The Strategy Unit.

Midlands DSU Network
Decision Support Centre

Midlands DSC
Guide to Evaluation Design, Principles and Practice

This is an interactive PDF. To navigate, use the arrow buttons on each page or locate a specific section using the side menu.

Midlands DSU Network
Decision Support Centre

This guide has been developed by the NHS Midlands Decision Support Centre (DSC), for the NHS Midlands Decision Support Unit (DSU) Network. It is written for people at all levels with an interest in evaluation – whether you are thinking about commissioning an external evaluation or conducting your own evaluation in-house.

The guide takes you through all aspects of designing and conducting evaluation – whatever your resources – and presenting the results.

The DSC and DSU Network was developed to raise the standard of evidence and analysis across health and care systems in the Midlands. Evaluation is an essential part of quality improvement as well as a standalone activity to learn about what works, and why.

The guide begins with an introduction to evaluation and how it should be used, before taking you through the stages required to develop and deliver a high-quality approach, with signposting to further resources.

The guide will be further developed over time to include examples from the DSU Network.

For further information about how the DSC can support your evaluation, including training and development across all of the elements outlined here, or to provide us with comments about the guide and how it could be improved, see our **website**.

What is evaluation?

- What is evaluation?
- How should evaluation be used?

What is evaluation?

Evaluation is a process for learning about what works and how outcomes can be improved.

The UK government guide to evaluation defines it as:

> *"a systematic assessment of the design, implementation and outcomes of an intervention. It involves understanding how an intervention is being, or has been, implemented and what effects it has, for whom and why. It identifies what can be improved and estimates its overall impacts and cost-effectiveness."*
>
> *(Magenta Book, 2020, p.1)*

In practice, it can range from the evaluation of a single service to a complex change programme. It might be formative – providing rapid-cycle learning to improve implementation throughout delivery – or summative – providing learning about the impacts achieved. The best evaluations do both.

This guide provides a process for understanding the type of evaluation that you need and designing an appropriate approach – one that is proportionate to the intervention (or programme) and the available resources.

It provides a set of key considerations and principles – with links to useful resources – rather than a single prescriptive approach. This is because an effective evaluation is one that is specifically designed for your needs; and there are lots of practical tools and resources to help you.

> *"While evaluation comes in many shapes and sizes, its key purpose is to help us to develop deeper understanding of how best to improve health care."*
>
> *(The Health Foundation, 2015, p.5)*

How should evaluation be used?

Evaluation is used for learning across the planning and delivery cycle, for example:

- When we are planning a new intervention or programme: to consider the current evidence base and design an intervention that we can evaluate
- When we are innovating and trying something new: to test and learn, and evidence effectiveness when implemented at scale
- When we want to know what lies behind the (strong or weak) performance of a service: and how it can be improved
- When we are piloting one or more approaches.

Evaluation is most effective when it is embedded in an intervention (or programme) from the start. In this way, we can ensure that we have the right information we need to make judgements at the end, as well as learn and reflect during implementation to adapt our service as necessary to deliver the outcomes that we set out to achieve.

How do I develop my evaluation design?

- Understanding your theory of change
- How to develop a logic model
- Identifying your aims, objectives and research questions
- Identifying the data you need to collect
- Identifying the purpose of your evaluation
- Identifying the timescales for data collection and analysis

There is no one size fits all approach to evaluation. But to follow best practice, the starting point is to understand your theory of change.

Understanding your theory of change

All interventions and programmes have different elements. And **a theory that if we do X, we will achieve Y**. Sometimes this is taken from our previous experience, or from what we have learnt from elsewhere. Sometimes it is based on clinician, patient or other expert views about what needs to happen, or change from our existing practice. Wherever it comes from, all programmes and interventions have a theory of change. When we understand this, we can develop an evaluation design to test whether we achieve what we set out to – and what are the key facilitators and barriers behind what does and doesn't work.

There are different approaches to 'theory-based evaluation' but all are based on this principle. Government and NHSEI guidance recommend the use of a logic model to capture the theory of an intervention or programme as the basis for an evaluation framework (and they are recommended by a wide range of organisations including the US Center for Disease Control and Prevention).

How to develop a logic model

A logic model is a diagram that provides a one page summary of the programme or intervention. As well as informing the evaluation, they are also useful for project planning. For instance, the process of developing the model can identify elements that have not been thought through or where there are differences of opinion that can then be resolved. **A logic model sets out the:**

Context: This is important because the organisational, policy and practice context can affect whether we achieve what we set out to. Changes in an organisational structure, new policy requirements, or new services being commissioned are all examples of things that can change and influence what our initial theory was based upon.

Rationale for change: This is at the heart of our theory. What is the problem that we are seeking to address, and why?

Inputs:	Activities:	Outputs:	Outcomes:	Impacts:
These are the resources that will be necessary. This almost always includes money, but other resources are also usually required such as in-kind contributions from partners, physical space, kit, or (parts of) FTE posts.	These are the things that we are going to do to deliver the programme. They are usually grouped into different themes or strands – for example there may be activities in primary care, for workforce development and patient engagement.	That our activities will deliver. Outputs are usually things that we can count. What we will notice changing? How many people will be involved?	Are the things that we are aiming to improve. They are what we expect to be achieved by the programme.	There is always some work to do to separate outcomes and impacts. Impacts are best understood as the wider, longer-term changes that we expect our outcomes to contribute to. Outcomes are directly attributable to what we will deliver. Impacts are wider, at a system or societal level.

Assumptions: All theories are based on assumptions – examples include the contribution of partners, availability of funding, recruitment of related posts, or patient or clinician take-up. They are often related to the context. Recording the assumptions means that we can test them in the evaluation; and take account of things that are important if they do not happen.

It is unlikely that you can fit all of the outputs and outcomes on the one page logic model if this is the case, use an additional page. For instance, 'Reduced health inequalities' as a high-level outcome, which is reflects a range of more specific outcomes linked to particular target communities.

Keep the logic model under review to ensure that it reflects the programme/intervention as it develops. It is usual for delivery to be amended over time as we learn about what works. This is important because the evaluation needs to focus on what is delivered, not on what we thought we would deliver at the outset. Capturing why delivery changes is important for learning.

You should also consider whether you need to develop logic models for each of your themed groups of activities, outputs and outcomes. This depends on how complex your programme is (an intervention can usually be captured in a single model). These **'nested' logic models** provide a further level of detail about the programme strands and can help focus the evaluation.

Finally, because the model is only a summary diagram, **a narrative should be written** to explain the theory in more detail. Use the headings of the model's elements (context, activities, etc.) to structure the narrative. It can also provide more space to list the different outputs and outcomes, where there isn't space to include them all in the diagram.

- **Here is a detailed guide on logic models in evaluation.**
- **Here is a 'how to' guide on producing a logic model.**

Identifying your aims, objectives and research questions

The **aim** of your evaluation is the overarching question that you are seeking to answer. The **objectives** are more specific statements of what you need to find out to answer the **aim**. When we are evaluating a programme or a complex intervention there will usually be multiple aims, each with their own **objectives**. These are closely related to the programme or intervention's theory of change.

The evaluation will seek to test if the programme activities deliver the outcomes identified. It is unlikely that the evaluation can focus on all aspects of the programme or intervention.

A logic model helps us to refine the aims and objectives by helping us to focus on the key aspects of the programme or intervention that we need to test. The logic model and theory of change make explicit: **What is the programme aiming to achieve? What needs to happen to deliver these aims?** By focusing on the key activities – the key mechanisms to deliver outcomes – the logic model informs the evaluation design.

For example, one aim may be to improve patient outcomes. The objectives may be: to improve their experience of services; to provide a more timely care pathway; and to improve clinical outcomes.

When you have developed your aims and objectives, you can develop your **research questions**. These are the lines of enquiry that you are going to pursue and that you will collect data to answer, to establish if the programme or intervention is working as intended.

To continue the example, if the programme objective is to improve patient experience your questions would include: What are patient's views of the new service? Do different groups of patients have different experiences, and why? Are there any ways in which the service could be improved?

Because your logic model identifies the theory of change, your questions should be structured to test that theory – are the activities being delivered as intended? Are they bringing about the expected changes, and any unexpected ones?

Your questions will enable you to collect data that provides learning about what works, for whom, and why.

Identifying the data you need to collect

Now you have identified the questions you need to answer, you can focus on the data that you need to collect.

The process of developing your logic model will have identified the outputs and outcomes that you need to measure. Usually, outputs are things that can be counted – things delivered, people trained, patients seen. Outcomes are things that will change and can require more time and attention to define.

Routine or administrative data

Many of the things you need to measure will be readily available within existing system records, some will be easily collected through adaptations to these systems or simple structures established to monitor the delivery of the programme or intervention – administrative data that is routinely collected. This is commonly quantitative data and is often referred to as monitoring information, or MI.

We discuss ways of identifying the most appropriate measures **here**.

New data sources

Outcomes are more likely to require new data sources. Some will be demonstrated by existing data, such as changes in the types of patients seen, in the clinical outcomes for target groups or increased efficiency compared to business as usual. But some will be related to experiences, or patients and staff and thus more qualitative.

We discuss ways of identifying the most appropriate data collection methods **here** and **here**.

Identifying the purpose of your evaluation

There are two broad functions that an evaluation can fulfil – formative evaluation and summative evaluation, or a combination of both. Deciding the purpose of your evaluation is an important step in developing your final design.

Summative evaluation

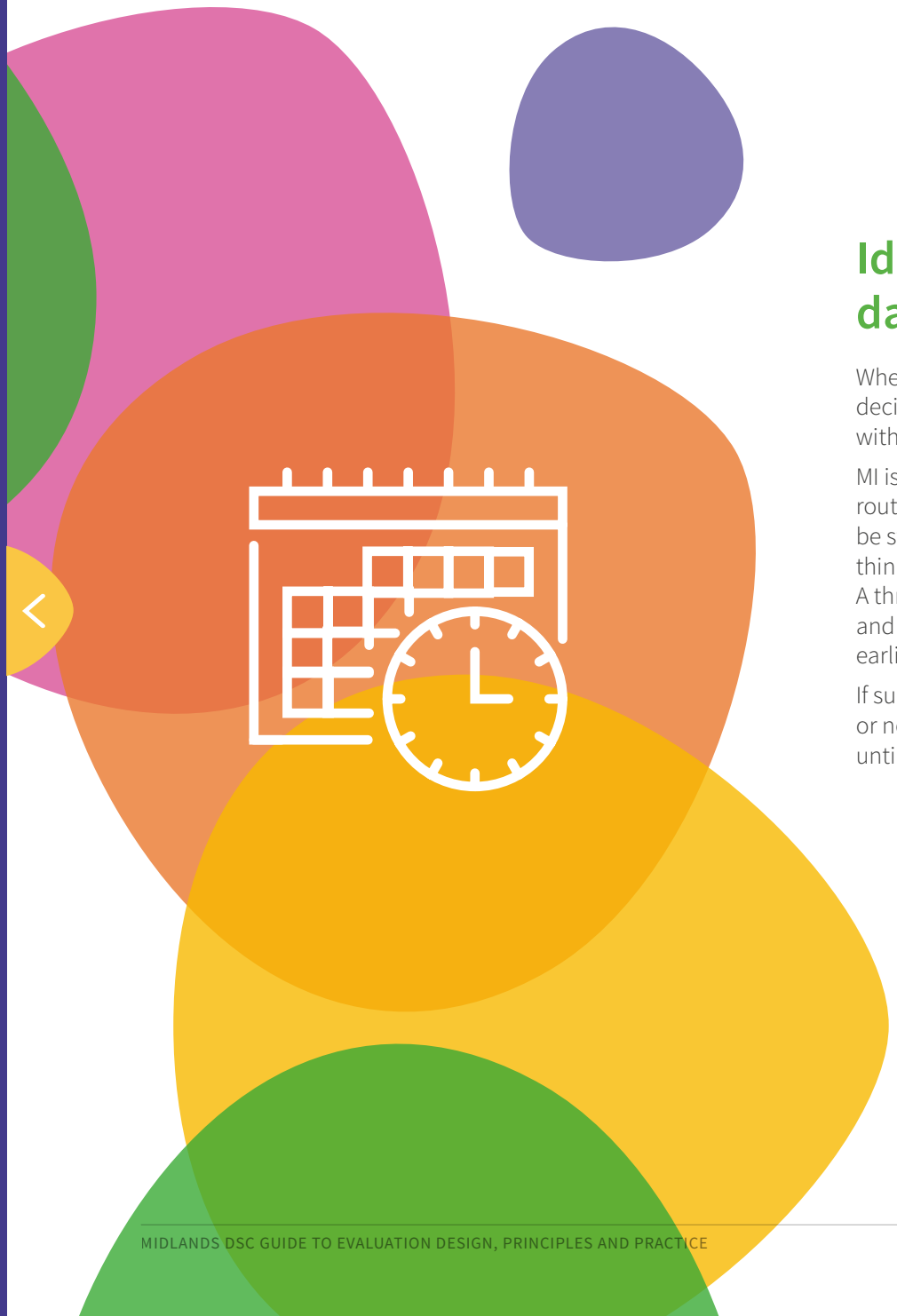
A summative evaluation provides learning at or towards the end of the programme or intervention and commonly focuses on outcomes. Sometimes referred to as impact evaluation, it is used to make investment decisions – is the programme or intervention a success? Should it be continued/rolled out? Is it value for money? Summative evaluations have a strong quantitative element, exploring the outcomes delivered with qualitative work to explore the narrative around the delivery that took place.

Formative evaluation

A formative evaluation provides learning on an ongoing basis, to inform the development of the programme or intervention. It is used to shape delivery – is the programme or intervention working as intended, and why or why not? What needs to change? Formative evaluation usually include a strong qualitative element to explore the processes behind the outputs being delivered (MI) and the outcomes that are emerging or that are expected from the outputs. Because outcomes take time to deliver, quantitative analysis is often limited to MI.

Formative and summative evaluation

The best evaluations fulfil both functions. They provide formative learning during delivery, and summative conclusions at the end or in the final stages. Evaluations of this type provide the strongest understanding of what did and didn't work, rather than relying on the partial accounts that can emerge in a summative design (where, for instance, key people involved in the early stages have moved on; or the detail of what happened becomes disputed or reconstructed by key stakeholders).



Identifying the timescales for data collection and analysis

When you have agreed the purpose of your evaluation, you can decide on when you are going to collect and analyse your data within the overall timescale for the programme or intervention.

MI is often analysed on a monthly or quarterly basis as part of routine project management. Other data collection will need to be structured to meet the evaluation purpose. If it is formative, think about when it would be a sensible time to review progress. A three-year programme might benefit from learning at 6 months and then each year. A short six-month pilot would require much earlier review.

If summative, when will a decision be required about whether or not to continue beyond the initially funded period – if we wait until the end of the three years of delivery, will that be too late?

Who should be involved in the evaluation?

- Who should contribute to the evaluation design?
- Who will need to provide information for the evaluation?
- Consider co-production and participatory approaches to engage stakeholders effectively

So far in this guide we have set out the steps to take in developing your evaluation design. But this is not a sole endeavour. The best evaluations are developed with key stakeholders in the programme or intervention. Working with stakeholders ensures that your evaluation meets their needs, and also secures buy-in – stakeholders are going to need to provide time and resources. Here we talk about the two ways you will need to involve stakeholders – in the design, and in your data collection.

Who should contribute to the evaluation design?

So far in this guide we have talked about the things you should do, without making any reference to who the 'you' is. Whilst there should always be a lead for the evaluation – as with any project, and we return to project teams **here** and governance **here** – a range of perspectives should be included in developing the evaluation design.

A key stage is in the development of your logic model, as discussed **here**. To ensure your theory of change reflects the perspectives of different participants in the programme or intervention; and, importantly to help identify and resolve where there are differences.

Working with stakeholders to develop the design can range from **co-production**, to more straightforward consultation. Co-production takes time and commitment, and resources. It might not always be practical. Whatever approach is taken, the important principle is that there is meaningful involvement of stakeholders – that the right people are consulted on their ideas, and that those ideas are listened to and acted upon.

Think about:

- Who has been involved in designing the programme or intervention, even where they have now moved on or will not be involved in delivery?
- Who is responsible for managing the programme or intervention, both at operational and strategic levels?
- Who has been involved in commissioning the programme or intervention?
- Who are the wider strategic stakeholders?
- Who are the partners in delivering the programme?
- Who are the intended beneficiaries of the programme?

All of these groups will have perspectives on what the programme is setting out to achieve and what the focus of the evaluation should be. Some will also be important gatekeepers – providing access to data and people during the evaluation (**see next section**). And finally, these stakeholders will be audiences for the evaluation; some will be absolutely key – they will need to take actions on the basis of your conclusions. So asking them about what they want from the evaluation will ensure you deliver what they need.

Not everyone needs to be involved in every stage. But think about who from these broad groupings should be involved in the logic model development.

As you develop your set of outcomes, aims and objectives, think about who from these groups can be consulted to confirm, refine and develop them to a final set.

More information on engaging stakeholders can be found on the Better Evaluation **website**.

Who will need to provide information for the evaluation?

There are two main groups of people who you will need to provide you with information – with data – for the evaluation.

People who provide access to data

You will be drawing on monitoring and other routine or administrative data. Before you begin, you need to identify who these people are and agree with them both that this data can be provided, and how it will be provided. There is more detail on this **here**.

There are also people who are gatekeepers to groups of people, for instance patients, that you will need to involve. Again, engaging these people during the design stage will help to ensure your evaluation runs smoothly and you don't hit snags when things that were assumed don't happen.

People who will provide data – your participants

There will be a wide range of people who will provide their perspectives on and experiences of the programme or intervention. The list of potential contributors of this qualitative data is the same as the list above of people to consider involving in the design of the evaluation.

Think about:

- Who has been involved in designing the programme or intervention, even where they have now moved on early in delivery?
- Who is responsible for managing the programme or intervention, both at operational and strategic levels?
- Who has been involved in commissioning the programme or intervention?
- Who are the wider strategic stakeholders?
- Who are the partners in delivering the programme?
- Who are the intended beneficiaries of the programme?

Different groups of people will provide data for the different outcomes of your theory of change, and will provide important contributions to your understanding of process – the narrative behind the outputs and outcomes that are being delivered.

Consider co-production and participatory approaches to engage stakeholders effectively

Co-production

Participation

Co-production and other meaningful involvement of stakeholders can take many different forms – it includes citizen involvement, participation, engagement and consultation.

'Co-production is a way of working that involves people who use health and care services, carers and communities in equal partnership; and which engages groups of people at the earliest stages of service design, development and evaluation. Co-production acknowledges that people with 'lived experience' of a particular condition are often best placed to advise on what support and services will make a positive difference to their lives. Done well, co-production helps to ground discussions in reality, and to maintain a person-centred perspective.'

Coalition for Collaborative Care, 2020

Co-production is not always required, practical or achievable. Meaningful involvement by the right people at the right time is the important guiding principle. Be careful not to make claims about coproduction if it is not possible – it takes time and resources.

Co-production is most effective when it is part of the way that work happens rather than as an add on. The Coalition for Collaborative Care, and NHSE, advocate a model of five values and seven steps to realise this ambition. The five values are centred around ownership and understanding, a culture of openness and honesty, a commitment to power sharing, clear communication and a culture of respect. The seven steps use techniques such as getting agreement from senior leaders to champion co-production, using open and fair approaches to recruit a range of people and having systems that recognise their contributions. Identifying areas where co-production can have a genuine impact was also cited as a key step as was ensuring citizens are involved early on and ensuring they receive training to understand what co-production looks like and then reviewing progress.

This **guide** for health and social care from SCIE provides a wide range of resources.

Consider co-production and participatory approaches to engage stakeholders effectively

Co-production

Participation

Participation should benefit all those involved – contributing to the design of solutions that improve outcomes, raise awareness of services, and provide self-development opportunities.

NHSE's **Framework for patient and public participation in public health commissioning** provides guidelines for how patients and the public are involved in the commissioning of public health services. 'Patient and public participation' is defined as service users, patients, children and young people, carers and families and those with lived experience as well as the wider public and stakeholder organisations representing these networks and communities.

In addition, NHSE's guide to **Diverse and Inclusive participation** offer a range of practical steps:

- Build on what exists already, rather than set up new mechanisms unless local stakeholders require it
- Using existing links to health, social care and health improvement services utilises the relationships these bodies may already have with communities
- Build a shared understanding of what participation and involvement looks like with staff and stakeholders
- Explore a range of ways for people to be involved, both on and offline
- Take into account different needs – cultural, linguistic, religious, communication and accessibility.

How do I design the evaluation?

- Types of evaluation design
- Evaluating digital health

When you have worked with your stakeholders to develop your theory of change logic model, aims, objectives and research questions, you can begin to design your methods for data collection and analysis.

When designing the evaluation, it can be helpful to consider the existing evidence. For example, when others have attempted to deliver a similar change, what outcomes did they achieve? Or, did they identify specific aspects of their service redesign that prevented successful implementation? Support with evidence review and access to a library of existing evidence, including from across the DSU Network, is provided by the **DSC Evidence and Knowledge** service.

Types of evaluation design

Different evaluation designs are used to answer different types of research question.

Process evaluations | **Impact evaluations** | **Mixed methods**

This type of evaluation answers questions about whether the programme or intervention is being delivered as intended, and what is working and why. A process evaluation uses primarily qualitative methods, although some quantitative data collection might be included, for example to provide a descriptive understanding of service activity. Process evaluations are usually formative, in that they provide learning during delivery and before impacts are established.

The main qualitative data collection methods and their purpose are explored in detail **here**:

- Surveys: A qualitative survey looks for the characteristics of diversity in the sample rather than counting the variation or distribution.
- Document Analysis: Relies on data that has already been collected and reported. Documents could include: business cases, programme reporting, case studies, website information.
- Interviews allow for the exploration of individual perspectives for the change. Findings can be used to improve, stop or roll-out the delivery of the change.
- Focus groups are essentially group interviews. In a focus group setting, an evaluator takes on more of a facilitator role rather than interviewer.
- Observations: Here the evaluator plays a passive role, systematically observing and then describing events, behaviours, and artefacts within the social or natural setting.

When you have worked with your stakeholders to develop your theory of change logic model, aims, objectives and research questions, you can begin to design your methods for data collection and analysis.

When designing the evaluation, it can be helpful to consider the existing evidence. For example, when others have attempted to deliver a similar change, what outcomes did they achieve? Or, did they identify specific aspects of their service redesign that prevented successful implementation? Support with evidence review and access to a library of existing evidence, including from across the DSU Network, is provided by the **DSC Evidence and Knowledge** service.

Types of evaluation design

Different evaluation designs are used to answer different types of research question.

Process evaluations | **Impact evaluations** | **Mixed methods**

This type of evaluation identifies the results or effects of a programme/intervention, focusing on the aspects that are usually found on the right of a logic model. It measures the outcomes and longer-term impact from implementing the intervention, including on the beneficiaries' or target group (for example patients, or the workforce) changes in knowledge, experience and behaviours.

- An impact evaluation typically makes use of both quantitative methodologies, where the quantitative data analyses can be advanced, for example taking a quasi-experimental design to attribute impact. More detail on quantitative methods is provided **here**.

When you have worked with your stakeholders to develop your theory of change logic model, aims, objectives and research questions, you can begin to design your methods for data collection and analysis.

When designing the evaluation, it can be helpful to consider the existing evidence. For example, when others have attempted to deliver a similar change, what outcomes did they achieve? Or, did they identify specific aspects of their service redesign that prevented successful implementation? Support with evidence review and access to a library of existing evidence, including from across the DSU Network, is provided by the **DSC Evidence and Knowledge** service.

Types of evaluation design

Different evaluation designs are used to answer different types of research question.

Process evaluations	Impact evaluations	Mixed methods

This is our recommended approach for the highest quality evaluation, blending both formative and summative approaches. It addresses both the process and impact aspects of the evaluation and draws on both qualitative and quantitative methodologies to provide both evidence of impact and narrative understanding. Findings from each methodology are combined or '**triangulated**'.

Mixed method evaluations have a dual purpose:

- Clarification: broadening and deepening the understanding of the processes through which the programme achieves its outcomes
- Corroboration: strengthening the reliability of data, and the validity of the findings and recommendations.

Evaluating digital health

As digital products and their use in healthcare practice becomes more commonplace there are some unique features to consider when evaluating the products and their implementation.

In addition to our general evaluation approach, the following are additional useful resources:

- **Public Health England** has provided step by step guidance for evaluating digital health products. They are intended for anyone developing or running a digital health product, including a service or campaign which is provided through an app or website.
- The **NASSS Framework** has been developed from the synthesis of evidence as to why technological/digital interventions face **Non-adoption, Abandonment, Scale-Up, Spread**, and Sustainability in healthcare settings. It acknowledges the complex systems within which technology programmes are implemented and guides consideration of a combination of factors: the condition or illness that the technology supports, the technology itself, the supply and demand side value, the behaviours required by users/adopters, the organisation(s) involved, the wider system context and the potential for embedding and adapting over time.

What governance arrangements does my evaluation need?

- Good project management
- Good risk management

Establishing good governance for the evaluation ensures that it is well managed and able to respond to any changes in the context for, or delivery of, the programme or intervention. It ensures that the evaluation keeps to the timetable and can identify and respond to any risks.

There are two elements to good evaluation governance. The first is the structure for the direct management of the evaluation. The second is the strategic structure around it, with a steering group of stakeholders who can support the evaluation activities and the response to any unexpected changes or emergent risks.

Good project management

A simple structure provides clear roles and responsibilities for the management of the project. The table below provides a summary of what good governance looks like.

Role	Skills and Characteristics	Responsibilities
Project Director	• Senior expertise in evaluation and/or the topic	• Overall responsibility for the delivery of the evaluation to time and budget • Quality assurance at each stage of the evaluation
Project Manager	• Good organisational skills • Evaluation knowledge and experience	• Day-to-day management of the evaluation • Identifying risks • Liaising with the Project Director
Project Steering Group	• Strategic stakeholders – senior roles in participating organisations • Commissioner or sponsor of project • Members of the programme/intervention delivery team • Representatives of patient or target beneficiaries	• Shape and steer the evaluation as it progresses • Share and reflect on emerging findings • Quality assure research tools and outputs • Help unblock barriers/facilitate progress/ address risks

The Magenta Book (the government guide to evaluation) provides more detail on governance and is an accessible style, in Section 5.

Good risk management

As with good project management, good evaluation risk management shares common features with risk management in other contexts.

Good risk management begins at the outset of the evaluation, when potential risks are identified and planned around. For example, the risk that data won't be provided on time or that a particular group of patients will be difficult to access – in the planning stage, you should consider the different strategies that you will take to ensure these activities take place according to your timetable. Risk management identifies theses risks and the steps that will be taken to address them; and progress towards them is monitored. The Steering Group can be a useful place to discuss alternative strategies, such as how to work with gatekeepers. Where delays are inevitable, they can also agree that final outputs will be delivered later than originally expected.

Risk management is not, therefore, static. New risks can emerge at any time, which were not foreseen. For example, a provider's reorganisation may introduce new pathways or new internal structures meaning that roles change and a gatekeeper committed to the success of the project moves on.

The Project Manager must be alert to these changes as risks to delivery and work in the first instance with the Project Director to identify the potential or evident consequences to the evaluation and strategies to address them. These should be shared with the Project Steering Group who, as above, should be able to help identify the strongest response as well as agree to any delays to or impacts on the final outputs.

An example risk table can be found **here**.

What ethical issues do I need to think about?

- GDPR and data security
- Consent
- Taking an inclusive approach
- Safeguarding
- Ethics in analysis and reporting
- Ensuring the safety of researchers

Any form of research involving participants as a source of primary data requires ethical considerations to protect and promote the dignity, rights, safety and wellbeing of participants whether it be the public, patients or staff. This is best practice and a legal obligation.

Evaluations do not require ethical approval. But you have a duty to conduct your evaluation ethically and there may be local research leads that you need to negotiate with to access patients or other stakeholder who will require evidence of your approach to ethics.

The NHS Health Research Authority's (HRA) **UK Policy Framework for Health and Social Care Research** describes the principles and responsibilities to apply when conducting health and social care research. It sets out when formal ethical approvals are required by Research Ethics Committees. The **HRA decision tool** can help you decide if ethical approval is required.

If research for the evaluation is taking place on an NHS site it may be advisable to consult local R&D offices to understand their criteria and requirements of approvals.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

GDPR and data security | Consent | Taking an inclusive approach | Safeguarding | Ethics in analysis and reporting | Ensuring the safety of researchers

GDPR and data security

All activities of handling personal data should comply with the General Data Protection Regulation (GDPR) in the UK. The **7 GDPR principles** describe how to handle data to meet the statutory obligations. The **7 Caldicott principles** further sets out how to handle patient identifiable data. It is vital to state clearly in the consent process in what situations confidentiality may need to be broken. This may become necessary when there are patient safety and/or safeguarding risks identified.

There is more detail on information governance **here**.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

Consent

Informed consent enables participants to make an informed decision about their participation. In this section we are referring to qualitative research but the same principles apply – for any quantitative data you are accessing you must ensure that it is either non-identifiable or that consent has been secured for use in evaluation.

Information provided to participants in advance should include: the purpose of the evaluation, the organisations involved in conducting it, the source of funding, how findings will be used, any potential adverse impacts of their participation, the nature and duration of involvement, how data will be used, stored and who will have access and the measures that will be taken to safeguard anonymity, confidentiality and privacy of participants. It should also include contact details for both the evaluation team (usually the project manager) and for the commissioning or host organisation. It should be clear that participants are able to contact either of these people to find out more, or to raise any concerns that they have.

Participants should be given sufficient time to consider the information and raise any questions/concerns about involvement and be free from coercion. Participants who are patients or service users must know that refusal to take part will not affect the services they receive in any way.

There is an example participant information sheet **here** and an example consent form **here**.

Consent should be understood as an ongoing process and participants should understand that they have the right to refuse or withdraw participation at any time without explanation.

The standards for consent involve an active, positive opt-in process requiring explicit consent to clear and specific statements. The conditions of valid consent that satisfy GDPR principles can be found **here**.

The HRA also provide **guidance for attaining consent** for: adults not able to consent for themselves; children and young people; emergency research; and, deceased people. Additional advice concerning those lacking mental capacity can be found **here**.

The HRA have also set out considerations for **electronic methods of attaining consent**.

Gaining informed consent to use information that is found in the public domain varies and, in some instances, this can be difficult to establish. For example, using information from social media requires attaining appropriate permissions from the content creator. If you decide to use online data collection methods that draw on pre-existing material people have uploaded (as opposed to material you have specifically asked them to generate following a consent process), you will need to carefully consider the ethical issues. This **Association of Internet Researchers' document** discusses these issues.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

GDPR and data security | Consent | **Taking an inclusive approach** | Safeguarding | Ethics in analysis and reporting | Ensuring the safety of researchers

Taking an inclusive approach

We explored ways of engaging stakeholders and ensuring meaningful participation earlier in the guide. These principles apply to your research methods as well. This begins by considering any potential impacts on all participants; any barriers to participation; and any resources required to address these barriers.

It is vital to ensure that all information is accessible by the target group, this means using plain everyday language and keeping processes simple and clear. Check understanding as part of your consent process – not only that the information has been received but that it has been understood.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

GDPR and data security | Consent | Taking an inclusive approach | **Safeguarding** | Ethics in analysis and reporting | Ensuring the safety of researchers

Safeguarding

The limit on confidentiality for research participants is where a safeguarding concern emerges. Researchers should not try to address the issue. They must make a note of the concern and report it through the appropriate channels; and must tell the participant that this is what they are going to do and why. These instances are very rare, but prior to data collection beginning you must establish who the key contact is for reporting a safeguarding concern in any setting or system you are working in.

Where you are working with patients, service users or members of the community it is a good idea to prepare a list of local services that can provide support should any broader – non-safeguarding – issues emerge. For example, the researcher may be asked where help can be provided for a particular condition. As a researcher in a position of power you have an ethical duty to contact the participant, with their agreement, after the interview or focus group (or other discussion) with the information they have requested.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

GDPR and data security | Consent | Taking an inclusive approach | Safeguarding | **Ethics in analysis and reporting** | Ensuring the safety of researchers

Ethics in analysis and reporting

Wherever possible, you should share your findings with participants, even if it is a short summary. It is important that participants can see what you have done with the information (data) and time they have provided. This fosters a sense of accountability, cultivates candour and demonstrates care. In a participatory approach, your research participants and stakeholders can contribute to data analysis – for instance sharing emerging findings and checking interpretation.

Evaluations should seek to meaningfully engage stakeholders to understand their perspectives. The design and management of evaluations should aim to remove bias and maximize objectivity by not reflecting personal or sectoral interests, whilst fostering professional integrity and respecting the rights of institutions and individuals. Information should be handled in confidence and be sensitive to the beliefs and customs of the local social and cultural environments.

You can find more resources on ethics and quality standards **here**.

How to address ethics in your evaluation design

It is vital all evaluations are conducted according to high ethical standards regardless of whether formal ethical approvals are required or not. The core ethical considerations for evaluations are discussed below.

GDPR and data security | Consent | Taking an inclusive approach | Safeguarding | Ethics in analysis and reporting | Ensuring the safety of researchers

Ensuring the safety of researchers

As well as your participants you must also consider your or your researchers' safety. This is particularly important when it comes to conducting interviews or observations in private settings. The Social Research Association's Safety Guidelines presents a useful code of practice. These can inform risk assessments and minimise potential dangers.

As well as physical safety risks, emotional safety precautions should also be put in place to support researchers. The nature of interviews can be sensitive and require an emotional labour from the interviewer which needs to be supported. Debriefs/ post interview reflections following such work with appropriate supervision is highly recommended.

What resources will my evaluation need?

- Consider the team and skills that will be required
- How to think about the resources that will be required

When you are arriving at your final evaluation design, you need to consider the resources that will be required and the resources that you have available. There will be interplay between what would like to do, and what you're able to do.

Consider the team and skills that will be required

Earlier, we outlined the **key roles for good project governance**.

- Think carefully about who has the skills and experience to be the Project Director, and Project Manager? Do they have the time available that the roles will require – and what needs to happen to keep that availability over the length of the evaluation (for instance, agreements with senior managers)?
- Who has the technical skills and experience to deliver the quantitative and qualitative aspects of the evaluation – data collection, analysis and reporting?
- Who is able to present the findings in an accessible way?

There may be support and development needs, which the DSC can provide through our **training programme** and **evaluation support**.

How to think about the resources that will be required

There is no standard way to identify the costs of the different elements of your evaluation, but there are some useful considerations that can guide you. Discuss your ideas with your **stakeholders as you plan your evaluation**, as they will be able to help think about how you can access particular groups of participants. Remember that as well as the time to conduct the data collection and analysis, time is needed to plan the work in detail and agree research tools such as interview guides with the steering group or other key stakeholder.

For quantitative work, the size of the data set and types of analysis required will be key. Discuss the time that will be required to clean, validate and process the data with the analysts who will be undertaking the work. It may be that you need to amend your design if the work is going to take longer than you expected, or require more person hours than are available.

This is explored in more detail **here**.

- For qualitative work, there are some more general considerations:
  - > You can assume it is possible to conduct around six face-to-face or telephone interviews in one day, or the equivalent of one day where they are spread out. With telephone interviews this is easier to plan for, because there is no travel time. With face-to-face interviews, as well as travel to a venue there can be travel between venues; or you may need to depend on a provider or partner arranging interviews for you and it is unlikely (but not impossible!) that it can be arranged for six patients, for example, to all be interviewed in one place in a single day.
  - > A focus group will need two people for facilitation – whether in person or online. As with interviews there are travel times to consider. A group will usually last one to one and half hours, and its unlikely that you'll be able to do more than two a day in person and three online.
  - > Some qualitative work may be with vulnerable people and about sensitive topics. These are likely to take longer than those that are more straightforward.
  - > Some qualitative work will take longer to arrange, as with the example of patient interviews that require the support of a provider or partner.
  - > Remember to think about the time it will take to write-up the notes from your qualitative work. It will take at least an hour and probably an hour and half to write up the notes from an hour long interview.
  - > And there is the time for analysis, with different approaches taking different time (depending on their rigour).
- Think about the resources any partners or stakeholders might need to commit, and whether they are able to. This includes providing access to people and places, but also things like attending the evaluation steering group. Make sure your partners are able to commit these resources before you arrive at your final design.

How do I carry out qualitative research?

- Identifying your sample
- Recruiting your sample
- Using case studies for an in-depth understanding
- Action research
- Working with patients and vulnerable groups
- Qualitative research methods
- How to do qualitative work during COVID-19 restrictions
- Creative approaches
- How do I analyse my qualitative data?
- Software and practical tools

In this section we set out the key considerations for you when developing the qualitative elements of your evaluation. Starting with thinking about the sample of participants, and then through the key methods.

Identifying your sample

During the development of your logic model, research aims and objectives and work with stakeholders you will have **identified the groups that you need to engage in data collection**. Now you need to think about how you can develop a sample of those groups to invite to participate in the evaluation.

In qualitative studies, small sample sizes can yield large amounts of data. For example, a study which explores staff views of a service improvement initiative in one hospital department may only need to 10-12 interviews or 2-3 focus groups with staff working within the service. Qualitative samples are not statistical samples as in quantitative methods. For example, if you are exploring the experiences of a minority group who are rarely engaged in research, it may only be possible to engage a small number of participants but what they tell you will be extremely valuable because these are voices seldom heard.

A qualitative sample of 30-50 is considered sizeable and would be required where the intervention being evaluated affects more people. For example, a service improvement that is introduced hospital-wide and involves different services/departments. This number of interviews may also be necessary if the service improvement is system-wide or introduced across different hospitals in a region.

A qualitative sample size that is larger, say 100+, is indicative of a complex evaluation often conducted over a longer time period, where there may be repeated interviews with the same sample to see how their experience changes over time.

There are some additional practical considerations that can influence sample size:

- **Population heterogeneity:** a diverse population necessitates an increased sample size
- **Selection criteria:** the more criteria that are considered important in a sample, the larger the sample will have to be
- **Groups of special interest:** if the study population has groups that require in-depth study, it will require a larger overall sample
- **Budget and resource:** the scale of the budget or evaluation capacity will place limits on sample size
- **Saturation:** it is good practice to sample until 'saturation' is reached; this is when later participants confirm previous themes/findings as earlier ones and no new or additional insight emerges.

Recruiting your sample

A number of routes can be used to recruit your sample of participants.

- **Purposive sampling:** the sample has features or characteristics of interest, e.g. diabetes patients who are all trialling a self-care app to be evaluated.
- **Convenience sampling:** the sample is chosen according to ease of access, e.g. choosing to sample from Site X rather than Site Y for the same intervention because stakeholders at Site X are more engaged
- **Opportunistic sampling:** opportunities for sampling that arise during the course of the data collection, e.g. an evaluation of ways of working within the voluntary sector might identify community groups that were not previously identified.
- **Snowball sampling:** asking people who have already participated to identify other people who also have experience of the same intervention
- **Maximum variation sampling:** seeking participation from as diverse a range of participants as possible, e.g. in evaluating population-based interventions, ensuring representation from as many different groups as possible.

Using case studies for an in-depth understanding

Case studies are often used in qualitative (and mixed methods) studies to provide a holistic and in-depth understanding of a phenomena or intervention, sometimes over time. A 'case' could be an individual, group, organisation, geographical area or event.

Data is gathered using a variety of methods (both qualitative and quantitative). The use of mixed/multi methods in case study evaluation provides an advantage in allowing more meaningful and robust measurements, analysis and interpretation. However, there are also disadvantages of cost and inter-dependency, created through multiple different data collection and analyses which could impact on evaluation timelines.

Case studies offer a more comprehensive and in-depth understanding (known as 'thick description') of the entity being evaluated, the context in which it is situated, and the characteristics of the participants (individuals, services, organisations). It is therefore ideal for researching broad questions and complex social phenomena. Whilst a primary purpose of case studies is to understand the unique aspect of the case(s), good contextual understanding can enable transferability of learning to other 'cases'. The case study can also provide a detailed illustration of the overall evaluation findings – a real world example of the experiences, for instance, that have emerged as a key theme from participants.

Action research

Action research is also known as Participatory Action Research. This approach co-opts participants into the study in order to directly bring about learning and change for their specific contexts. In healthcare settings, this form of self-reflective enquiry can help healthcare practitioners improve their own practices, their working environments and the experiences of colleagues and patients.

The rationale is that healthcare professionals are continually informally evaluating and changing the way they deliver care, and through an action research approach they develop and apply a set of formal and structured skills for improvement including planning, observation and listening, evaluation, and critical reflection.

Working with patients and vulnerable groups

If you are working with people with different needs, and potentially vulnerable people such as patients or marginalised groups, there are ethical issues to consider – as explored **earlier**.

Additionally, working with seldom heard groups also requires considered approaches to ensure diverse and inclusive participation. The NHS have published a series of guides for stakeholder engagement with a range of providers including **Sexual Assault Referral Centres**, and **Service user involvement (drug and alcohol)**.

In order to assess the potential consequences from participation of vulnerable people, ESRC guidelines stress that researchers

'should make the participants aware of the limits to confidentiality when eliciting consent, and decide whether verbal or written consent will be more appropriate and protective of the participants' interests.'

(Research with Potentially Vulnerable People, 2020).

Think about:

- How you can make sure that any barriers to inclusion are addressed – for instance, holding any meetings, events or focus groups in accessible venues; providing materials in community languages and offering language support
- Working with representative groups – patient charities or organisations working with particular minority groups – to identify and understand barriers to participation and how these can be addressed
- Working with these groups and with your steering group to test interview topic guides for sensitive topics and language, to ensure that issues are explored with empathy
- Providing funds to cover any travel or other costs incurred, such as childcare
- Providing incentives to participants – such as a high street or supermarket voucher – in acknowledgement of the time given to your research.

Qualitative research methods

Qualitative interviews

Interviews are the most common method in qualitative evaluation. They should be understood as "a conversation with a purpose" (Burgess, 1984). A qualitative interview is an opportunity to gain an in-depth understanding of the experiences and perceptions of people involved in designing, delivering or who are using a programme or intervention.

Interviews can be conducted in a variety of different ways – face-to-face, by telephone, on a video call – and individually or with a group (known as focus groups). When taking part in an interview, participants must have given **informed consent**.

The type of interview you choose to implement for your evaluation will vary, and fall within three categories: structured, semi structured, and unstructured.

Structured interviews

Structured interviews provide the least depth and detailed information. They are usually done when there is a well-developed understanding of the intervention or programme and a narrow focus to the evaluation. The **questions** are highly structured and a very similar to a survey. They are quick to conduct, usually have response categories to choose from and have few open-ended questions. The purpose is to minimise variation and maintain consistency; the format is standardised, so each interviewee is asked the
same series of questions in the same order. This will produce consistent data that you can compare across several respondents but will lack detail as the format does not allow for any exploration or elaboration.

For structured interviews, you can record responses straight on a prepared note template, as answers are likely to be short and concise.

Semi-structured interviews

This is the most widely used method of data collection in qualitative evaluation, and the 'conversation with a purpose' that is most effective – they create space for the participant to give, and elaborate on, their experiences and perspectives.

We recommend semi-structured interviews for your qualitative research.

An interview topic guide is prepared before the interview – and should be piloted with and **signed off by stakeholders**. The topics should be related to your research questions. They are then supported by follow-up questions – or prompts – to further explore participants' answers.

Make sure that your topic guide includes a preamble, introducing the interviewer and the purpose of the research. Start with questions about the interviewee – their role in the programme or the services they have used – before moving into questions about the experiences and perceptions.

The questions are used as a guide, providing a clear focus but also allowing for flexibility. The skill in semi-structured interviews is in guiding the conversation so that the interviewer is responding to the participant's answer, whilst exploring the issues in the topic guide. As with any conversation, the discussion can go in unpredictable ways. The participant should be free to take the discussion wherever they think is most relevant; the interviewer must balance this with keeping to the research questions. Interviewers must spend time familiarising themselves with the topic guide before interviews take place, so that they are able to engage with the discussion and not focusing on the questions.

Notes can be written during the interview, but it is best to record and then transcribe the interviews for **analysis**.

How to record your interview

If you are conducting face-to-face interviews, use a digital recorder or mobile phone. Place the device near to the participant to minimise background noise.

If you are conducting telephone interviews, you can attach the recorder (some devices) or if using MS Teams or other online telephone you can use the built-in recording facility.

Similarly, if using video (or online call), use the built-in facility. MS Teams offers recordings of both video and voice calls.

Having permission to record the call must be part of your process of securing ongoing **consent**.

Focusing on the conversation and not your topic guide will allow you to develop a rapport with the interviewee and listen attentively without worrying about jotting down notes – although you should make a note of something that is said that you want to return to later in the discussion or as a follow-up question.

There is an example topic guide **here**.

Unstructured interviews

Unstructured interviews work best when the subject under investigation is not fully understood and is still in development. They are unlikely to use this approach in evaluation. They can be done as a precursor to structured or semi- structured interviews, to find out more about the topic. But in practice, such exploratory interviews are most effective as semi-structured discussions as you will have some clear issues that you wish to explore. They are more informal and free flowing, with a minimal agenda. Similar to semi-structured interviews, notes can be made but recording is preferred to maintain rapport.

Focus groups

This interviewing technique is used to conduct semi-structured interviews with several people simultaneously. It can be a simple and quick way to collect data for a qualitative evaluation as it provides comparisons between experiences of different group participants. However, it takes skill to manage the group conversation; and it can also mean that people are not as open as they would like to be, because they are worried about what other members of the group might think or there is something that do not want to share more widely. Because of this, think carefully about whether you are going to be exploring any sensitive or potentially sensitive issues, which are likely to be unsuitable for a group discussion.

A focus group is facilitated by a lead moderator who will guide the conversation, as with a semi-structured interview. Group interactions are used as a key method to gather experiences collectively. Rather than asking each person to contribute individually, people are encouraged to talk to each other, respond and comment to generate a discussion. One of the skills for the moderator is ensuring that everyone gets an opportunity to contribute and the discussion is not dominated by one or more voices.

Group dynamics are always at play so the right size for people to feel comfortable is important. Six-seven participants is generally the optimal size of a focus group but there might be reasons to have smaller or larger groups. Who you choose to include will depend on the research topic, what is identified as important to the evaluation and the relationship dynamics at play? For example, if a focus group is conducted to evaluate a new healthcare service, it will be useful to separate patients and healthcare professionals or separate even further between doctors and nurses.

For recall and to facilitate the conversation effectively, it is best to have a separate note taker as well as record the focus group.

Observation

Observation involves the researcher spending time at the setting under evaluation, observing behaviours and interactions. It allows the evaluator to understand, experience and capture the context that is under evaluation. The first-hand experience provides an opportunity to learn things that people might be unwilling or think irrelevant to discuss in an interview. It will require engagement in a series of activities including documentation, informal interviewing, and reflection. You will need to consider whether this type of evaluation is compatible with your aims and objectives – is it important to see how things work in practice – and the heavy time commitment it usually entails. Another disadvantage to consider is that behaviours are likely to change simply because the participants are being observed, and thus do not act in a typical way. Observation might be an ongoing aspect of the evaluation to explore how things change over time (and developing a rapport so that participants are more likely to behave naturally, after a time); or a short encounter (for example, attending a ward round) to understand day-to-day delivery (where an ideal type of behaviour may be presented).

The detailed notes of any observations made including verbatim quotes will need to be recorded as soon as possible.

How to do qualitative work during COVID-19 restrictions

With social distancing requirements and other restrictions to address local spikes, there are likely to be limitations on the amount of face-to-face qualitative work you can do.

For colleagues in the same workplace, it will be possible with adherence to social distancing, assuming that you are able to use offices or rooms with sufficient space and ventilation.

Where it is not possible, and where you wish to limit travel or the use of public transport, there are a variety of ways in which you can conduct qualitative research remotely.

- Telephone interviews are a common method without COVID-19 restrictions – allowing for a greater number of interviews to be conducted within the available time, due to the lack of travel time required.
- The response to the pandemic has meant that many more people are now used to using online tools for meetings and this means that interviews via MS Teams and Zoom are now a common approach.

Conducting interviews via telephone or video conference during COVID-19 is little different to before the pandemic and the same best practice principles should be followed. But you should bear in mind that people may be at home, rather than the office, and allow for distractions such as pets, children and deliveries! As with any interview conducted remotely, ensure that your participant is in a private place where their contributions will remain confidential; and remember that there may be topics that participants do not want to discuss if there is the chance they will be overheard. As with any research, discuss with your participants what they are and aren't comfortable to discuss.

Creative approaches

There are a range of other approaches to qualitative research, which might be useful during COVID-19 restrictions.

These include:

Written diaries/journals – where participants keep a physical or electronic diary (video diaries are also possible but can be time consuming to analyse and bring additional potential concerns over confidentiality)

Using smart phones to capture voice memos or take photographs that can be written or talked about (**this page** is from a market research company but gives a short summary of how smartphones can be used to capture experiences)

Interviews conducted via **email** (an email conversation) – with the snappy title of epistolary interviews

Graffiti walls that can be placed in offices or other locations to capture perspectives

How do I analyse my qualitative data?

It is important to consider how you will analyse the data collected right at the start of planning your evaluation. This allows you to design your data collection approach with data analysis in mind; so that you have right resources in place and plan for your analysis and reporting. In qualitative approaches the data collected is words or text. What is most important to bear in mind is that you must take a structured approach.

Broadly, you are going to review your data – written notes or transcripts of interview audio recordings, most commonly – to identify the themes that emerge. Start with a sample of the data, making notes of the key themes. You can organise these under your research questions. Ideally, more than one member of the evaluation team will look at the same sample of data. Compare your notes to check that you are interpreting the data in the same way.

When you have agreed on a set of codes, apply them (again, more than one person should be involved) to a further set of data to check that they work in the same way and add any further themes that emerge.

You have now developed your coding framework and can apply it to your dataset.

A commonly used and accessible method is the Framework Approach. This enables you to cut and paste your data into a table so that you are able to see what your participants have said about each theme as well as what each participant has said across all of your themes.

An example Framework for qualitative analysis

	Workforce experiences				Patient experiences				*Continues...*
	Positive changes	Negative changes	Leadership	Equipment	Positive experiences	Negative experiences	Views of clinicians	Barriers to access	
Participant 1									
Participant 2									
Participant 3									
Participant 4									
Continues...									

There is more information and worked examples in this academic **paper**, written in a fairly accessible style.

For more guidance on coding and approaches to analysis, see these resources from **Better Evaluation**.

Software and practical tools

There are a range of Computer Assisted Qualitative Data Analysis (CAQDAS) packages on the market, all of which require some level of training and software. They are designed to support the organisation, management and analysis of data, they do not replace human interpretation. Think carefully and take advice from the **DSC** before purchasing any specialist software.

This **resource** helps you choose an appropriate package for your need.

How do I carry out quantitative research?

- Selecting the right method
- Demonstrating causality
- Selecting metrics
- Sourcing data
- Aggregate vs record level data
- Data linkage
- Information governance
- Data validation
- Different quantitative designs
- Quasi-Experimental Methods

Quantitative methods are generally used to measure progress within formative evaluations and to measure impact within summative evaluations. They are also used to measure the cost effectiveness of programmes and interventions in economic evaluations. These methods are most robust when they compare the intervention with what would have happened without it – commonly referred to as a counterfactual (or Business as Usual (BAU)).

Here we describe several common quantitative methods used to evaluate health and care interventions, when different methods may be applicable and their pros and cons. It is not a definitive guide and other methods are available.

The ones defined are:

- Pre-post Studies
- Randomised Control Trials
- Quasi-experimental study designs including:
  - > Interrupted Time Series (ITS) analysis
  - > Retrospective Matched Cohort Studies
  - > Synthetic Controls
  - > Regression Discontinuity

These and other methods are summarised in the Magenta Book Annex A.

Selecting the right method

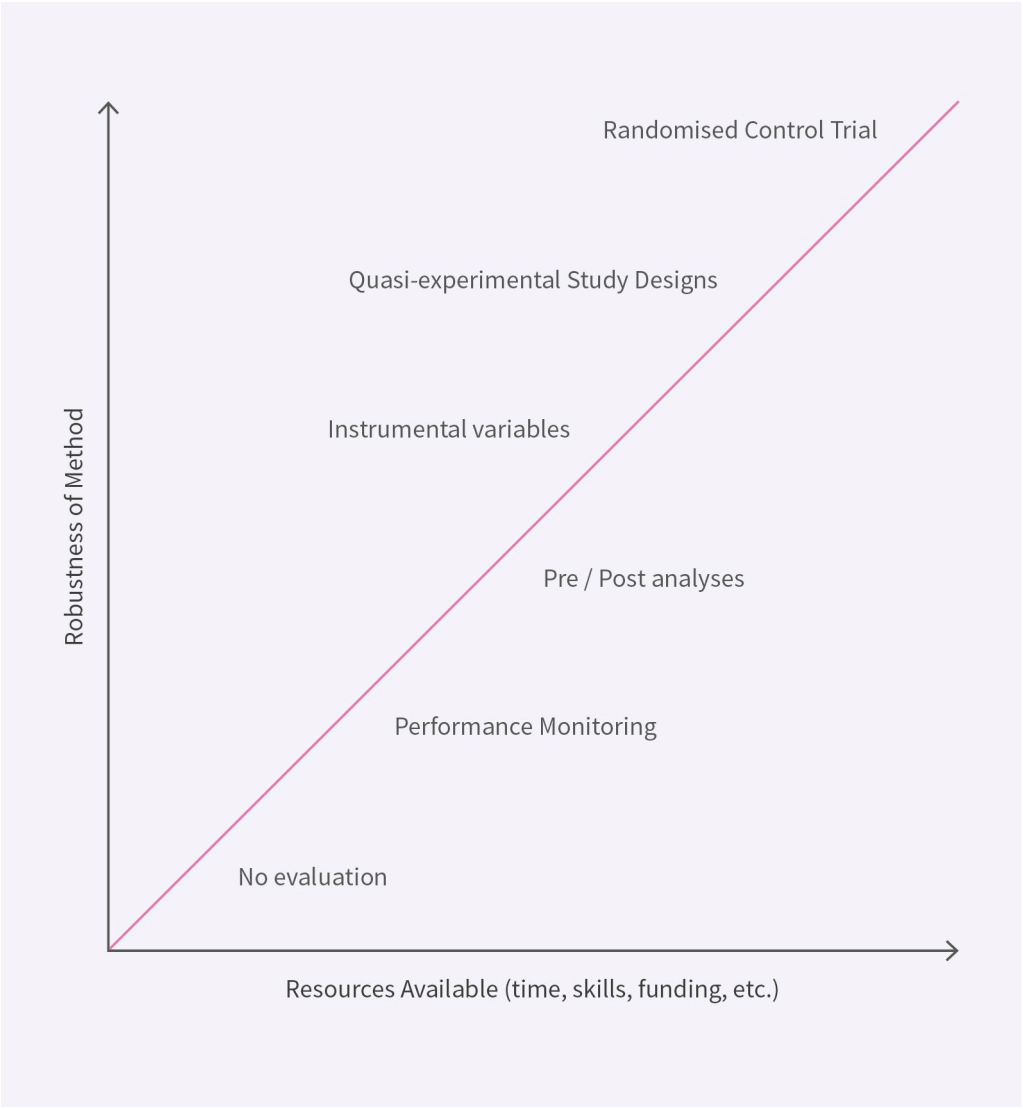
There are several factors that will influence the best method to apply, including:

- Available budget
- Level of evidence required to make decisions on continuing, ending or scaling up the intervention
- Available data
- The type of intervention
- Recruitment criteria for the intervention.

There is generally a trade-off between the resources required for the analytical methods and the robustness of the method. Randomised Control Trials provide the most robust evidence of the causal effect of an intervention but are generally very expensive and can be impractical, as they require randomisation of the intervention or treatment; whereas pre-post studies are relatively inexpensive but are unlikely to demonstrate causality.

The chart opposite shows ranks some of the most common methods used in terms of the costs and robustness of the method. Each method has its own strengths and limitations and no method is perfect.

Balancing available resources against robustness of method



The data available from local and / or national datasets will influence the methods available. Access to data on an appropriate counterfactual – a control group – is required for each method. RCTs and retrospective matched cohort studies require access to detailed record level dataset whereas other method may be possible using aggregate level datasets.

Some of the methods can only be applied if access to record level data is available (e.g. Retrospective Matched Cohort Studies and RCTs) whilst others can be done using aggregate datasets (e.g. ITS and Synthetic Controls). This is explored in more detail later.

When thinking about the counterfactual or control group we must consider:

- Whether the data is of the required quality and quantity for the analysis
- If the counterfactual is genuinely comparable to the intervention group
- The size of the intervention effect and whether it can be distinguished from 'noise' in the data.

The **Magenta Book** (section 2.2.4) provides a good introduction to supplement the outline here.

The type of intervention and the way it is implemented may preclude the use of methods that require randomisation if the whole population will be exposed and therefore no controls can be identified. Certain methods are more appropriate for interventions that are piloted in part of a locality with controls being identified from other parts of the area. Other methods require the impact of the intervention to happen immediately and not be delayed. For example, an ITS design requires the impact to start soon after implementation for it to detect the effect.

The table on the **next page** provides a summary of the methods described in this guide.

Summary of Quantitative Evaluation Methods

Randomised Control Trials | Pre post studies | Interrupted Time Series (ITS) analysis | Retrospective Matched Cohort Studies | Synthetic Controls | Regression Discontinuity

Randomised Control Trials

Overview	When to apply	Data required
• Gold standard • Compares cases with a randomly selected control group • Prospective analysis • Require ethics approval	• High cost interventions • High risk interventions • When a higher level of certainty around causality is required	• Record level • Cases and controls

Pros	Cons
• Demonstrates causality • Reduces selection bias	• Expensive and time consuming • Results may not be generalisable

Summary of Quantitative Evaluation Methods

Randomised Control Trials | **Pre post studies** | Interrupted Time Series (ITS) analysis | Retrospective Matched Cohort Studies | Synthetic Controls | Regression Discontinuity

Pre post studies

Overview	When to apply	Data required
• Compares cases before and after implementation • Retrospective analysis	• When other more robust options are not possible	• Aggregate • Cases only

Pros	Cons
• Easy to set up • Inexpensive	• Do not demonstrate causality

Summary of Quantitative Evaluation Methods

Randomised Control Trials | Pre post studies | **Interrupted Time Series (ITS) analysis** | Retrospective Matched Cohort Studies | Synthetic Controls | Regression Discontinuity

Interrupted Time Series (ITS) analysis

Overview	When to apply	Data required
• Compares cases with a projection at the point of implementation • Retrospective analysis	• Intervention can be clearly defined (at population level) • When experimental methods are feasible • Best with expected fast acting interventions on the outcome of interest	• Aggregate • Cases only (ideally routine data pre and post intervention at equal time points)

Pros	Cons
• Only requires data from the case/unit of interest • Easy to implement • More detailed assessment of longitudinal impact	• Need to account for autocorrelation[1], seasonality, secular trends[2] • Lack of a control population • Cannot be used to make inferences on individual-level results

[1] Autocorrelation refers to the degree of correlation between the values of the same variables across different observations over time. For example, if values that occur closer together in time are, in fact, more similar than values that occur farther apart in time, the data would be autocorrelated.- https://en.wikipedia.org/wiki/Autocorrelation

[2] Secular trends are long-term trends that develops or progress over many years

Summary of Quantitative Evaluation Methods

Randomised Control Trials | Pre post studies | Interrupted Time Series (ITS) analysis | **Retrospective Matched Cohort Studies** | Synthetic Controls | Regression Discontinuity

Retrospective Matched Cohort Studies

Overview	When to apply	Data required
• Compares cases with a matched control group • Controls are identified using matching variables • Retrospective analysis	• When record level data is available for both the intervention and control groups • When an intervention is already being delivered	• Record level • Cases and controls

Pros	Cons
• Uses existing datasets • Can be applied interventions already in place • Can provide intermediate results to track progress over time	• Requires data routinely collected at record level and permission to use and link this data. • May not be possible to match all members of the intervention group • Other factors, not accounted for in the matching, may explain differences between the intervention and control groups (unobserved confounding).

Summary of Quantitative Evaluation Methods

Randomised Control Trials | Pre post studies | Interrupted Time Series (ITS) analysis | Retrospective Matched Cohort Studies | **Synthetic Controls** | Regression Discontinuity

Synthetic Controls

Overview	When to apply	Data required
• Compares cases with a synthetic control group • Retrospective analysis	• When intervention can be clearly defined • Data for both the treated and potential control units are readily available	• Aggregate • Cases and controls

Pros	Cons
• Can estimate causal inference • Considered more robust than elementary Difference in Difference (DiD) approaches	• Hard to detect “small effects” on interventions • Requires larger amounts of data compared to ITS or other methods • groups (unobserved confounding).

Summary of Quantitative Evaluation Methods

Randomised Control Trials | Pre post studies | Interrupted Time Series (ITS) analysis | Retrospective Matched Cohort Studies | Synthetic Controls | **Regression Discontinuity**

Regression Discontinuity

Overview	When to apply	Data required
• Compares cases with a projection at cut off point for selecting cases • Retrospective analysis	• When a threshold, such as a risk stratification score higher than x, defines eligibility for an intervention	• Aggregate • Cases only

Pros	Cons
• Easy to conduct • Does not require randomisation • Can be combined with other methods if needed	• Only allows estimation of impact for individuals close to the cut-off • Assumptions must be made to assert causality

Demonstrating causality

The key objective of quantitative evaluation is to demonstrate causality, in other words did the intervention have a direct impact on the outcome being measured. The two main elements to consider when deciding if your study is likely to show a causal effect are the sample size (how many people were exposed to the intervention) and the effect size (how different is the outcome variable from the counterfactual). The larger the sample size the more likely smaller effects can be evidenced in the analysis.

Power calculations can be made to understand the sample size required to demonstrate causality. For evaluation, power calculations are likely to be done at the start of the study to determine the required sample size based on an expected effect size. This will help identify the scale and timescales for the intervention to ensure any impact can be evidenced. Online Tools, such as the **ClinCalc Sample Size Calculator**, can be used to perform these calculations.

Selecting metrics

An essential element of a quantitative evaluation is selecting the right metrics to analyse. They should be chosen based on the outcomes defined in your logic model. If you are undertaking a formative evaluation, metrics that measure the progress of the intervention (process metrics) may be needed alongside outcome metrics. These will provide an early indication of whether the innovation is being implemented as expected and is therefore likely to deliver the outcomes specified in the logic model. Summative evaluations are likely to focus mainly on outcome metrics.

There are several national outcome frameworks, with pre-defined metrics, available to source the right outcome metrics for your evaluation. These include:

- **NHS Outcomes Framework (NHSOF)**
- **Public Health Outcomes Framework (PHOF)**
- **CCG Outcomes Indicator Set (OIS)**
- **Adult Social Care Outcomes Framework (ASCOF)**

Process metrics are likely to be project specific and therefore will need to be developed locally. They should be developed at the start of the evaluation and monitored regularly.

All process and outcome metrics used in the evaluation should be:

S **Specific:** to measure the information required as closely as possible

M **Measurable:** to ensure that the information can be readily obtained

A **Attributable:** to ensure that each measure is linked to the project's efforts

R **Realistic:** to ensure that the data can be obtained in a timely fashion, with reasonable frequency, and at reasonable cost

T **Targeted:** to the objective population (also comparative population)

Metrics should be fully defined and agreed with stakeholders at the start of the evaluation. An evaluation metrics framework can be useful in defining the metrics and how they will be used in the evaluation. This is likely to include:

- Metric Name
- Metric Definition
- Numerator
- Denominator
- Data source
- Frequency of Reporting (monthly, quarterly, annually, etc.)
- Period Reported (monthly, year to date, snapshot, etc.)
- Reporting Level (trust, GP practice, CCG, etc.)
- Baseline Data (period, numerator, denominator and actual result)
- The outcome in the logic model the metrics measures
- Rationale for the metric and any additional notes

Sourcing data

When selecting the most appropriate metrics it is important to identify how you will source the data required to measure them. This includes identifying what data is available either nationally or locally, who holds the data, the format of the available data, the Information Governance (IG) measures required to access the data and the quality of the data. It is also important to understand if data can be sourced for any counterfactual or control group required for the selected method. If specific data is only being collected for people exposed to the intervention it is unlikely to be available for any control group.

Aggregate vs record level data

Different evaluation methods require different granularities of data and therefore it is important to understand what data is available before selecting your method. Aggregate data has been grouped and summed whilst record level data includes a single record for each subject. Record level data is more flexible but can be difficult to access. For example, an Interrupted Time Series study can be undertaken using aggregate data whereas a Retrospective Matched Cohort study requires record level data. An assessment of what type of data can be sourced within the available resources of the evaluation should be undertaken at the start of each project. This should consider:

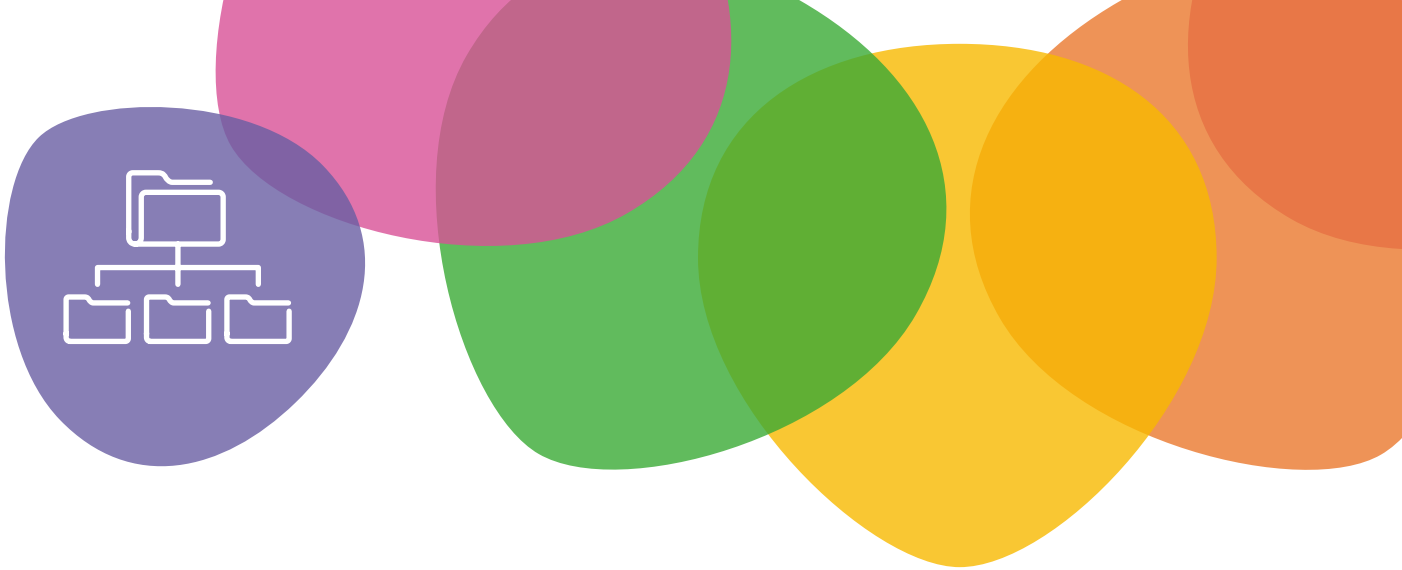
- The outcomes you are evaluating – some outcomes are readily available within national and / or local aggregate level datasets, but others will require record level data to precisely measure them.
- The type and implementation of the intervention – how the intervention is designed and implemented will influence the level of data required. For example, if the intervention is being piloted in part of the locality you are evaluating then it may be possible to use data other parts of the locality as the counterfactual.
- The availability of data (including for any counterfactual) – different data is available at different levels and therefore aggregate or record level data may be the most appropriate for your evaluation.
- The cost of sourcing the data- some record level national datasets have costs that will need to be built into the budget for the evaluation if they are required.
- The Information Governance (IG) approach – record level data is more sensitive and therefore requires additional controls that take time and will need to be planned into any project.
- The timescales for the project – record level datasets generally take longer to source than aggregate datasets and therefore may not be suitable for evaluations with short timescales.

There are several nationally defined record level datasets available that can be used within evaluations, including:

- Hospital Episode Statistics (HES)
- Secondary User Service (SUS)
- E-Referral Service (ERS)
- Mental Health Service Dataset (MHSDS)
- Community Services Dataset
- Cancer Outcomes and Services Dataset (COSD)
- Cancer Wating Times (CWT)

Useful aggregate level datasets can be found at:

- NHS Digital
- NHS England
- Public Health England
- Quality Outcomes Framework (QOF)



Data linkage

Quantitative evaluations often require outcomes to be measured across pathways. When evaluating an intervention in one part of the system that has an impact on another part of the system then data linkage may be required. This will allow you to track the outcomes individuals across the pathway. For example, if you are tracking the impact of a community service on acute urgent care then you will need to use a community dataset to identify the individuals exposed to the intervention and then link their data with an acute dataset to track their urgent care usage.

There are several factors that need to be in place to allow you to link person level data for local evaluations, such as:

- Data that is routinely collected at a person level
- High quality data is required from all partners to ensure the data can be linked
- A consistent Identifier in all datasets (e.g. NHS Number) that can be used for matching individuals
- An agreed Information Governance Framework including a legal basis to share each dataset, a defined purpose as to why the data needs to be shared and Data Sharing Agreements signed by all parties
- An understanding of who will be undertaking the data linkage.
- Senior buy-in from each partner to unblock any data sharing and information governance issues.

Linking datasets is relatively straightforward and requires you to have a single consistent identifier (pseudonymised NHS Number) in each dataset. This identifier can be linked using joins within a database. The way the data is linked will depend on the evaluation method but should at least include every person who had the intervention in the period you are evaluating.

To track outcomes across the linked datasets you need to be able to identify the overlaps in the datasets. In particular, you need to identify those people who were subject to the intervention in the period you are evaluating in all datasets. If you are undertaking a study design with a control group, you will need to ensure you can identify any outcome and matching variables in the linked dataset.

For example, if you are evaluating the impact of a new mental health service on acute emergency admissions using a linked dataset then it is important all the matching and outcome variables are available in the linked dataset for the new service and any potential controls. Any data collected only for the new service cannot be used – although it will provide you with information on use of the service, it won't establish causality in a quantitative evaluation.

Information governance

Access to the right data is important for all evaluation projects. At the start of each evaluation it is important to seek advice from an Information Governance specialist to ensure any data required can be shared legally. They will help you complete a **Data Protection Impact Assessment (DPIA)** to identify the data required, who holds the data, the legal basis for sharing the data, how it will shared, the risks associated with sharing the data and the mitigations required to minimise those risk. The is a requirement of the General Data Protection Regulations (GDPR) and the Data Protection Act (2018) and should always be completed at the start of any evaluation.

Responses to the DPIA can be translated to a Data Sharing Agreement (DSA) or contract to underpin any data required for the evaluation. This would specify all the details needed to share and process the data legally, including:

- The data sharing purpose
- The legal basis for sharing
- Details of all the data to be shared, the organisation sharing the data (Data Controllers) and the organisation receiving the data (Data Processors)
- Technical details of how the data will be shared, stored and accessed securely
- Details of how each organisations **Privacy Notice** will be updated
- Any procedures for dealing with **Subject Access Requests**
- Timescales for the data sharing, including details of how and when the data will be deleted at the end of the evaluation.

More information on GDPR, DPIAs and DSAs can be found on the Information Commissioners Office (ICO) **website**.

Data validation

It is important that any data is checked before it is used in the analysis. Any issues in the data when it is shared with you will impact on the analysis. The issues are multiplied if you are sourcing the same record level dataset from more than one organisation and/or if you are linking datasets. There are several issues that can occur when sourcing and linking record level datasets, summarised in the table on the **next page**.

Common Data Quality Issues

Issue	Solution
The NHS number is pseudonymised using a different key	Compare full list of Pseudo ID's in both datasets and check if any match. If no ID's match, then a different key has been used.
Null or blank NHS Numbers have been pseudonymised	List all Pseudo ID's in the entire dataset and count how many times they appear. A Pseudo ID may appear many more times compared to others. Check if activity makes sense or not. If in doubt, check with the pseudonymisation key holder who should be able to trace the Pseudo ID back to its original value.
The activity dates are inconsistent or not accurate	Check any date field to see if date formats are consistent. E.g. YY-MM-DD or DD-MM-YY.
Activity has been double counted due to many to many relationships	Find records where all contents exactly match, therefore appear as duplicates. If found, check how the data has been joined.
The data formats in different datasets do not match	When importing data check the field format is what it should be. If this has been missed at the import stage, then the field can often be reformatted in the database.
Records are missing in the linked dataset	Check with data provider how many records there should be in the data set and compare this to what you have in your linked dataset from that provider
Duplicate Pseudo ID appears	A person's birthday may have passed between two contacts therefore having 2 ages. You will need to decide which age to assign to the patient so to avoid creating two patients in the linked data. This can also happen if the postcode changes. Creating an index table of all unique IDs and then allocating the most appropriate demographic details (e.g. latest age and postcode) to these IDs can help mitigate this issue.
NHS Number collection is poor in one or more dataset	An assessment of whether NHS Number compliance is poor will need to be taken before these datasets are used. Datasets with poor compliance should either not be used or caveats need to be clear in the method.
There is inconsistent coding over time	Look for sudden shift in trend from one specific period to the next e.g. month/financial year
The datasets use different coding methods	List unique codes within a field with their descriptions to identify codes that do not match their descriptions set out in the data specification.

Different quantitative designs

Randomised Control Trials (RCTs)

Randomised Control Trials (RCTs) are the gold standard for quantitative evaluation, in which a number of similar people are randomly assigned to two (or more) groups to test a specific drug, treatment or other intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, a dummy intervention (placebo) or no intervention at all. The groups are followed up to see how effective the experimental intervention was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically to test the effectiveness of the intervention. In this way, the method addresses potential bias.

Because RCTs are the gold standard they are presented first in this guide, as they sit at the top of any hierarchy. But there are practical issues which mean that they are unlikely to be suitable for most, if not all, the evaluations your DSU or the network carries out.

When should RCTs be used?

A well designed RCT can provide strong evidence of causality when assessing the impact of an intervention on the measured outcomes. They are most appropriate for evaluating more expensive interventions and / or those that may have severe adverse impacts, such as drug trials.

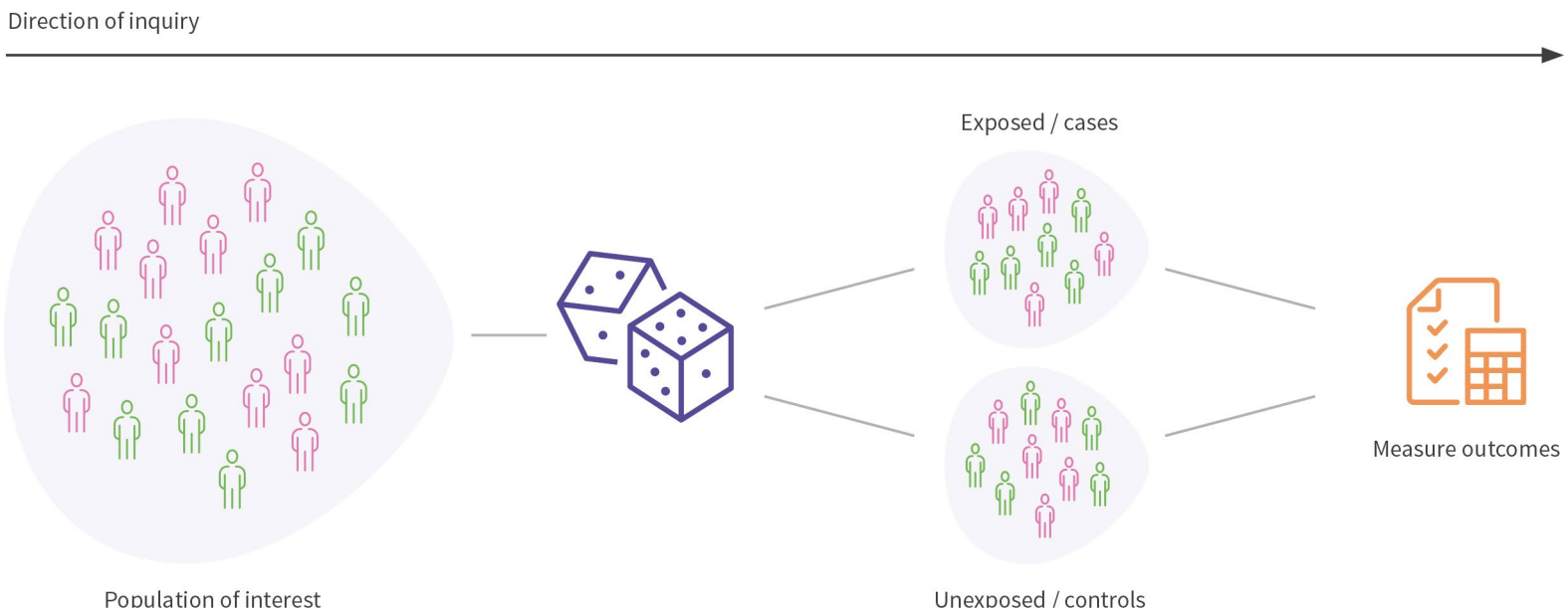
How to use RCTs?

At the start of any RCT you will need to check with the Health Research Authority (HRA) about the level of ethics approval required for the evaluation. Protocols need to be developed to identify who will be included in the trial

Each person who registers for the trial should be provided with detailed easy read information on the design and risks of the trial. They will need to consent to be involved in the trial.

Each person is then assigned to either the experimental group or one or more control group who are either provided with an alternative treatment or no treatment. The assignment is random and therefore every person has the same chance of being in any of the groups. Clustered randomisation can be used for interventions that may have different impacts on different population groups. This ensures a sufficient sample size in each population group for analysis. People are added to the right population group before being randomised to the experimental group or a control group.

Random Selection in RCTs



Different analysis methods will be applied depending on the design of the trial and the outcomes chosen. These may include:

- Logistic regression
- Analysis of covariance
- Survival analysis
- Difference in Difference analysis [link to DiD section]

Possible limitations

RCTs are costly and time consuming to set up and run. They need to be set up before the start of the intervention and are therefore not appropriate for retrospective evaluation. They require strict ethics approval to ensure they are run safely which can take time to be approved.

They are not suitable to any intervention where it is not practical to randomise who is included in the experimental group and who is not. For example, if an intervention is targeted at a whole population then it is not possible to choose who is in which group. Due to the strict eligibility criteria of RCTs the results may not be generalisable for groups that do not meet the criteria. It is not always possible to be confident that the control group is not affected by other, unconsidered, variables or factors.

Some people have concerns that by randomising the intervention you are denying potential benefits to the non-treatment group. Others argue that you should not implement an intervention at scale until it has been proven through a robust RCT design. If you are interested in exploring these debates in more detail this **paper**, written to promote RCTs in public policy, presents a persuasive (although not uncontested) case.

Pre-Post Studies

A pre-post study measures one or more outcome before and again after an intervention is implemented. They can be used to measure one group pre and post intervention or to compare more than one group before and after the intervention. This could include a group that did not have an intervention.

When should Pre-Post Studies be used?

Pre post studies are reasonably straightforward to undertake and require minimal data to run. They should be used where a more robust method is not practical and there is no requirement to demonstrate causality. These studies have the strength of temporality to be able to suggest that the outcome is impacted by the intervention, however, they do not have control over other elements that are also changing at the same time as the intervention is implemented.

How to use Pre-Post Studies?

Each outcome is measured before and after the introduction of an intervention. Any differences are then assumed to be related to the new intervention. The chart below shows an example of a pre-post study, comparing one outcome variable pre and post a new intervention. The result of the variable post intervention is compared to the pre intervention result to show the impact. Confidence limits should be used when reporting the analysis. A Confidence Interval sets out the range of values that can be confident our true value lies within. It is required when you are using an average from a sample population. Confidence limits are the upper and lower limits of the confidence interval and are usually set at 95%. This means that we are 95% sure that the average is within this range.

Example of a Pre-Post Analysis

If you are comparing the results against a similar system elsewhere then a Difference in Difference (DiD) analysis can be applied. DiD compares the change in outcome within the intervention group to the change in the outcome within the other system, over two time points.

The table below gives an example of how a results table may look for a difference-in-difference analysis. It illustrates that A&E attendances were reduced by 0.5 per person amongst your service users compared to the matched group.

Example of a Difference in Difference analysis

	Your Patients			Matched Cohort			
	Before (a1)	After (b1)	Change (c1)	Before (a2)	After (b2)	Change (c2)	Effect (c1-c2)
A&E attendances per person	4	3	-1	4	3.5	-0.5	-0.5

Requirements for Pre-Post Studies

The main requirement is ensuring that the 'pre', or baseline, data is collected prior to the start of the intervention. It requires aggregate level data for each outcome for a period before and after the introduction of the intervention.

Possible limitations

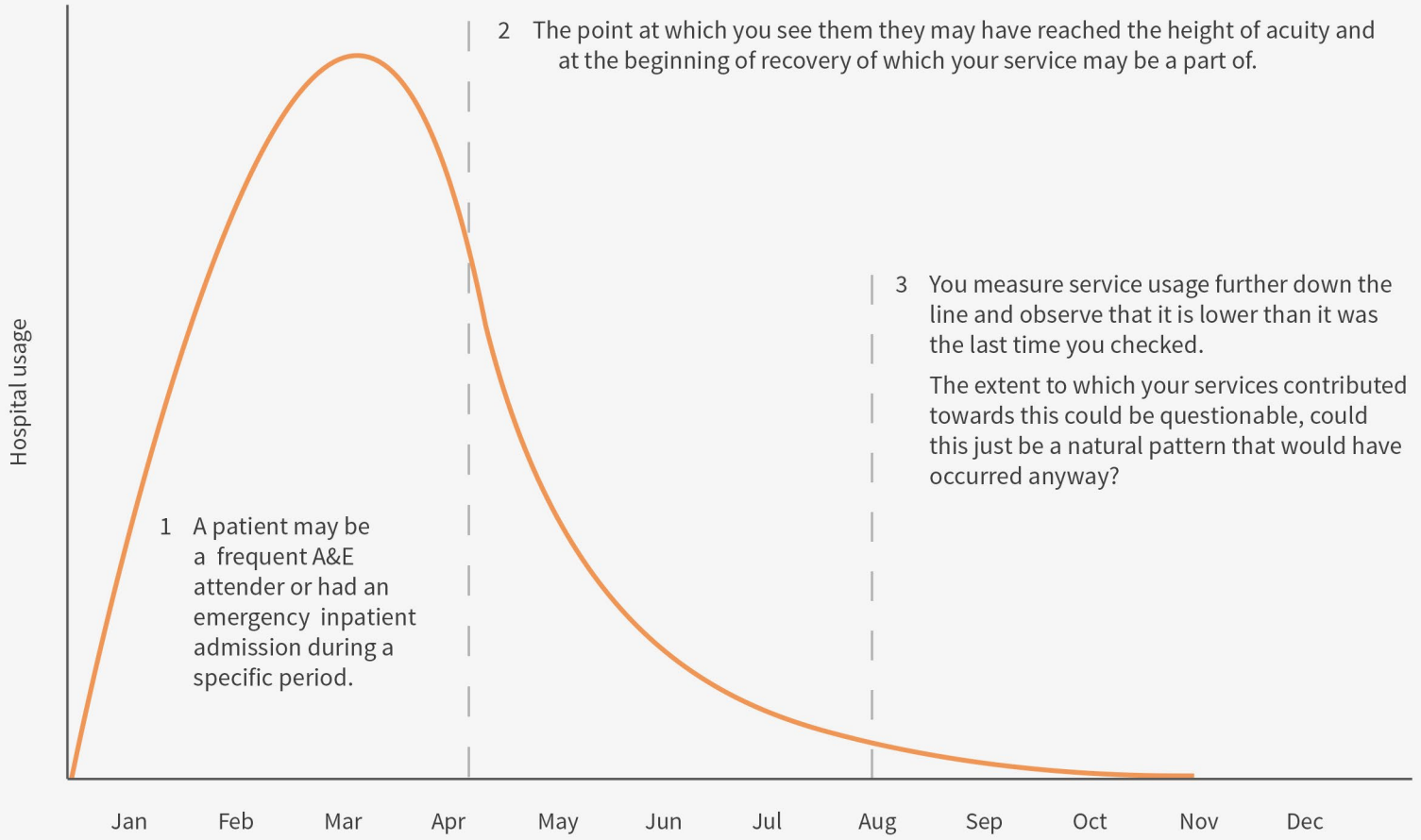
Pre-Post Studies do not demonstrate causality as they do not control for other factors that may have impacted on the outcome. People may have improved without any intervention. They are also at risk of 'regression to the mean'.

The issue of 'regression to the mean' can occur whenever something which varies over time is measured once and is then measured again at a later point in time. Observations made at the extreme the first-time round will tend to come back to the population average the second time round.

Regression to the mean is a challenge when an intervention is focused on specific types of patients (for example patients with high emergency care use). Say we look at people with frequent hospital admissions at present. On average, these individuals will have lower rates of unplanned hospital admissions in the future, even without intervention. So, if a therapist is working with patients who are currently having frequent A&E attendances, they may notice how the patients have fewer admissions over time. However, this reduction might well have occurred anyway due to regression to the mean, and it cannot necessarily be attributed to the input of the therapist. Regression to the mean occurs simply because after one extreme period, the next period is statistically likely to be less extreme.

The way to control for regression to the mean is to create a matched control group and look for differences between the intervention group and a similar control group. If you are using a before and after approach you should reflect that this may be an issue when reporting.

Regression to the Mean



Quasi-Experimental Methods

The gold standard for quantitative evaluation is a Randomised Control Trial (RCT). As already outlined, RCTs are generally time consuming and expensive to run (and there are debates about randomising potentially beneficial interventions); they are unlikely to be practical to evaluate most NHS service changes and innovations.

Quasi-experimental study designs represent a pragmatic alternative to conducting a RCT. Such designs arise from different ways of attempting to control for third (confounding) variables without using random assignment. These methods also help control for 'regression to the mean' which is an issue with pre post study designs.

The most appropriate approaches likely to be used are:

- Interrupted Time Series (ITS)
- Synthetic Controls
- Regression Discontinuity
- Retrospective Matched Cohort approach

Interrupted Times Series (ITS) analysis

Interrupted Time Series analysis (ITS) otherwise known as segmented regression analysis is an approach that can be applied retrospectively to measure how an outcome variable changes as a function of the onset of an intervention. Segmented regression analysis provides an assessment of whether there has been a statistically significant change in either the trend or the level of an outcome measurement.

When should Interrupted Times Series (ITS) analysis be used?

ITS can be applied to many population-level healthcare interventions where the intervention is clearly defined at a point in time. They offer a way to analyse the longitudinal nature of data which may often not be possible with RCTs, and given that they are often used in real-world scenarios, can have stronger external validity (they are more likely to be generalisable to different contexts).

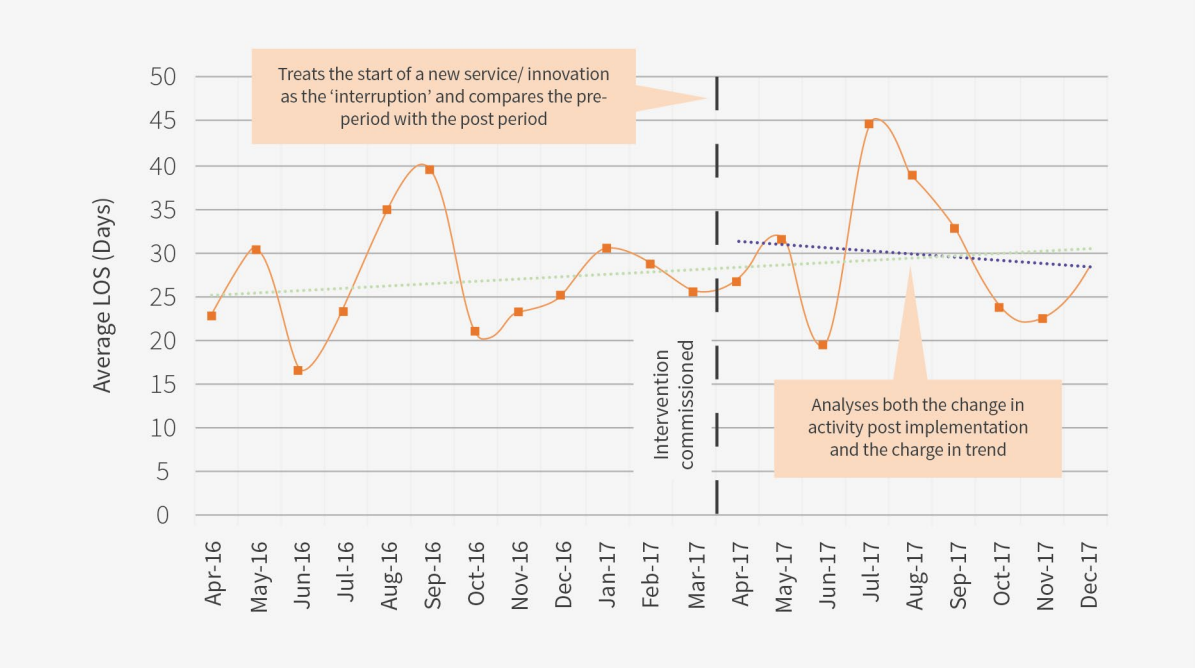
As a quasi-experimental method, ITS allows us to infer causation.

ITS can be widely understood by many practitioners without a quantitative background, and the results can be shown both numerically and graphically.

How to use Interrupted Times Series (ITS) analysis?

In an ITS design, repeated measurements of an outcome variable are taken before and after an intervention (interruption) is introduced. Statistical methods are used to assess whether the intervention has had a significant effect on the time course of the outcome variable that is greater than any underlying trend. The use of multiple data points as opposed to a standard pre/post test design reduces some of the threats to validity that can affect other non-experimental designs.

Example ITS Analysis



There are several statistical techniques that can be used for the ITS depending on the characteristics of the data, the number of data points available and whether autocorrelation is present. Those techniques include simple linear regression, autoregressive integrated moving average model (ARIMA) and Bayesian structural time series model.

Requirements for an Interrupted Times Series (ITS) analysis

The dataset should feature continuous measurements of the variable/outcome of interest over time, ideally at equal intervals. There is no fixed number of data points expected, but it should be noted that power calculations could be affected if there are too few time-points, or if the ratio of data points pre and post is unbalanced by a significant degree. Thus, routine data sources are usually the best avenue for collecting your data.

Other considerations for the use of ITS include the time point of the intervention being clearly defined. This clear demarcation between the pre and post-intervention time frames is necessary for segmentation. The intervention does not need to be on a single point in the time-series – it can be gradual, with this change modelled as a "slope" post-intervention.

The outcome variable works best when it is expected to change hastily after an intervention is applied, or after a short lag.

Possible limitations

ITS analysis may not be appropriate if interventions are gradual rather than abrupt and/ or if the causal effect of an intervention is delayed in time.

Threats to internal validity – the confidence that we can have in the causality established – include:

- Simultaneously occurring interventions ("co-interventions")
- Changes in the composition of study population
- Changes in the measurement of the outcome.

However, these threats can be mitigated by the inclusion of appropriate control groups/variables.

Retrospective Matched Cohort Studies

A Retrospective Matched Cohort Study compares outcome metrics between like for like groups of patients where one group has been exposed to the intervention (cases) and the other has not (controls). The difference quantified between both groups is indicative of the impact of the service.

The Nuffield Trust have written a more **comprehensive guide** that provides a lot more detail that our outline here.

When should Retrospective Matched Cohort Studies be used?

Retrospective Matched Cohort Studies use existing data and are therefore easy to set up where access to record level datasets is available. This means that several outcomes can be tracked using the same dataset. They do not require the same strict inclusion criteria as RCTs so can be used to evaluate interventions in 'real world' settings.

They are less expensive and time consuming than RCTs and can be applied retrospectively. Therefore, they can be carried out on interventions that have already been implemented. They can be applied at multiple points in time to track the progress of an intervention.

How to use Retrospective Matched Cohort Studies

Retrospective Matched Cohort Studies use data to match people exposed to the intervention with similar people who were not. They generally include demographic (age, gender, ethnicity, deprivation, etc.), need (e.g. long-term conditions, etc.) and prior service utilisation (e.g. A&E attendances, emergency admissions) variables. All matching criteria should be agreed prior to the analysis starting to avoid the controls being found to show the right result. There are two main options used in healthcare evaluations for matching the cases and controls:

- Matching several of the underlying characteristics at once, without attempting to summarise them into a single figure (e.g. demographic, clinical and prior utilisation variables)
- Matching using a Propensity Score (e.g. Risk Stratification Score).

The analysis needs be run following the full implementation of the service and a sufficient period to allow the new service to be delivering the expected impacts. This will help ensure the sample size of patients seen and the effect size of the service are sufficient to statistically show impact should this be present.

The cases group would comprise of patients that have received the intervention during the study period. Selection of the control group would depend on how the intervention is implemented and would either be selected from another similar local or national geography where the intervention has not been implemented.

Cases and controls would then be matched on either a set of locally agreed variables (e.g. age, gender, deprivation, prevalence, A&E attendances in the previous year, etc.) or a propensity score.

Despite matching there may remain slight underlying differences between the intervention and control groups, so an analysis may be required to compare the changes within the two groups relative to their baseline – and test whether the change in outcome found in the intervention group is greater than that found in the control group. A difference-in-difference approach compares the change in outcome within the intervention group to the change in the outcome within the control group, over two time points.

Requirements of Retrospective Matched Cohort Studies

Retrospective Matched Cohort Studies require access to record level datasets that are sufficient to identify controls for everyone who is exposed to the intervention. This may include data linkage across multiple organisations. Robust IG processes and senior level support from all organisations are required for any data linkage project.

The matching variables should be agreed with clinicians and managers involved in the intervention to ensure they capture the most appropriate factors, such as the eligibility criteria, to ensure the matched controls are as close to the intervention group as possible. Protocols for dealing with multiple matches, where more than one control per case is identified, and non-matches, where no control can be identified, are required and should be documented when writing up the analysis.

Possible limitations

Retrospective Matched Cohort Studies are only appropriate where the matching and outcome variables are routinely collected at a person (record) level. The method is likely to require data linkage and therefore you will need to be able to access all the data required for the matching and outcome variables and have IG arrangements in place to link the data. This can be costly and time consuming.

In some cases, not all members of the intervention group can be matched with a similar person in the matching pool. These should be excluded from the analysis but reflected in your reporting.

The matching process can create very similar groups based on the agreed matching criteria, but there might be other, hidden factors that explain differences between the intervention and control groups (unobserved confounding). If these factors are not built into the matching process, they can create bias in the analysis. Any possible confounding factors should be included in your reporting.

Synthetic Controls

Synthetic Control (SC) methods have become an increasingly popular choice of programme evaluation, building upon previous "difference-in-difference" analyses, including interrupted time-series.

Synthetic Controls use a combination of multiple control units as the counterfactual, testing whether the intervention had any statistically significant effect.

When should Synthetic Controls be used?

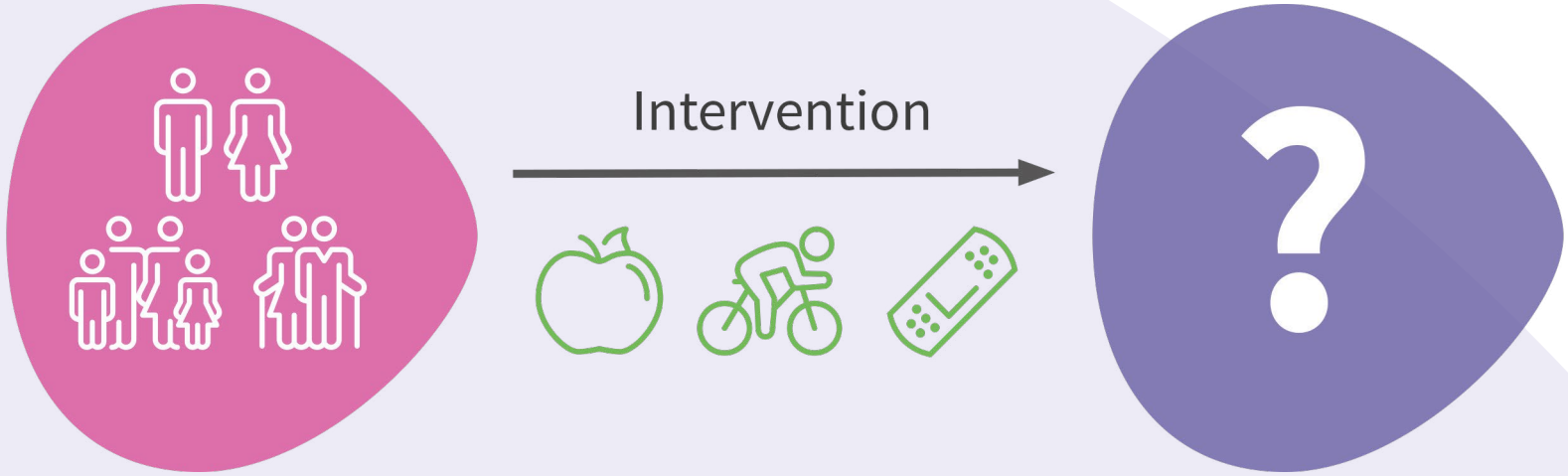
Synthetic Controls give us the opportunity to extend observational studies. Unlike other difference-in-difference approaches, Synthetic Controls do not explicitly rely on parallel pre-implementation trends.

How to use Synthetic Controls

1) Ensure the conceptual framework behind theory of the intervention is well understood

The researcher must understand what the main outcomes of interest are, and ask what variables could be (influencing (confounding) the apparent effect? What segment of the population does the intervention target? Can we clearly define boundaries between the target and control groups?

Synthetic Controls: Conceptual Framework

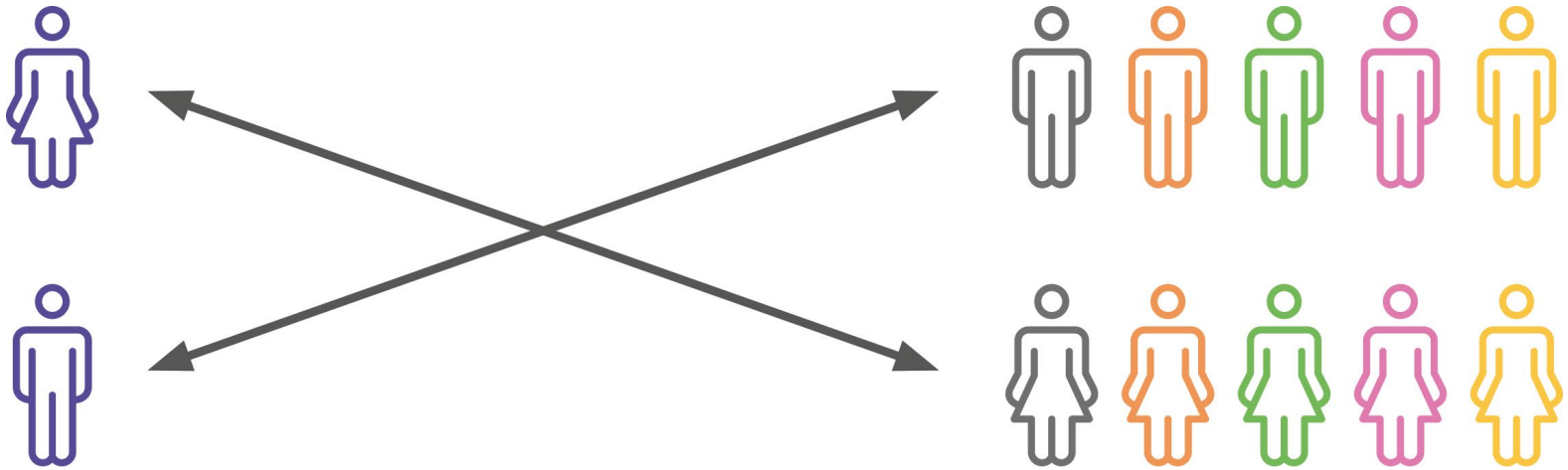


2) Identify the possible donor control units

The pool of potential controls (also known as the donor pool) should reflect the characteristics of the treated Unit, in terms of pre-intervention values, and the existing theory/conceptual model behind the analysis. For example, if you are studying the effect of an intervention on a GP practice, it would make greater sense to include other GP practices geographically close in the donor pool. The counterfactual is constructed from a selection of these donor units.

The similarity of pre-intervention values between the donor units and the treated Unit, should be similar not only with the main outcome of interest, but also with other predictor values which could influence the outcome.

Identifying the Right Donor Pool



The data is then collated. The dataset for the donor pool units and treatment units should be in a standardised "panel" format, whereby observations for both dependent and independent variables are longitudinal.

Then, initiate the Synthetic Control procedures. Afterwards the "synthetic" counterpart of the actual unit is created, made up of a weighted combination of control units from the "donor pool". These fitting and weighting calculations are done through numerical optimisation procedures. These will be done by the package chosen to run the analysis.

Initiate outcome analysis – compare the postintervention data of the treated unit and its "synthetic unit".

Run any robustness checks – placebo analysis is frequently used – this involves performing the analysis as if other units in the donor pool were the treated Unit, to generate a distribution of effect estimates.

Packages for Synthetic Controls are now available in many software/languages, including Stata, Matlab, R and Python. The **Midlands DSC Analytics Department** can provide support.

Requirements of Synthetic Controls

For synthetic control methodology to be valid, several key requirements must be met.

Potential control units in the donor pool are usually selected for similarity in terms of geographical proximity, or with similar characteristics, whether that is beyond the matching variables. Subjective assessment of which units to include in donor pool can also be made, driven by your theory.

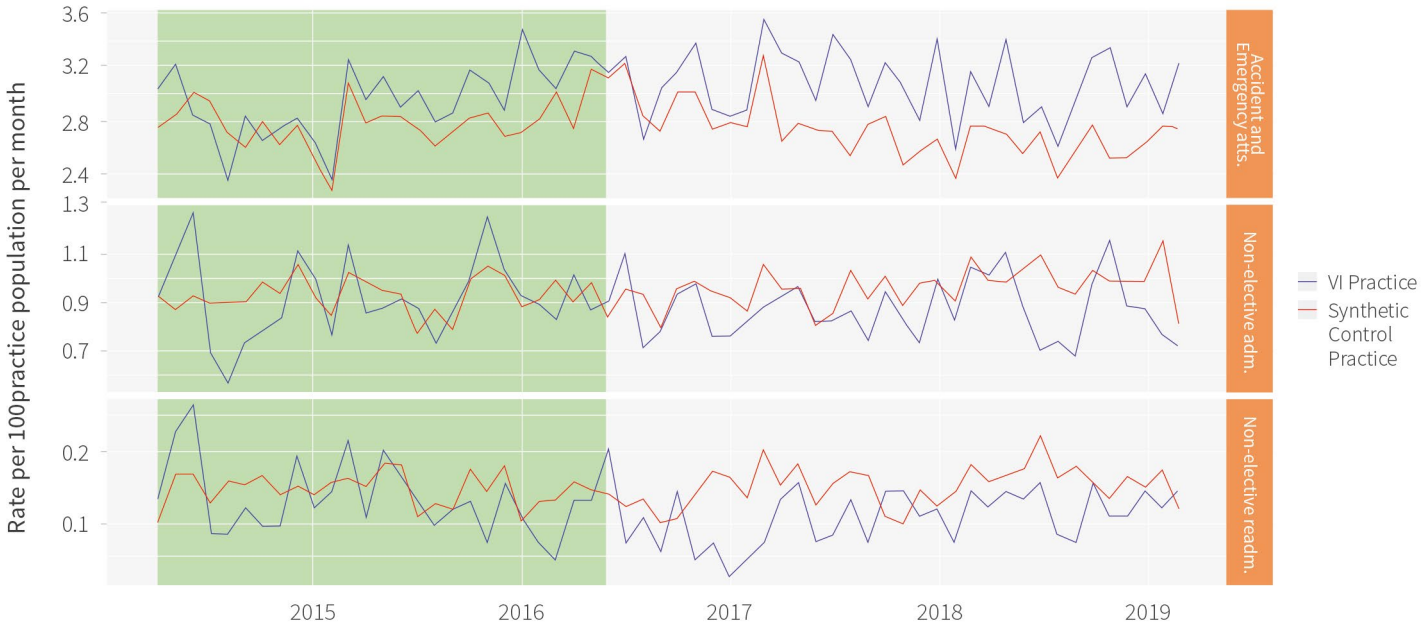
There should be no "contamination" of effects of the intervention into potential control areas. Any positive or negative impact of the intervention on actual units should not have an indirect impact on the outcome for control units.

No events that might differentially affect the outcome of interest in the treated or potential control units in both pre and post-intervention periods. In addition, no anticipatory effects. - untreated units should not be expected to receive the intervention in the future and begin changing their behaviour prior.

Examples in Healthcare Evaluations

At the Strategy Unit, we recently completed an evaluation for the Royal Wolverhampton Trust, examining the effects of their Vertical Integration programme through synthetic controls. This involved testing the GP practices involved in the pilot scheme, through creating synthetic counterparts from geographically adjacent GP practices not involved with the programme as the donor pool.

Impact or Vertical Integration on the Chosen Outcome Metrics



For one of the GP practices involved in the intervention (blue line), this plot shows the time-series of the data for our three outcome variables, for pre (shaded green) and post-intervention periods, along with its synthetic counterpart (red line).

The components of the red line would compose of several practices inside the donor pool, as seen on the left, weighted based on optimisation procedures.

Matched GP Practices that make up the Synthetic Control

Synthetic GP	CCG	Weighting
1	Birmingham and Solihull	0.301
2	Birmingham and Solihull	0.241
3	Birmingham and Solihull	0.198
4	Birmingham and Solihull	0.158
5	Sandwell and West Birmingham	0.101

Possible limitations

Data availability needs to be consistent across both the treated and control units. Incomplete or missing data can bias the results.

As Synthetic Controls is still an emerging method, comparisons with other difference-in-difference methods show that it is not a perfect solution for all settings. Careful checking of assumptions, consideration of other statistical techniques, availability of data, all need to be considered when deciding what is the best analytical tool to use.

Regression Discontinuity Design (RDD)

Regression discontinuity design (RDD) is a quasi-experimental approach which is used when the assignment to treatment or programme participation depends on one of the independent variables crossing a specific threshold. For example, we want to estimate the impact of drug A on the risk of Stroke and we know that drug A is prescribed for people with high levels of cholesterol. We also know that higher cholesterol increases the risk of stroke, therefore, those who are taking drug A have a higher risk of stroke regardless of the effectiveness of the drug. As assignment of treatment is not random, a quasi-experimental design needs to be used.

When should RDD be used?

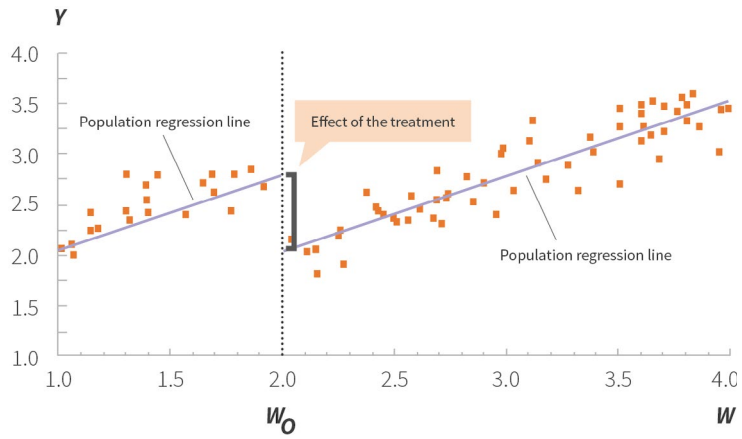
In situations when the assignment to the treatment or control group depends on clear thresholds – as in the example above, the groups are not comparable and, as a result, the results are likely to be biased due to unobserved factors.

How to use RDD?

The idea of RDD is to compare the group which is just above the threshold and the group which is just below the threshold. As long as the relationship between the threshold and the independent variable is discrete (e.g. cholesterol level and assignment of drugs) and the relationship between a variable of threshold and an outcome variable are continuous (e.g. cholesterol and the risk of stroke), considering observations within a small interval around the threshold is similar to having a randomized assignment.

There are two types of RD: sharp RD (all units above the threshold are participating in the programme) and fuzzy RD (crossing the threshold influences the probability of being in the treatment group, but there are additional determinants). The regression model for the sharp RD includes a dummy variable of a programme participation, a variable which determinates threshold and a set of control variables. The ordinary least square method can be used to estimate the coefficient in front of a programme participation, which is the indicator of the effectiveness of the programme. The picture opposite is adapted from **Stock and Watson (2015)** and shows the effect of the treatment on the outcome variable.

Example RDD Analysis



In the case of fuzzy RD, there are other unobserved factors which are likely to affect the programme participation. To avoid model error, instrumental variables can be used. In the example of cholesterol and drug A, we need to find an instrumental variable which affects the probability to be assigned to have a drug (relevant instrument) but does not affect the risk of stroke (exogeneous instrument). If such an instrument can be found, the programme participation variable can be replaced by its estimation.

Possible limitations

One of the main limitations of the RDD is the number of observations near cut-off point, which can be low. Moreover, some authors argue that extrapolation of the results to the general population might require more assumptions, because RDD is considering only a very limited cohort of the sample with characteristics close to threshold. There is also a risk that the assignment score around the cut-off will be manipulated to encourage more people be in the treatment group.

How do I assess costs and savings?

- What are the different types of economic evaluation?
- How do I evaluate costs and cost savings?
- Publicly available sources of cost data

When you have identified the outcomes that have been achieved – from the programme data linked to your logic model and, where appropriate, a quantitative impact analysis – you can consider the costs and cost consequences of the programme or intervention. This is usually referred to as economic, or value for money evaluation.

What are the different types of economic evaluation?

An economic evaluation is a comparative analysis of the costs and consequences (usually cost savings) associated with the intervention or programme. The consequences that you include in the analysis are the outcomes identified in your logic model, and a monetary value is assigned (for example, for inpatient bed days saved or for each appointment avoided). The costs are the monetary and other resources identified as inputs in the model (for example, programme funding or proportions of FTE posts). Economic evaluation draws on the **impact evaluation** analysis, which provides quantified evidence of the outcomes achieved. A monetary value is assigned to these outcomes.

There are different types of economic evaluations. You can find more detail on each of the approaches **here**.

Cost-benefit analysis

Cost-benefit analysis compares the costs of interventions and their consequences (health outcomes and others) – assigning monetary values to provide an assessment of costs and outcomes of the intervention compared to an alternative (the benefits).

Cost-effectiveness analysis

Cost-effectiveness analysis is a comparison of the costs of interventions and the value of their outcomes. It provides a cost per outcome rather than a comparison with a previous or similar intervention.

Cost-utility analysis

Cost-utility analysis is a type of cost-effectiveness analysis in which health and care interventions are compared in terms of the delivery of quality-adjusted life years (QALYs) outcomes. This outcome measure captures the length of life of a patient given a health condition and it is adjusted to reflect the quality of life. QALYs are a health outcome measure that can be used to evaluate any health intervention. Therefore, cost-utility analyses are particularly useful in resource allocation decision-making. The National Institute for Health and Care Excellence (NICE) uses this measure to decide whether an intervention is cost-effective.

Cost-consequences analysis

In a cost-consequence analysis, costs and outcome values are reported separately; an overall measure of efficiency is not produced.

Cost-of-illness studies

Cost of illness studies are an approach to understanding the economic burden of a specific disease to a society. The cost drivers are those for the healthcare system, for patients and other, indirect costs. Healthcare costs include diagnostics tests, drugs and management of disease by healthcare professional. Patients' costs include any of those that they incur to obtain the necessary treatment – travel time and expenditure for over-the-counter medication. Lastly, indirect costs include the loss of productivity due to time off sick. Because these studies do not compare healthcare outcomes and cost of different intervention, they are not considered to be economic evaluations. But they are useful for providing the rationale for a new intervention; or wider context for understanding cost-effectiveness or cost-consequence.

You can find more information **here** (from the CDC in the US, but providing a useful outline and worked example) and **here**.

Return-On-Investment analysis

Return-On-Investment (ROI) analysis compares in monetary values the benefits of one intervention and the costs of delivering that same intervention. The benefits accounted for are not only health gains but also costs saved (e.g. reduced admissions).

PHE's Health Economics and Modelling team (HEMT) from PHE have produced ROI tools in different areas such as falls prevention and musculoskeletal conditions, among others.

How do I evaluate costs and cost savings?

The type of costs included in an economic evaluation partly depends upon the type of analysis you undertake. The most common perspectives used are NHS, social care and a broader, societal perspective. An NHS perspective accounts for costs such as medicines and management of diseases; this includes GP visits, hospital admissions and administration, among others. A social care perspective includes social services and related costs. **NICE** recommends a perspective of 'NHS and personal and social services'. The societal perspective accounts for the impact of an intervention for the whole of society. This would include costs such as transportation for access to treatment, over the-counter purchases, co-payments, informal care time or time off work. **The Green Book** recommends the societal perspective for public sector interventions. The perspective you take will depend on the scale of the intervention and the focus of the evaluation. Single-funder initiatives are likely to be limited to the NHS or perhaps NHS and social care. Shared, system-wide initiatives may be more appropriate for a societal perspective.

There are several sources of information available that you can use to identify the value of the outcomes that have been delivered. Input costs, beyond programme/intervention funding, should be calculated from in-house resources (for example, average or specific salary costs).

Public Health England have a useful guide to comparing costs and outcomes of a range of different conditions and services **here**, as well as a range of supporting materials including summaries of economic evidence and these **Spend and Outcomes Tools (SPOT)** for local authorities and CCGs.

Publicly available sources of cost data

Hospital Costs

Hospital costs can be found in **Reference Costs**. This source is based on Healthcare Resource Groups (HRG), which are standard groups of clinically similar treatments and use comparable levels of healthcare resource, enabling you to account for the complexity of the patient's healthcare needs in monetary terms. This is a rich source of information that includes: inpatients costs (elective, non-elective, day case), critical care, outpatients and A&E, radiotherapy and chemotherapy, renal dialysis, spinal injuries, rehab, maternity, audiology, physiotherapy and dietetics, and many more.

Primary Care, Social Care and Community Costs

Personal Social Services Research Unit (PSSRU) produces the **Unit Costs of Health and Social Care** report every year. This rich source provides costs for: primary and social care staff contacts and hourly rates, hospital medical staff hourly rates, community-based staff hourly rates and costs across social and care including services for drug and alcohol misuse, people with learning disabilities and older people. **Reference Costs** offers costs data for community settings but is very limited.

Pharmaceutical Costs

British National Formulary offers costs and typical regimens for licensed drugs (generic and proprietary).

How do I design a survey?

- Designing surveys is not easy
- Using pre-existing surveys
- How to design your survey

Surveys can be good way to collect data from a larger group of people than is possible through qualitative methods. But there are some pitfalls to avoid. A survey is often seen as a simple way to get information from lots of people; but it is easy to do it badly and to create a lot of work to analyse the responses.

If you are thinking about putting together a survey, consider whether it is the right approach for the outcome you are trying to measure, or whether there is already an existing survey that you can use. If you do decide to go down the survey route, are you clear about your study objective? Do you know who the target population are? And do you propose to send the survey out to everybody or a sample of the population?

For example, if you want to learn about the experiences of BAME patients across a particular area, how would you go about forming the right questions to ask? How would you approach the population? And how general or specific will your population need to be?

Designing surveys is not easy

It is important to remember that surveys are a scientific method and not just a set of questions. A good definition of a survey (albeit not the only one) is provided in this (huge, detailed) **handbook** on survey methodology:

> *"a research strategy in which quantitative information is systematically collected from a relatively large sample taken from a population".*

Though they are often seen as an easy option to quickly and cost-effectively gather information, surveys can also present problems such as lack of depth, generalisability, and response rate. When using a survey you need to think about the population you want to collect data from, what a sample of that population might look like, what questions you need to ask and how these should be worded to get the information that you need, how you are going to distribute the survey (and how this might encourage or act as a barrier to participation) and how you are going to analyse the results – including taking account of who has responded.

Having said this, a survey can be a great way of collecting information - when it is well thought through, piloted so that you can test whether or not it works as you intend, and you have planned your resources for analysing the results.

Using pre-existing surveys

In some instances, large pre-existing surveys are already widely in circulation, are proven to be reliable measures and capture the necessary data with a large enough sample size suitable for evaluation; or there are surveys that have already been carried out in your area, which can be repeated or adapted.

Using pre-existing surveys is cheaper than collecting new data. But using an off-the-shelf survey isn't always possible. Is there a survey out there that meets your needs? Is it high quality? Is it collecting data for the same or similar purpose? Think carefully about whether any existing survey is available and whether it will be an appropriate source of data or data collection tool for your research questions.

There are a wide range of validated questionnaires – tools that are proven to deliver reliable results. Some require payment. Free measures include The Warwick-Edinburgh Mental Wellbeing Scale **(WEMWBS)**; the Health and Safety Executive's Management Standards Indicator Tool for **work-related stress**; the General Practice Physical Activity Questionnaire **(GPPAQ)** and the NHS Patient Activation Measure **(PAM)** for people with Long Term Conditions.

This **guide to measures of patient experience and patient outcomes** (produced by The Strategy Unit for the Dudley MCP Vanguard) provides an introduction and overview of a wide range of validated measures and their advantages and disadvantages.

How to design your survey

These sections provide some some points for consideration if you are planning on implementing a survey. The list is a non-exhaustive, and where possible links to publicly available resources have been provided for more reading.

- **The survey questions**
- **Survey format**
- **Sampling**
- **Analysing your results**

The survey questions

Before creating a set of questions, think carefully about which of your research objectives and questions you are going to answer. There are two types of question (the people who complete the survey are called 'respondents'):

- Open questions – where the respondent is free to provide any answer they wish
- Closed questions – where the respondent chooses from different options (multiple choice).

Take time thinking about how you can structure your survey questions to collect the data that you need. Who are you collecting the data from – will they understand the question? Does it use technical terms or abbreviations? Is it in appropriate language? The wording you use can make it inaccessible for some sections of the community to understand, for example for participants who speak little English.

Avoid using too many open questions. They create a lot of work for your analysis. If you don't know much about people's views on the topic and thus how to create closed questions, do some qualitative work to explore the issues in depth with a small group of people and this will provide you with the range of views that you can turn into the different options (your multiple choices) for your survey question.

When using closed questions, think carefully about the options you are going to present. Is there any qualitative work that has taken place with a small group of your target population, which you can draw on to see what potential answers might be? Is there any other research or literature out there on the topic? Do you have a narrow set of options that you are seeking to explore views on?

Getting your questions right takes time. Involve a range of people who are familiar with the topic and target population – colleagues, stakeholders, and/or members of your Steering Group – to refine them. And then test or pilot the survey with a small group of your target population to see if they work in the way that you intended.

If you are not experienced in writing survey questions (or even if you are) it is worth looking at these **resources** from the UK civil service for questions in health and care. The resources are part of a 'harmonisation' initiative, which aims to provide more robust survey data through common, harmonised, approaches. There are clearly written, tested questions available that you may find useful (for instance, on **general health**, and **long term conditions**).

The **Magenta book** sets out four rules that are worth keeping in mind when designing your questions

- Can the respondents understand the question?
- Do they understand it in the same way that you do?
- Are respondents able to answer the question?
- Are they willing to answer the question?

Along with question wording, think about the flow of the questionnaire.

There are a range of widely used online tools such as **Survey Monkey** that include **guidance** on designing your survey (and analysing the results).

Survey format

There are lots of tools available online that help you design a survey that's right for you. Keep in mind that not everyone has access to the internet, or to smart phones so it may be suitable for you to send a hard copy out in the post, or to make hard copies available in settings where your target population will be – for instance, GP practices or community centres.

Sampling

A key consideration for any survey is of course who you are going to ask to complete the survey. It may be that you want to distribute the survey as widely as possible amongst the target population – all the patients of a practice or the users or a service – or it may be that you want to distribute it to a sample of the target population – all residents of a particular area.

Sampling is a complicated, technical area. Sampling methods fall into two broad categories: probability sampling and nonprobability sampling.

Probability sampling methods involve selecting respondents at random from a sampling frame (i.e. a list of all respondents in the population of interest). Ultimately, you need to avoid "coverage error", of which the most common form is under coverage. Coverage error occurs when a part of the population is systematically omitted (under coverage) or over-represented (over coverage) in a sample. For example, if a GP sent out a survey to a sample of all her patients (the population), but only picked the top 200 from an alphabetically ordered list, that would systematically exclude everyone with surnames beginning with letters further down the alphabet.

Non-probability sampling methods include quota sampling – seeking out particular quotas for different types of respondents – and convenience samples – those people that are easy to reach.

The main methods for each of the two types of sampling method are described briefly below. For more detailed reading, see the **Magenta Book** (section 4).

Analysing your results

For any open questions you have used, your analysis should follow the principles of qualitative analysis described in this **section** of the guide.

For closed questions, you will be able to produce charts that show the number of responses for each question and the types of respondent for each answer – how many women or men, people of different ages or with different conditions, for example, chose each option.

It is essential that you take account of who your respondents are. If you have used a convenience sample, look at the spread of your respondents across different groups. If you have used a sampling approach, then you can apply statistical techniques to establish the statistical significance – the confidence you can have in the links between your respondents and their answers, for example that men are more likely to think X – of your findings.

We recommend that you use a tool such as **Survey Monkey** to conduct and analyse your survey, because they include a wide range of **easy to use approaches** to understanding your data and findings.

How do I report my findings?

- Understanding the audience
- Different outputs for different stakeholders
- Triangulation
- Executive summary
- Writing in an accessible style
- Using diagrams and pictures helps communicate your findings
- Make sure you include full details of your method
- Include quality assurance of your final outputs

Earlier in the guide we explored the different purposes of evaluation – providing formative learning (informing delivery); and summative learning (about outcomes achieved). The way in which you report your findings will be informed by these purposes. Formative findings should be shorter, presenting the evidence to date and focused on recommendations for implementation. Summative findings should be fuller, providing a narrative of implementation over time and providing conclusions about effectiveness and recommendations for future delivery.

Reporting your findings in an accessible way:

- Builds awareness of the programme
- Demonstrates transparency and accountability of the evaluation
- Communicates value of the programme to commissioners/funders
- Shares good practice with external stakeholders
- Provides findings that facilitate change and improvement.

Evaluation reports should address the three core aspects of any evaluation – what works, for whom, and why?

- Describe the theory of change (using the logic model to illustrate) – how the purpose and activities are intended to deliver the intended outcomes.
- Describe the implementation of the intervention of programme, using your qualitative data, including anything that changed and why, and the outcomes from monitoring information and your quantitative analysis.
- Discuss the learning from the analysis in a conclusion – what have been the key drivers of or barriers to success? What are the conditions for success? What are your recommendations for implementation, further delivery or investment decisions? It may also be appropriate to present a final logic model of how the intervention or programme worked in practice, as a model for replication or rollout at scale.

Understanding the audience

Earlier in the guide we discussed the importance of **working with stakeholders** in developing the evaluation and involving them in your steering group. They way in which you structure and present your reporting should be informed by the needs of the audience – as discussed with your stakeholders. Who will read the report? Who will use the report?

Think about:

- What kind of information is required, e.g. should the emphasis in the reporting be more on what was found or what was done?
- What format would meet the audiences' needs – a short slide-pack of headline findings, or a fuller and more discursive Word document?
- How will the information provided in the report be used by the reader? For example, it is likely that the recommendations (from a credible evaluation) may be directly actioned by the programme stakeholders, citing the evaluation report. The findings will need to be clearly evidenced.

As the evaluation stakeholders are usually the primary audience, it is good practice to agree the report format at an early stage. It is also good practice to involve wider stakeholders and beneficiaries at other points of the reporting process. For example, report drafts can be shared to enlist stakeholders in co-production of the findings and development of recommendations. One note of caution though – conclusions and recommendations need to be grounded in the data and not biased by individual interests, so work for a consensus about what the findings mean.

Different outputs for different stakeholders

Due to the range of different internal and external stakeholders it is likely that you will need to report in different formats for different audiences – both a fuller report (in Word) and a summary of findings (in a slide-pack format). Think about developing a (short) communication and dissemination plan as part of the planning process. This will help identify the type and timing of the outputs, matched to the different stakeholder groups. For instance, there may be a particular Board or other decision-making group that require an output for a particular date.

The communication and dissemination phase at the end of the evaluation is a two-way process intended to support use of the evaluation findings for improvement and decision making.

Think about:

- How you can use a range of dissemination methods, including written information, electronic media, and workshops.
- Developing clear channels for stakeholders to feedback and interact with the outputs
- Drawing upon existing dissemination resources, relationships, and networks to the maximum extent possible while building new resources as needed by users.
- Provide linkages to resources that may be needed to implement the information.

Triangulation

In reporting your findings you are both presenting the data and interpreting what this means – in the context for the programme and your findings, and the needs of your audience. The process of triangulation is the drawing together of different sources of evidence to provide robust conclusions. This is why we recommend mixed-method evaluations.

This **paper** from the BMJ discussed triangulation in more detail. In summary, triangulation of findings can bring three important benefits:

- The results may converge and lead to the same conclusions, this increases the validity of the finding through verification by different analyses
- The findings may be complementary to each other but highlight different aspects, thereby providing a more holistic insight
- The results may be divergent or contradictory, this can lead to new and better explanations and a revised programme theory.

Executive summary

This is a short section that is usually found at the beginning of the final evaluation report. Its purpose is to provide the reader with an outline of the report's content without having to read the entire document. It can also serve as a standalone document, sharing key aspects of the evaluation with a larger external audience.

The exact length and components of an executive summary can vary but it is typically 1-3 pages in length, provides an overview of the programme and evaluation, summarises findings and provides key recommendations. It does not include technical details about data collection methods used but may include a figure that illustrates a key finding or a table of the main recommendations.

Writing in an accessible style

Writing your report in a clear, accessible style will support wide engagement with your findings.

The Plain English Campaign provide helpful **guidance** on key principles to apply, including:

- Keeping your sentences short
- Using active verbs
- Using 'you' and 'we'
- Using words that are appropriate for the reader
- Don't be afraid to give instructions
- Avoiding nominalisations (using a verb as a noun)
- Using lists where appropriate

Using diagrams and pictures helps communicate your findings

Think about how you can present your evaluation findings in different ways. Tables and charts can help to present findings in an accessible way, and give more impact. You might have access to infographics or communications specialists; talk to them about their ideas even if there is no resource to involve them directly.

In the main body of your report, there are some key ways in which using visuals can help illustrate your findings.

- Quantitative data, such as survey findings or the results of modelling are often best presented in a table, chart or other diagram to communicate often complex messages – or a range of different data – simply and clearly.

There is a guide to presenting numerical data **here** (ironically, quite wordy but it is fairly accessible to the non-specialist).

- Qualitative data or key messages from mixed method evaluation can be communicated clearly using infographics. Infographics are used to tell a story, or a range of different key messages together in one place. However, they are likely to require a specialist or (a small amount) of resources if they're to be done properly.

There is a short guide with some examples **here**.

- Your report can also be illustrated with pictures, for instance photos of places or people (with their written consent). If using pictures that you haven't taken yourself or that aren't provided by research participants (for example, maps or commercial posters), check that you have permission.
- As well as simple tables and charts there are other simple techniques such as using a roadmap to show a journey, or a simple timeline.

Make sure you include full details of your method

It is essential that you include the detail of your method in your report, so that everyone is able to see how you arrived at your findings – and thus have confidence in them – and so that they can be learnt from for future projects. For quantitative work, the methods may be replicated. For qualitative work, they may provide learning about engaging particular groups. It is important that you are transparent in describing what you did and why; what went well and what, if anything, didn't work or go to plan so that any limitations on your findings are clear. For instance, if you do not achieve a good balance of ethnicity across your participants, then there will gaps in what you have found out and the conclusions that you reach.

You shouldn't include full details of your method in the full report. Provide the key information about who participated and the data analysis that was undertaken. Remember to include any changes that you made from the original design, and why.

A simple structure for presenting your methodology in the report would be:

- Aims and objectives
- Research questions
- Methods
  - > Qualitative – including a breakdown of participants
  - > Quantitative – including an outline or description of the key techniques or approaches applied
- Limitations – any reflections on the limitations of the study (for example, any issues with participant recruitment or the quality of the quantitative data).

An Annex can be used to provide full details. Include copies of topic guides used in qualitative interviews, your logic model and a fuller set of charts from data analysis than you will in the main body of the report. Setting out this detail means that the evaluation can be peer reviewed, and other evaluators (and members of the DSU network) can learn from the work you've done. It also means that, when provided in an Annex, the focus of the report is the discussion of the evaluation findings.

- For any qualitative work include:
  - > A breakdown of the participants (but not so that they can be identified – describe the number of participants for each broad group, e.g. Patients, Managers, Senior Leaders, Ward Nurses, GPs, etc.), and how they were engaged
  - > Your approach to data analysis
  - > Your research tools – interview and group topic guides; **consent and participant information sheets**.
- For any quantitative work include:
  - > A description of the datasets and your approach to information governance
  - > The formulas used and modelling
  - > Fuller analysis than can be included in the main report – the tables and charts that there is not room for or that are not appropriate for the main discussion but that provide supporting information.

You might find this **guide** to evaluation reporting (from The Center for Disease Control (CDC)) helpful if you would like further reading.

Include quality assurance of your final outputs

Earlier, we described the importance of including **good governance of your evaluation** and how this provides quality assurance during design, development and delivery. The arrangements should also provide for quality assurance of your final outputs, to ensure that they meet the requirements of your audience.

Before you submit your report for this review, you should invite review from someone from outside of the day-to-day delivery team to quality assure the report. This might be a senior manager or someone else with knowledge of the research methods you have employed. Or it could be someone external, including someone from **the DSC**, who can act as a critical friend. The quality assurance review should include both a technical review of the methods employed, and review of the narrative including proof reading.

Make sure that you build time for quality assurance into your workplan.

Where can I get further support?

- Midlands Decision Support Centre

Midlands DSU Network
Decision Support Centre

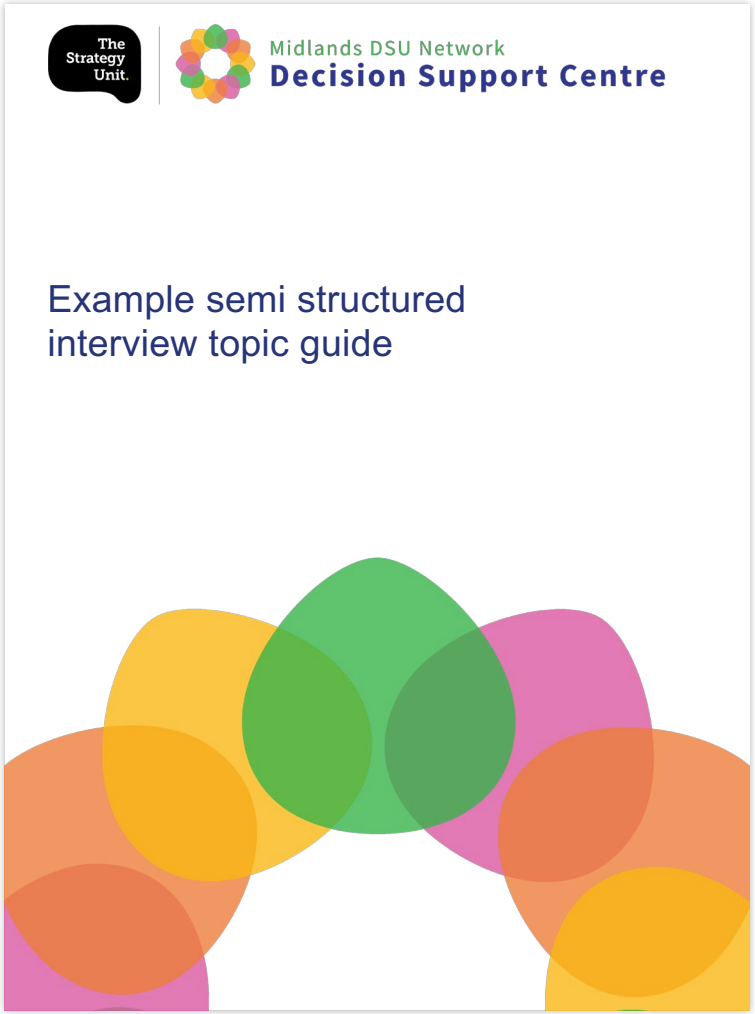
The Midlands DSC Evaluation Department that has produced this guide can provide members of the DSU Network with a range of further support:

- Discussion of, and expert guidance on, any aspect of this guide and what it means for your evaluation – planned or in progress
- Training – either bespoke or from our development programme
- The Midlands DSU Evaluation Network brings together those practicing and interested in evaluations, to share resources and learning – including facilitating joint-DSU evaluations
- Quality Assurance of evaluation tools, methods, analysis and reporting
- Support with developing evaluation briefs and commissioning external evaluation
- Providing a blended DSU and DSC evaluation team (working together to conduct the evaluation)
- Providing an independent evaluation as an external provider.

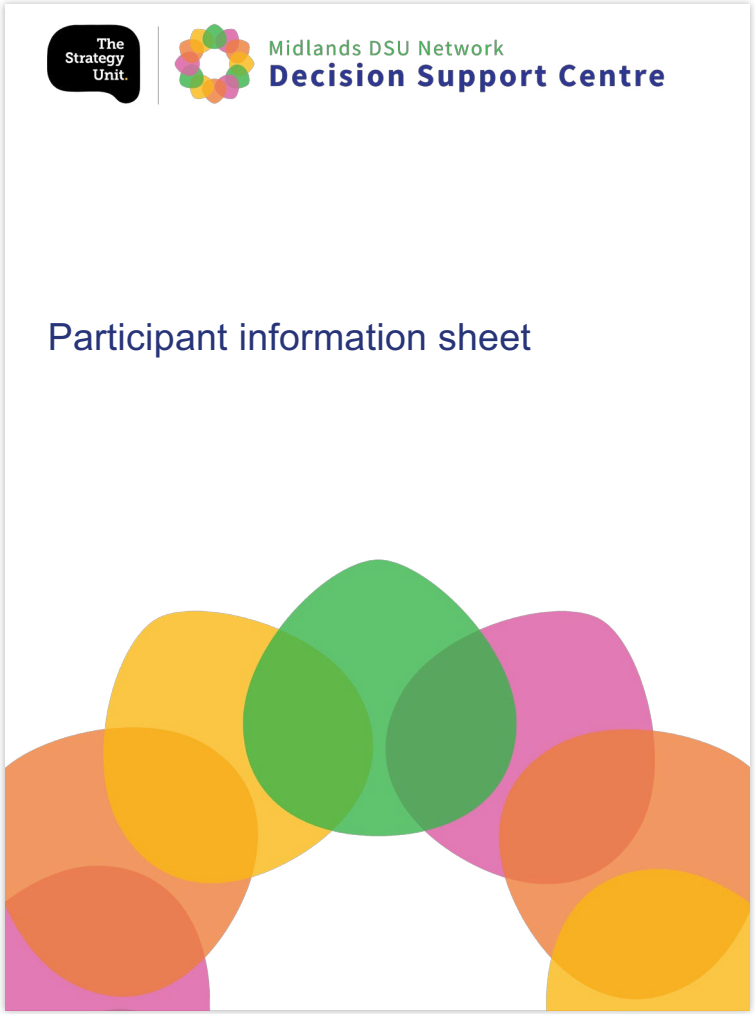
Supporting resources

- Example semi structured interview topic guide
- Participant information sheet

Example semi structured interview topic guide



Participant information sheet



Prepared by:

Paul Mason

Abeda Mulla

Anastasiia Zharinova

Gillian Cope

Mahmoda Begum

Mike Woodall

Shammas Rahim

Beatriz Rubio-Huete

Victor Yu

Zaheera Teladia