and sometimes process this more fluently than information that is inconsistent with their prior beliefs. The client's pre-existing beliefs or attitudes towards you, the topic or object of uncertainty might influence the effects of any uncertainty communication and, in turn, how the client passes this information onward (**van der Ples et al. 2019**). Therefore, in order for uncertainty to be communicated effectively, messages should be tailored to clients depending on knowledge of these factors.
- **Guiding your clients on how to use your analysis** – An inquiry into the **governance of statistics (pdf)**, published in July 2019, by the Public Administration and Constitutional Affairs Committee (PACAC) produced a recommendation that government statisticians 'should do more to guide users how to use their statistics, explain how they are typically used, outlining their strengths and weaknesses, providing commentary and advice' (point 46). Applying this to all analysis, even when information is passed onward from the most accessible and least ambiguous client, the onus is on you to ensure clarity of message, something which is greatly aided by this type of communication.
- **Client Expertise/Knowledge** – The complexity of uncertainty messaging in all contexts needs to be tailored towards the level of expertise of your client i.e., their ability to correctly understand and use the information. Providing messages to a low-expertise client that are too complex carries the risk of them accidentally distorting the information due to a lack of knowledge, irrespective of ambiguity/accessibility levels. Conversely, delivering an overly simplified message to a high-expertise client represents a missed opportunity to communicate more nuanced uncertainty messages.
- **Communicating uncertainty within the presentation of analysis** – Uncertainty caveats can be embedded directly into analysis in a variety of ways, for example, by presenting figures with confidence intervals. It then requires a deliberate action from the immediate client to remove this information, whereas it would be much easier to deliberately or accidentally drop footnotes or other contextual information that is not embedded into the analysis.

7. Conclusion

This toolkit is designed to provide guidance to analysts on how to incorporate uncertainty analysis into their work, as well as setting out some high level advice to decision makers, so they know what to expect when commissioning analysis and interpreting analytical results.

See the Case Studies for examples of applying and communicating some of the techniques listed **here**.

Additional papers that may be of interest are presented in the **Further Reading** tab. Your department may also have an uncertainty group who can advise on specific techniques and their implementation.

This toolkit has set out good, not best, practice, as analysis and communication must always be tailored to the audience and decision being made.

Please get in touch if you have any comments by emailing: **AnalystsUncertaintyToolkit@homeoffice.gov.uk**
We also welcome your feedback via our **short feedback survey**

8. Case studies

Estimating the impact of evidence volumes in Legal Aid spend

The situation:

Prices in Crown Court Legal Aid work had been undergoing a sustained increase. Several drivers were identified that suggested that these increases would continue, such as:

- Increased electronic evidence
- Media focus on collapse of trials due to lack of evidence
- New guidance on the disclosure of data

However, there was no information to determine for how long these increases could continue. In addition, prices had risen particularly quickly in more recent months and may be an indication of an accelerating trend.

The analytical approach:

The existing model takes predicted court activity and calculates expected bill volumes, which a price forecast is then applied to.

Asking the right question:

- What should the price forecast be set at for Crown Court Legal Aid?
- What is the risk of over or underspending against this forecast?

Defining and identifying uncertainty:

- The timing and value of upcoming bills
- Potential changes in underlying cases mix
- Case volumes

Understanding and measuring uncertainty:

- Uncertainty around case volumes was discounted (**Dominant Uncertainty**) as the impact was expected to be small.
- No information was available to determine the scale of future increases, so scenario analysis was used to assess the impact of certain assumptions.
- The impacts of 3 price scenarios, representing price increases continuing for 3, 6 or 12 months, were estimated.
- Two additional scenarios were considered:
- An increasing level of prices. Due to insufficient evidence to support it being considered for the main estimate, it was logged in the risk register to highlight the potential variance.
- Prices levelling out immediately in case the full impact had already fed through. Due to the volume and timing of identified drivers this was felt to be very unlikely, so including it could be misleading and contribute to an optimism bias to the decision making.

Presenting and communicating uncertainty analysis

- Scenarios were presented taking the 3, 6 or 12 month assumptions for the central projection, with the higher and flat scenarios being logged in the risk register. The communication stressed that there was no information that would allow us to tell which scenario was more likely.
- Decision makers agreed that the range of scenarios considered reflected the range of reasonable outcomes.
- As no information was available on the which scenario was mostly likely, discussion focused on the risks that were associated with taking each scenario through for planning purposes, such as over or underspending against the overall budget.
- Scenario analysis was used to provide a range of values to illustrate the large amounts of uncertainty around the forecast without giving the impression that any particular outcome was more likely, allowing a decision to be made with the understanding of the risks involved.

Estimating uncertainty in prison population projections

The situation:

Prison population projections are the end result of a number of criminal justice system interconnected models. The uncertainty around each of these models had not been estimated, but understanding the uncertainty of prison projections was essential for capacity planning.

The analytical approach:

Court demand and sentencing and forecasts were used within a prison population microsimulation model, with policy forecasts overlaid.

Asking the right question:

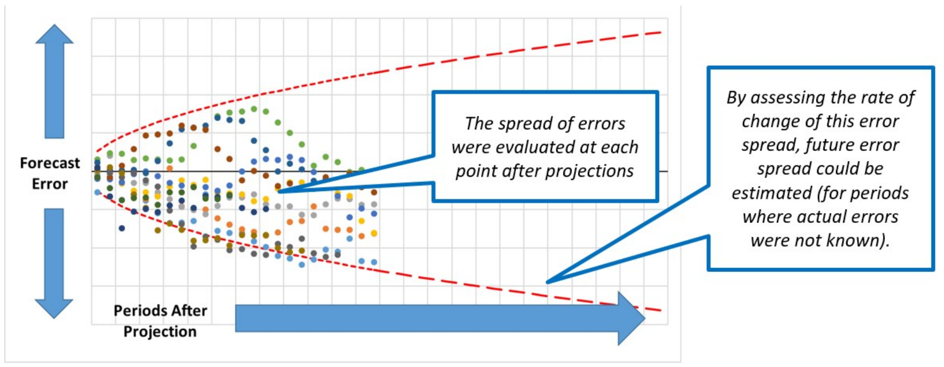
- How confident are we that the prison population will not exceed a particular threshold?
- Was there sufficient resource to estimate uncertainty using a traditional bottom-up approach (i.e. combining the uncertainty from all constituent models)?

Defining and identifying uncertainty:

- Court demand projections
- Sentencing assumptions (custody rate)
- Impact estimates of future policy changes
- Unknown impacts such as exogenous system shocks or policy changes not known at time of projection
- Estimation error (such as sampling bias) from simulation model

Understanding and measuring uncertainty:

- A top-down approach was chosen as the resource requirement for a traditional bottom-up approach would have been prohibitive.
- The variance of all the previous forecast errors after 1 time period were fitted to a chosen probability distribution (in this case a normal distribution was chosen). This distribution was used to estimate the uncertainty of the new forecast after 1 time period.
- This process was repeated for the variance at each subsequent time period.
- The rate of change of variance was used to extrapolate the variance for periods where no forecast error existed.



Presenting and communicating uncertainty analysis

- We presented our historical forecast errors, explaining their main drivers and how each contributes to our current uncertainty range. This helped our key customers to better understand the uncertainties inherent in the system we model, and therefore why we can't produce more precise forecasts.
- The other aspect was helping customers think about how they should use the uncertainty estimates. For example: for the prison estate, assessing the likelihood of prison population reaching different levels together with an assessment of risk appetite and contingency measures we could use. This allowed for an evidence based approach for how much capacity we should plan for, not simply planning to the central estimate.

Estimating changes to staff travel times following an office relocation

The situation:

Alternative government office locations were being considered and their on staff travel times.

The analytical approach:

Google maps API was used to estimate journey times from each home address to each potential location.

Asking the right question:

- 90 minutes was considered to be a 'reasonable' travel time.
- The distribution of journey times was an output of interest, but the key focus was "What proportion of staff will have a journey time of over 90 minutes?"

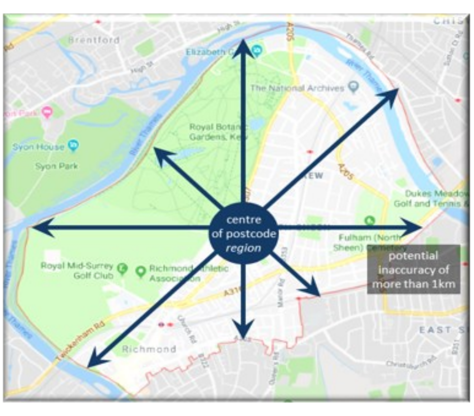
Defining and identifying uncertainty:

Missing Data:

- Only first half of staff postcodes were available.
- Not known who could feasibly travel by car.

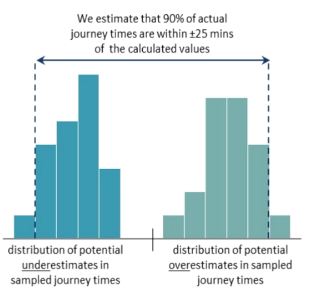
Assumptions:

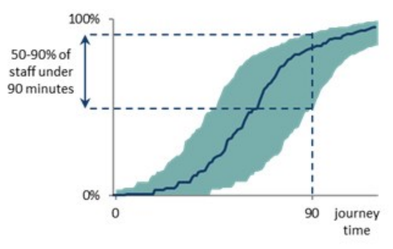
- Staff begin and end work at 9am and 5pm.
- Journey time assumes all transport is running to schedule.
- Assumptions for transfer times, walking speeds etc.

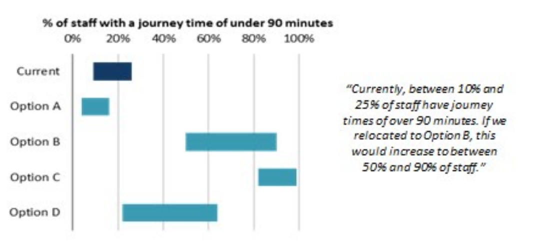


Understanding and measuring uncertainty:

- Through sensitivity analysis on each parameter, the use of partial postcodes was found to be the most significant source of uncertainty and so **Dominant uncertainty** was chosen.
- A **Monte Carlo** approach would have been a more complete analysis but was not viable in the available time.
- A sample of postcode districts were reviewed to calculate: - The journey time from the centre of the postcode; - The shortest possible journey from within that postcode; - The longest possible journey from within that postcode.
- The differences in these times were analysed across the sample to form distributions.
- Producing an uncertainty estimate of 90% of journey times to be within ± 25 minutes of the calculated durations.







Presenting and communicating uncertainty analysis

- A **cumulative distribution** of journey times along with uncertainty bounds clearly presented how many journey times were estimated to be within the crucial threshold of 90 minutes.
- **Bar graph ranges** were presented to the senior move managers, to show how journey times compared between location – the central estimate wasn't plotted to avoid the reader putting too much significance on it.
- Prose sentences were used to clearly present the key results.
- Unquantified uncertainties were communicated e.g. the assumption of no delays.
- The communication of the uncertainty made clear the case for obtaining exact postcodes, allowing the analysis to be repeated with less postcode uncertainty.

Predicting staff pay costs

The situation:

Each department has to manage their pay costs to ensure they stay within their budget, making decisions to recruit (or not) in the face of uncertainty over staff turnover, pay rises, etc.

The analytical approach:

Use a microsimulation to model staff members over time, with departures, promotions, and new recruits modelled stochastically.

Asking the right question:

- Initially 'What will our pay costs be at the end of the year for a given recruit plan?'
- Redefined as 'how likely is it that our costs will exceed our budget?', allowing a risk-based approach to decision-making, and also allowing uncertainty in the budget to be incorporated.

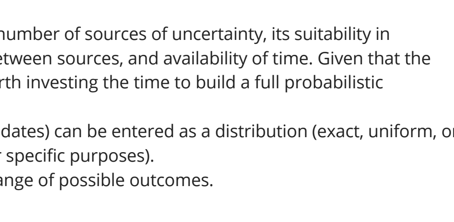
Defining and identifying uncertainty:

Unknowns:

- The pay budget including additional sources of funding not known until the start of the year

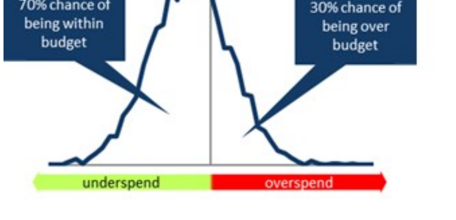
Assumptions:

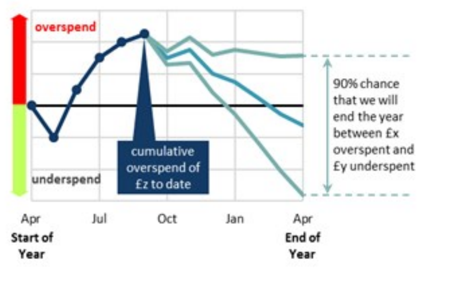
- Number of staff who will join or leave the department in future months
- Variation in starting salaries, pension etc. for new staffs
- Variation in pay awards based on staff performance



Understanding and measuring uncertainty:

- A **Monte Carlo** approach was used due to the large number of sources of uncertainty, its suitability in microsimulation, the (assumed) lack of correlation between sources, and availability of time. Given that the model was likely to be reused many times, it was worth investing the time to build a full probabilistic uncertainty assessment into the model.
- Most numerical assumptions in the model (including dates) can be entered as a distribution (exact, uniform, or triangular, along with some bespoke distributions for specific purposes).
- This resulted in a full probability distribution of the range of possible outcomes.





Presenting and communicating uncertainty analysis

- All major uncertainties were quantified, and presented (with any unquantified sources noted elsewhere).
- The key output was simply the likelihood of remaining within budget, using positive and negative framing (e.g. "We have an 70% chance of staying within budget, and a 30% chance of exceeding our budget")
- As the budget holder was an analyst, we chose to present the full **probability density profile** to illustrate the range of possible over-/under-spends.
- Alongside, simple prose descriptions of the 90% prediction intervals to describe the likely range of the over-/under-spend
- A further graph was used to show how performance against budget was predicted to evolve over the year, using **multiple line graphs** to show the 90% prediction intervals.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

9. Further reading

.

The Aqua Book: guidance on producing quality analysis for government, *HM Treasury, 2015*

Communicating quality, uncertainty and change Guidance for producers of official statistics (pdf), *Government Statistical Service, 2018*

Communicating Risk (pdf), *Parliamentary Office for Science and Technology, 2017*

Uncertainty and graphicacy; How statisticians, journalists, and designers reveal uncertainty in graphics for public consumption (pdf), *Alberto Cairo, 2017*

Useful links:

Click here to see the accessibility statement

Click here to return to home page

10. Decision maker summary

All analysis is uncertain

Analysis is based on a model of the real word. Whether we are looking at statistics about the past, measurements of the present, or forecasts of the future, there will always be a degree of uncertainty in the outputs.

Making decisions using uncertain information can be uncomfortable – decisions would be far easier if we knew for sure what the consequences of each choice would be – but this is an unavoidable position, so analysts aspire to provide honest and practical advice on uncertainty.

Why does it matter?

Taking account of uncertainty – and being seen to do so – is important for public trust. We must not pretend that the consequences of a policy or decision are certain: they are always uncertain to some degree. For each option, a range of outcomes are possible. Implying certainty about one particular outcome can damage public trust when things turn out differently.

Example

Red River Flood, Grand Forks USA, 1997

The National Weather Service (NWS) predicted, 2 months in advance, the Red River to crest 49 feet.

In response, the levees were built to handle a flood of 51 feet

The actual flood level was 54 feet

Had the NWS communicated their uncertainty (+/- 9 feet) the several $billion damages could have been avoided

Relying on best estimates may lead to the wrong decision being made

This uncertainty can be critically important to any decisions based on the analysis, as 'best estimates' are generally not enough to make an informed decision. For example, Option A may appear better than Option B when looking at the most likely outcome of each, but if the uncertainty in A is greater than in B then it might carry an unacceptable chance of much worse outcomes. To help consider the full range of outcomes, it may be useful to work with the analysts to ensure the uncertainty is framed in terms of the decision being made. For example, rather than looking at the central estimate, it may be more useful to consider a higher likelihood, such as the 90% confidence that the figure is within a certain range. Alternatively, to understand the probability of a policy being effective or not, considering the level of take up required for the policy to break even. This is particularly important when levels of uncertainty are very high and there may not be a single best estimate, and instead a decision must be made purely on the expected range.

The presence of uncertainty in analysis may not always inspire confidence, but the absence of uncertainty is even worse. Since uncertainty is ever-present, omitting it from analysis leaves it unacknowledged and unassessed. It could be of any size, leaving the true range of possible outcomes entirely unknown. A proper assessment of analytical uncertainty should reassure you that the analysts have considered the limitations in their data and methodology, as well as the inherent randomness in the world, in order to provide an honest assessment of the range of possible outcomes – rather than presenting misplaced confidence in an impossibly accurate estimate.

Decision makers have a critical role

Decision makers have a critical role working with analysts to agree what the analysis should focus on in relation to the decision being made (for example, a range around an estimate may be less useful than understanding what the percentage take up should be a policy for the policy to be a benefit), helping to identify sources of uncertainty, discussing how the results inform the decision being made.

Key points to remember

- Consider the full range of possible outcomes when using analysis to inform decisions, not just the 'best estimate'.
- Challenge analysts where information on uncertainty is absent or inadequate.
- Ask questions about how to interpret the uncertainty and its implications.
- Provide feedback to analysts on the usefulness and effectiveness of how they communicate uncertainty.

We welcome your feedback

This toolkit has been written for analysts to help them understand and assess the uncertainty in their work, and then to communicate that to the users of their analysis in an effective and helpful way. Feedback from those users – those who commission the work and make decisions informed by it – is crucial to helping refine and improve this guidance. We welcome any comments by emailing: **AnalystsUncertaintyToolkit@homeoffice.gov.uk**.

Useful links:

Click here to see the accessibility statement

Click here to return to home page

Accessibility statement

This website is run by the DfE, MoJ and Home Office. We want as many people as possible to be able to use this website. For example, that means you should be able to:

- change colours, contrast levels and fonts
- zoom in up to 300% without the text spilling off the screen
- navigate most of the website using just a keyboard
- navigate most of the website using speech recognition software
- listen to most of the website using a screen reader (including the most recent versions of JAWS, NVDA and VoiceOver)

We've also made the website text as simple as possible to understand.

AbilityNet has advice on making your device easier to use if you have a disability.

How accessible this website is

We know some parts of this website are not fully accessible:

- the tab names in the browser needs informative title tags
- a focus state is required to increase the constrast when using the tab key
- the website needs a "skip content" tab so that the tab doesn't scroll through everything at the top before going further down the page
- the floating menu on the left hand side in sections 2-5 needs to be keyboard accessible
- the menu bar we requires a visually hidden text that says "menu"
- we need to define language (= English) for each HTML page for the benefit of a screen reader
- all images need amending so that the screen reader doesn't read out the file name
- the content isn't as usable when you enlarge the text
  We were advised by accessibility colleagues that a sitemap was not required in this instance.

What to do if you cannot access parts of this website

If you need information on this website in a different format like accessible PDF, large print, easy read, audio recording or braille:

- email: **AnalystsUncertaintyToolkit@homeoffice.gov.uk**

We'll consider your request and get back to you as soon as is practicably possible.

Reporting accessibility problems with this website

We're always looking to improve the accessibility of this website. If you find any problems not listed on this page or think we're not meeting accessibility requirements, contact us by email on: **AnalystsUncertaintyToolkit@homeoffice.gov.uk**

Enforcement procedure

The Equality and Human Rights Commission (EHRC) is responsible for enforcing the Public Sector Bodies (Websites and Mobile Applications) (No. 2) Accessibility Regulations 2018 (the 'accessibility regulations'). If you're not happy with how we respond to your complaint, **contact the Equality Advisory and Support Service (EASS)**.

Technical information about this website's accessibility

This website is partially compliant with the **Web Content Accessibility Guidelines version 2.1** AA standard, due to the non-compliances listed below.

How we tested this website

This website was last tested in January 2020. The test was carried out by accessibility experts in the Ministry of Justice and changes were made by analysts in the Department for Education.

What we're doing to improve accessibility

We are actively trying to get resource for a CSS/html expert to overcome the issues listed in the "How accessible this website is" paragraph above. This will help to ensure that this site is fully accessible. We will continue to monitor the accessibility of this website on an on-going basis and we plan to fix any accessibility issues reported to us.

Useful links:

Click here to return to home page