Public Service Transformation: Introductory guide to evaluation

March 2014
Public Service Transformation Network

This document/publication is also available on our website at http://publicservicetransformation.org/

If you have any enquiries regarding this publication, please contact us at
http://publicservicetransformation.org/contact

For all our latest news and updates follow us on Twitter: https://**twitter**.com/**Service**Reform

March 2014

Contents

Preface

Transforming services to deliver better outcomes for the people who use them is the most important challenge facing the public sector today. In doing so, we must also redesign services in a way that improves value for money for the taxpayer. That will require better engagement with service users to understand what they value most and better replication of best practice so that we learn from what works.

Proper evaluation is an essential business tool. Public service providers should think about evaluation from the outset when redesigning services. Effective evaluation provides powerful insights into why some approaches are more successful than others and how to deliver better services within constrained budgets. Evaluation provides decision-makers with the evidence they need to support new ways of providing services and to make the best possible case for securing the resources needed to deliver them.

The Public Service Transformation Network has produced this 'Introductory Guide to Evaluation' following broad consultation with local service providers and central government. The guide has been developed to support public service providers engaged in transforming and redesigning services. It also provides information about general support and specific resources that can be accessed to support better evaluation of service transformation.

We would like to thank all those who helped to produce this guidance. It will be updated based on feedback about how it can be improved and more material added as and when there is a need and opportunity.

Mohammed Aziz
Senior Policy Adviser
Public Service Transformation Network

What is this guide about and who is it for?

This guide is for anyone involved in evaluating the effectiveness of work to transform public services. It has been developed by the Public Service Transformation Network in consultation with central government and local service providers.

The guide sets out some guiding principles and key considerations to help local providers to measure the impact of changes to services. It is not intended to provide a comprehensive evaluation framework for use in all contexts.

Effective evaluation will enable local providers to ensure the highest possible quality, credibility and accountability for services that have been redesigned. It will also enable multi-agency partnerships delivering new services to demonstrate value for money and how redesigned services will make a real difference to local communities.

Guiding principles for the evaluation of service transformation should be that evaluation is:

- Locally led.
- Proportionate to the resources expended, outcomes expected, scale of change and associated risks.
- Robust enough to provide assurance to key partners, locally and nationally.

The NAO report "**Case study on integration: Measuring the costs and benefits of Whole-Place Community Budgets**", noted that:

> *"While projects can be appraised before implementation it takes time for their impact to be established in practice, so there needs to be a strong commitment to monitoring and evaluation over the longer term."*

The information contained in this guide complements additional guidance on cost benefit analysis: Supporting Public Service Transformation: Cost Benefit Analysis Guidance for Local Partnerships.

The guide is split into 5 sections:	
1	**Introduction to evaluation** Explores some of the basic considerations and principles underpinning evaluation.
2	**Undertaking the evaluation** Sets out some practical issues to consider when planning and undertaking an evaluation.
3	**Common pitfalls** Identifies some of the common pitfalls and how these can be avoided.
4	**Externally commissioned evaluations** Provides advice to get the most out of externally commissioned evaluations.
5	**Resources and reference materials** Includes a compendium of evaluation reference material and resources.

Part 1: Introduction to Evaluation

1.1 What is evaluation?

In simple terms evaluation is about answering the questions: how do you know what worked and where is the evidence? Proper evaluation is an important tool to determine the merit of an activity or approach and provides insights into the value of a new approach, concept or way of working. In the context of service transformation, evaluation of a proposed new or changed service can provide the evidence to demonstrate whether the service represents value for money and delivering the expected outcomes.

The Magenta Book, central government's core resource on evaluation methods, defines evaluation as follows:

> *"Evaluation examines the actual implementation and impacts of a policy to assess whether the anticipated effects, costs and benefits were in fact realised. Evaluation findings can identify "what works", where problems arise, highlight good practice, identify unintended consequences or unanticipated results and demonstrate value for money, and hence can be fed back into the appraisal process to improve future decision-making."*
>
> Magenta Book, (page 7, paragraph 1.3)

Evaluation is about monitoring and collecting the right information to make judgements on the quality, value or importance of the service transformation project or programme. Process evaluation is generally most relevant during the implementation of service change or soon afterwards, when assessing changes against specific success criteria, whereas an impact evaluation is most relevant over a longer timescale. Evaluators can use a range of methods, including qualitative and quantitative to assess the process, impact, costs and benefits of different service designs.

Why conduct an evaluation?

Any service redesign will normally have been informed by an initial assessment of the likely costs and benefits. The analysis of costs and benefits associated with this type of initiative should inform the evaluation. Cost benefit analysis may also illustrate what data may already be available.

Evaluation is also a key tool for assessing whether changes deliver the expected outcomes and benefits with the assigned resources. This should include an assessment of whether the beneficiaries are those that were originally intended and the identification of any unintended consequences.

Generally evaluations assist in understanding the consequences of changes to services by:

- Enabling both local and national partners to measure the economic and social impact of re-designed services.
- Monitoring progress in the implementation of proposals against projected savings targets, outcomes, timescales and milestones and understanding what elements of service implementation may be working or not working.
- Tracking emerging outcomes against locally and centrally recognised measurements.

- Generating ideas for improving the service or for improving future service transformation through better co-design, consultation, implementation, cost-effectiveness, and management structures.
- Providing accountability to investors and stakeholders as to the value of the change or intervention - including tangible ‘proof’ of success and value for money to continue supporting a redesigned service.
- Helping decision-makers at all levels to make better choices about service provision, objectives, and the allocation of resources.

For instance, the table below indicates some of the key questions an evaluation can help to answer. This will depend on the type of evaluation undertaken (see section 1.4):

What were the achievements of change?	**Why did things work or fail?**
▪ Did the change do what it was supposed to do? ▪ What resources were used? ▪ How did activities undertaken meet the goals of the service transformation?	▪ What were the reasons for the successes and failures? ▪ How realistic were the goals and objectives? ▪ Were there any flawed assumptions?
Has all of the change made a difference?	**What could be done differently in future?**
▪ How did the change impact on knowledge, attitudes, skills, and behaviour? ▪ How did the change affect the target population, the wider community, participants, providers and others? ▪ Who benefitted from the change and in what ways?	▪ What more effective interventions could be used? ▪ How were organisational systems/processes adapted during the course of the change to become more effective? ▪ How could the evaluation findings be used to contribute to future service changes?

Table: 1

1.2 Measuring success

To undertake an effective evaluation, data and information on the impact of the new or transformed service will need to be analysed. A good evaluation will focus on assessing whether the initiative or project met the success criteria and whether the change can be attributed to that project or initiative. The success criteria will probably be defined by stakeholders and should align with a project, programme or other management approach.

Without clear success criteria it is difficult to focus the evaluation and determine whether an initiative was successful. Success criteria will inform the type of data and information used for an evaluation. For example, a new service design which aims to reduce demand will need to measure the degree to which the number of service users has reduced as a result.

Another approach may be to manage multiple services through a single hub. This will have cost implications in terms of buildings and resources as well as requiring some information on whether this is seen as an improvement from the perspective of the service user. This should be supplemented by objective data, for example, whether the time demand on different services has been reduced. This is because user surveys alone are not always considered a reliable source of evidence. However, surveys on their own can still provide valuable insights.

Box 1: An example of an interim evaluation measuring success

Care Package Facilitator, Swindon Borough Council

The 'Care Package Facilitator' is a new post appointed by Swindon Borough Council to help reduce the number of children and young people at risk of going into high cost care placements by providing innovative personalised care support packages. The main objectives and success criteria for the post are:

- Reduction in costs through savings and demand mitigation.
- The number of children and young people given suitable alternative support.

An interim evaluation was commissioned four months into the facilitator being in post to provide an independent assessment of the success of the initiative. The key evaluation questions were:

- Was the project achieving the desired outcomes? If so, why and how?
- What were the estimated cost savings/avoidance so far?
- Is it possible to scale up the project? If so, how?
- How can evaluation be built into the remainder of the project?

The evaluation approach included a mix of process and impact evaluation. A review of existing project data and trends in demand provided quantitative analysis, while in-depth interviews with stakeholders provided qualitative analysis.

Trends in demand were assessed by analysing historical numbers of children in care, combined with analysis of the number of placements budgeted for 2013/14, versus the number of cases that the post-holder has worked with. This gave an indication of the potential number of cases with which the facilitator worker could work in the immediate term and an indication of how caseload might increase in the future.

A key challenge was a lack of evidence regarding the success of the project to date, due to its relative infancy, and whether it was possible to clearly attribute successes to the facilitator post-holder. This was overcome by using in-depth interviews which asked individuals if their situation might have been different without the intervention.

The evaluation found the work to be successful with considerable in-year cost avoidance for the authority of approximately £0.5m. Key recommendations from the evaluation were used to inform the future work of the facilitator, and Swindon's Children and Families service. These included:

- Extending the facilitator's's contract.
- Continuing to monitor the estimated cost avoidance and starting to monitor the number of aggregate care weeks saved compared with the baseline.
- Introducing a clearer case referral process as many children had initially been referred on an ad-hoc basis.
- Further monitoring of impacts and a clearer approach to establish attribution to the activity of the facilitator to determine scalability of the work across the service.
- Using the Transformation Network's cost benefit analysis model to monitor future cost avoidance.

Enquiries to Sarah Hardwick, Evaluation Lead: shardwick@swindon.gov.uk

1.3 The evaluation process: key stages

The table below details the key steps to take when performing a typical evaluation. Some of the actions may occur concurrently depending on the skills and resources available. More detail on each of the steps can be found in Part 2: Undertaking the evaluation.

Steps	What is involved
Evaluation scoping	▪ Establish and agree the purpose of the evaluation: what you want to know? ▪ Identify and meet with stakeholders and investors to clarify the evaluation's purpose(s), goals, objectives and expected use. ▪ Will the evaluation be externally commissioned or conducted in-house? ▪ Identify a budget for the evaluation.
Evaluation preparation	▪ Develop an agreed and shared logic model describing the connection between the activities of the service and the expected results. ▪ Identify indicators and kinds of information which can be used to measure the intended changes and benefits. ▪ Identify the type of evaluation to be used based on the aims of the evaluation, budget and resources. ▪ Think about appropriate but propotionate sample size and whether control group/s will need to be set up. ▪ Develop specific evaluation tools and methods centred on the evaluation questions so that outcomes and impacts are attributable to the change initiative. ▪ Identify potential participants and sources of information. ▪ Allocate data collection tasks and begin monitoring processes.
Monitoring and Data Collection	▪ Identify data sources, and costs associated with different approaches. This should include baseline and trend data. ▪ Information collection needs to be practical, systematic, ongoing, accurate and ethical. ▪ Use initial findings to recommend changes to service design and implementation. ▪ Review and refine data collection tool and methods.
Analyse and interpret the data	▪ Summarise and analyse data as it is collected; summaries made during the evaluation can save work at the end. ▪ Identify key patterns, themes and issues by analysing and aggregating data. ▪ Share and develop findings iteratively amongst the evaluation team, including team members responsible for different methods. ▪ Identify what worked and what didn't, why and in what circumstances. ▪ Prepare a summary or top-line report of findings.
Communicate the results	▪ The final report should be practical, easy to understand and include strategies for future success. ▪ The report will need to factor in the interests of the audience e.g. investors, partners and users.

Table: 2

1.4 Types of evaluation

Evaluations can have different purposes at different stages in the service delivery cycle. Most evaluations will use a combination of:

- Process evaluation – a means of tracking progress and the effectiveness or appropriateness of the delivery process.
- Impact evaluation – an assessment of the difference a service change has made against agreed outcomes.

Process evaluation is usually more straightforward since it is mainly a question of gathering views and information on how well the delivery has been carried out and any lessons arising.

Impact evaluation is more complex depending on the innovative nature of the service change and may require the application of experimental or quasi-experimental assessment methods. This is discussed in more detail in part 2 of this guide.

The table below sets out more detail on what is likely to be involved under each type of evaluation:

Evaluation type	Purpose	Stages of service change
Process evaluation	**A process evaluation typically studies the development of the change and the looks at activities and processes that took place in designing and implementing change:** ▪ **Targeted at improving the structure, management, or implementation of service change.** ▪ **Engaging with stakeholders and feeding back findings in real time.** ▪ **Helps to explain why certain outcomes might have occurred.** ▪ **Use findings to inform scoping of impact evaluation and inform decisions on whether longer term evaluation is practical.**	**Development and Design** **Implementation and Testing**
Impact evaluation	**Asks questions about the impact of a specified change on a specific group of people. It should seek to assess how the impact compares to the original objectives, or to some other programme, or to doing nothing at all:** ▪ **Measuring the changes (short term or long term impacts/outcomes) brought about by change, including positive/negative and intended/unintended effects.** ▪ **Impact evaluation makes an overall judgement about the value of change.**	**Delivery and Scaling up**

Table: 3

Part 2: Undertaking the Evaluation

2.1 Evaluation scoping

Defining the scope of an evaluation amounts to asking the question:

What is going to be evaluated?

This is a vital part of planning any evaluation because without being clear about aims and objectives it will be an endless task to look at every aspect of a change or series of changes. When answering this question, it is important to determine which of the change or series of change processes and other evaluation issues need to be explored and which need to be examined in more detail. It is normally helpful to adopt a relatively strict definition of the scope of the evaluation.

What can be achieved by the evaluation will be influenced by the goal and by the availability of data, time constraints and available resources. In order to reach conclusions, the evaluation should be confined to an examination of the change or series of changes and their most essential interdependencies with other public policies.

Stakeholder engagement and partnerships

Evaluation activity for local service transformation should be locally led and co-designed in collaboration with a range of partner organisations. All partners should understand exactly what the evaluation will measure and how, so that they are comfortable with the level of evidence it will provide. This is particularly important if partners are being asked to invest in a service redesign based on the outputs of the evaluation.

Information sharing

As part of the monitoring regime local providers should consider and plan for information sharing across a range of stakeholders including delivery partners, providers and sponsors. Some common challenges include:

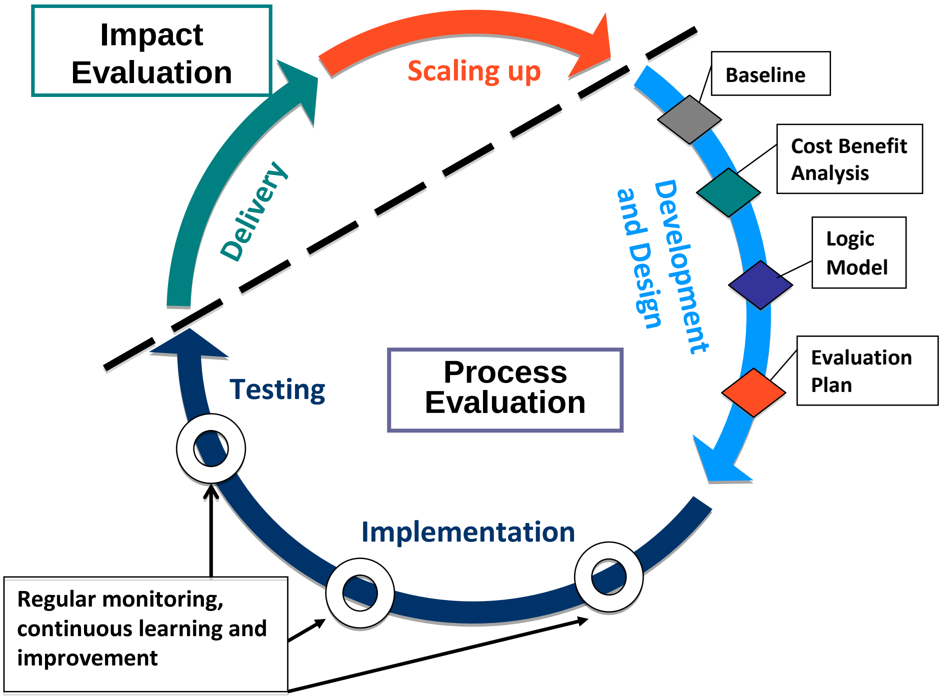
- Confidentiality and data protection issues.
- Data quality, such as cohort data inconsistencies and gaps.
- Data format and incompatible systems.
- How to keep personal data anonymous.
- Secure storage of information.
- Limited resources devoted to information sharing.
- Lack of systematic processes for information sharing.
- Lack of trust between organisations.
- Informed consent and wider ethical issues.

Further information, tools, resources and case studies to help public sector organisations share information effectively, efficiently and safely can be found on the Improving Information Sharing and Management website.

The evaluation cycle

Evaluation should not be considered a stand-alone activity. It should be thought of as a set of linked tasks that are undertaken from the start and throughout a period of transformation. The diagram below describes the evaluation cycle linked to the stages of service delivery and type of evaluation.

Figure: 1.1



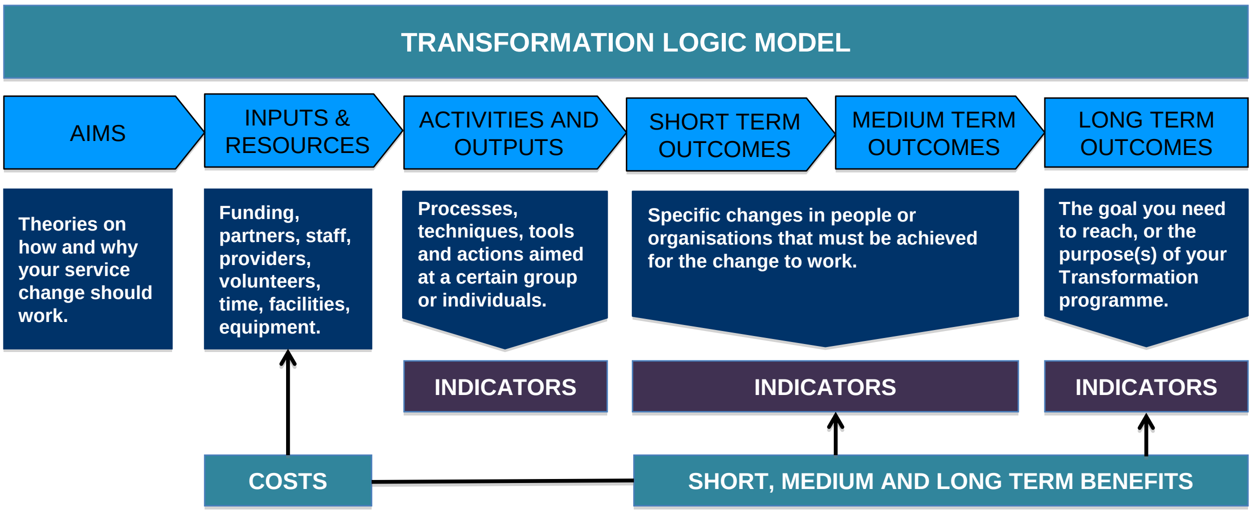
2.2 Evaluation preparation

Local providers developing service transformation proposals should factor in evaluation right from the beginning of the service design and development stage. This will enable better service delivery as it will enable projects or programmes to demonstrate how and why a service change can achieve intended outcomes. A useful tool that can help both the initial service development and evaluation planning stages is a transformation logic model, also known as ‘intervention logic’ or ‘programme theory’.

A **transformation logic model** describes a transformation programme in a diagram and a few simple words. It shows a causal connection between the need identified, the intervention to address it and the impact on individuals and local communities. A transformation logic model is a useful framework to identify outputs and outcomes that will feed into the evaluation. Ideally, this should be developed to support the design of the policy or service intervention. If a transformation logic model does exist it should be tested to help hone the evaluation process. If one does not exist, now is a good time to develop one. This should not be a technical exercise, but needs to involve those that are designing and implementing the service intervention.

A logic model can be used internally as a tool for supporting the monitoring and evaluating projects, and externally as a way of summarising the project’s overall purpose for partners and stakeholders. The diagram below sets out the key elements of an evaluation approach to assessing transformational change.

Figure: 1.2



For further details on developing the logic model see:

- The HM Treasury guide: Magenta Book, page 41 and 54.
- University of Wisconsin: developing programme logic models.

For practical guidelines and diagrams on the theory of change see Nesta's Open Workshop guidance on planning for evaluation.

Practical considerations when planning the evaluation

When deciding how to conduct an evaluation consideration should be given to practical issues, and in particular whether the evaluation and methods used are achievable.

Key questions include:

- How can you ensure that key decision makers receive the information in the most useful, cost-effective way?
- What are the levels of expertise of the evaluator(s)? Do you need to build capability, for example improving statistical expertise to support effective impact evaluations?
- Which evaluation method(s) will provide all of the information required, including capturing sufficiently accurate information to enable attribution to be made?
- What types of data, qualitative or quantitative, will be needed?
- Will the evaluation be done internally, perhaps by staff developing the intervention to be evaluated, or by external evaluators, or by a mixture of?
- What are the budgetary constraints?
- What are the time constraints?
- In which form will the intended participants best understand the evaluation questions?

The evaluation plan

Evaluation should ideally be planned at the same time as a service is being developed or redesigned. The service development or redesign should guide the evaluation by clarifying what activities will be undertaken and based on this what indicators or information will be used. When planning an evaluation consideration should be given to how the evaluation can support ongoing change by feeding into and informing business plans.

The evaluation plan should build on the logic model and outline what information will be collected, when and by whom, what question the information will answer, and the audience that will receive the evaluation. There may be several audiences and each may have unique requirements that should have been identified through the projects initial stakeholder analysis. For example, the management team, partners and investors may be interested in whether the intervention achieved its intended outcomes, was effective and if there was a return on investment (e.g. £ per measurable indicator). Other stakeholders and the customers may have additional questions about service quality that they want answered through evaluation.

The evaluation plan should include an outline of the preferred research methods to be used. An evaluation plan should therefore include:

- Introduction and background to the project or programme.
- Type of evaluation.
- Level and type of support provided through the change or series of changes and number of beneficiaries.
- What is to be evaluated? In particular, details of who the evaluation is for and its aims, objectives and targets that it will need to report against, and how the evaluation evidence will be used.
- Monitoring data and other information and feedback available for the evaluation.
- Past evaluation evidence, or evidence from evaluations of similar projects or programmes whether internal or external.
- Timescales – start date and final deadline.
- Management roles and responsibilities.
- Whether a survey of other primary research with beneficiaries is required.
- Whether the evaluation is to be conducted internally or by external consultants.

It can be helpful to plan an evaluation using a template that reflects the considerations listed above. Local providers may find it helpful to use or adapt the template at **Annex A.**

2.3 Monitoring and data collection

It is important to consider what to monitor and what data to collect early on in the development stage of a service transformation proposal. The type of evaluation approach chosen will inform the type of monitoring and data collection activity to undertake.

Monitoring should be an ongoing process throughout the life of a service transformation project or programme. Collecting information as systematically as possible will help to monitor progress, make necessary changes and track outcomes for the intended groups for whom the service change is designed. Developing a monitoring strategy could be really useful.

As part of the monitoring and data collection process it may be necessary to collect a wide variety of data from different sources at different stages of service delivery. The most important information is likely to be:

- **Outputs** – the direct and measurable results of a program's activities or services, often expressed in terms of units (hours, number of people or completed actions). Generally, more than one ouput is necessary to produce a final outcome.

- **Outcomes** – the changes, benefits, learning or other effects that occur as a result of activities. These are often expressed in terms of an increase in understanding, and improvements in desired behaviors or attitudes of participants. Outcomes can be anticipated or unanticipated and may be positive or negative. A single outcome is often the result of multiple outputs.

Types of data

Evidence is the information gathered to make robust assessments about service changes. It is often referred to as either:

- **Quantitative information** - things that can be counted such as how many people stopped re-offending after participating in a programme.

- **Qualitative information -** how the service is experienced by those involved, generally gathered through personal testimonies and eyewitness accounts.

A mix of both kinds of evidence will ensure the evaluation is robust and worthwhile.

The Magenta Book states that there are four main types of data, which, if planned for and collected in a systematic way, can support both process and impact evaluations:

- Existing administrative data that has not been collected specifically for the evaluation.
- Long term, large scale, often longitudinal, structural survey data managed locally, by central governments or the Office for National Statistics.
- Monitoring data or performance management data that are already being collected to support the administration of the policy.
- New data collection needed to support the evaluations information needs.

New or additional data, quantitative and qualitative, required to support the evaluation can be collected using a range of methods. Some key methods are listed below in table 4, section 2.4.

Checklist for collecting data

- [ ] Decide which evidence will be collected before, during and at the end of the project.
- [ ] Make sure the evidence relates to the kind of things that partners and stakeholders need to know.
- [ ] Be sure the methods chosen are proportionate to the value and size of the project.

- Make sure the methods are manageable, given the time and costs involved.
- Make sure they take account of equal opportunities and ethical issues.
- Secure permission from those providing the evidence that it can be used.
- Work out how the data will be kept safe and confidential.
- Consider what comparators can be used or need to be developed.

2.4 Research methods

There are many methods that can be used to conduct evaluations. Robust evaluations often use a mix of several different methods.

The methods chosen to evaluate will vary and depend on many factors. Some key factors are:

- The design of the service intervention.
- The evaluation objectives, in particular whether it is aiming to be able to attribute change to the project or programme.
- The stage of the change or intervention being evaluating.
- The level of resources available for the evaluation.
- The risks surrounding the service change or intervention.
- Type and scale of service change or intervention.

Most evaluations will apply a mixture both quantitative and qualitative research methods. Some examples are given below:

Quantitative research methods: Commonly used to answer 'how much?' or 'how many?'	**Qualitative research methods:** To answer 'who?', 'what?', 'how?' and 'why?'
▪ Closed answer questionnaires. ▪ Observation counts. ▪ Analysis of administrative data. ▪ Expenditure data. ▪ Experimental and quasi-experimental.	▪ Interviews. ▪ Focus groups. ▪ Participant observations. ▪ Case studies. ▪ Stories, blogs and diaries.

Table: 4

Evaluation needs to be robust enough to reflect the nature and scale of change being proposed in order to stand up to public scrutiny. There is increasing emphasis on experimental and quasi-experimental approaches to evaluation which can add to the overall robustness of evaluations focusing on the impact of the service change initiative.

If the focus of an evaluation is to understand the effect of a change it is most likely that an impact evaluation will need to be undertaken (see table 2 'types of evaluation' in section 1.4 above).

Impact evaluation assesses the changes that can be attributed to a particular intervention, such as a project, programme or policy. It assesses both the intended changes and, ideally, the unintended ones. In contrast to outcome monitoring, which examines whether targets have been achieved, impact evaluation is structured to determine whether the changes can be attributed to the intervention. An example would be how changes in participants' well-being compared to no action being taken. This involves counterfactual analysis, a comparison between what actually happened and what would have happened in the absence of the intervention. Impact evaluations seek to answer cause-and-effect questions. In other words, they look for the changes in outcome that are directly attributable to a project or programme.

Attributing changes in outcome means that impact evaluations need to pay particular attention to the evaluation methods used. These methods are described in the following sections.

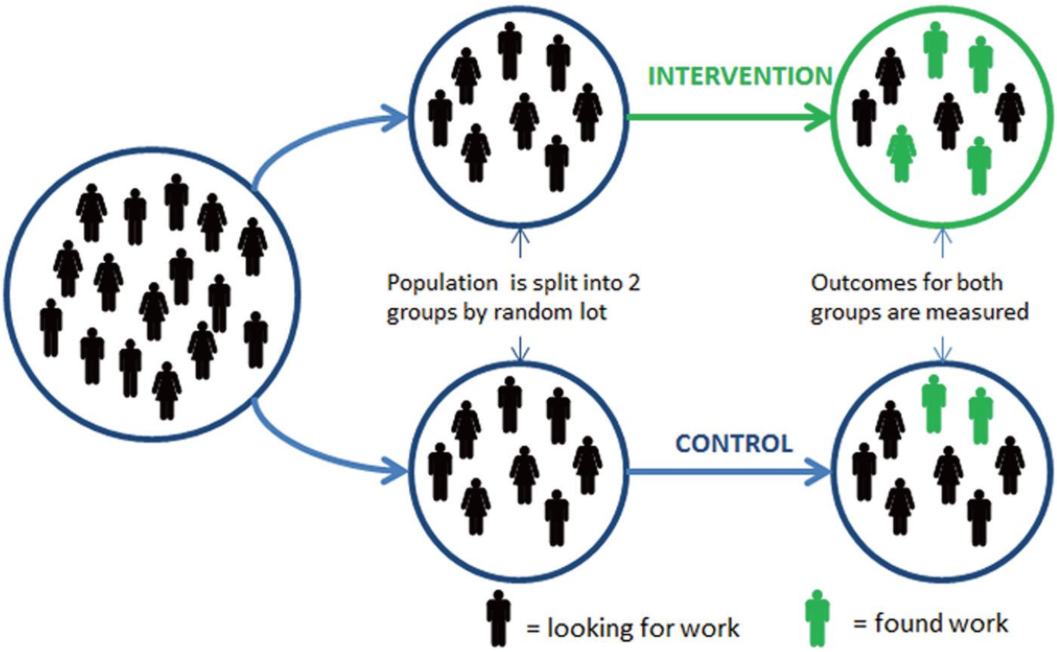
Experimental

Randomised Control Trial: is a process in which a number of similar people are randomly assigned to two or more groups to test a specific change. One group, the experimental group, receives the change being tested while the other, the control group, receives an alternative service or the existing service.

The groups are monitored to assess differences in outcomes. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This reduces bias and enables comparison of the effectiveness of a change against what would have happened if nothing had changed.

The diagram below demonstrates the application of a randomised control trial to a new Back to Work programme

Figure: 1.3



Source: Test, Learn, Adapt. Cabinet Office, 2012

If the service change or intervention is amenable to randomisation, this can be a very effective evaluation method, free from bias and may not be as expensive to implement as some other methodologies. It does however require careful planning and appropriate analytical capability.

Quasi-Experimental

The basic structure of a quasi-experimental evaluation involves examining the impact of an intervention by taking measurements before and after it is implemented. A quasi-experimental design is one that looks a bit like an experimental design but lacks the key ingredient of random assignment. Below are some examples of quasi-experimental design taken from the Magenta Book supplementary guide "Quality in policy impact evaluation".

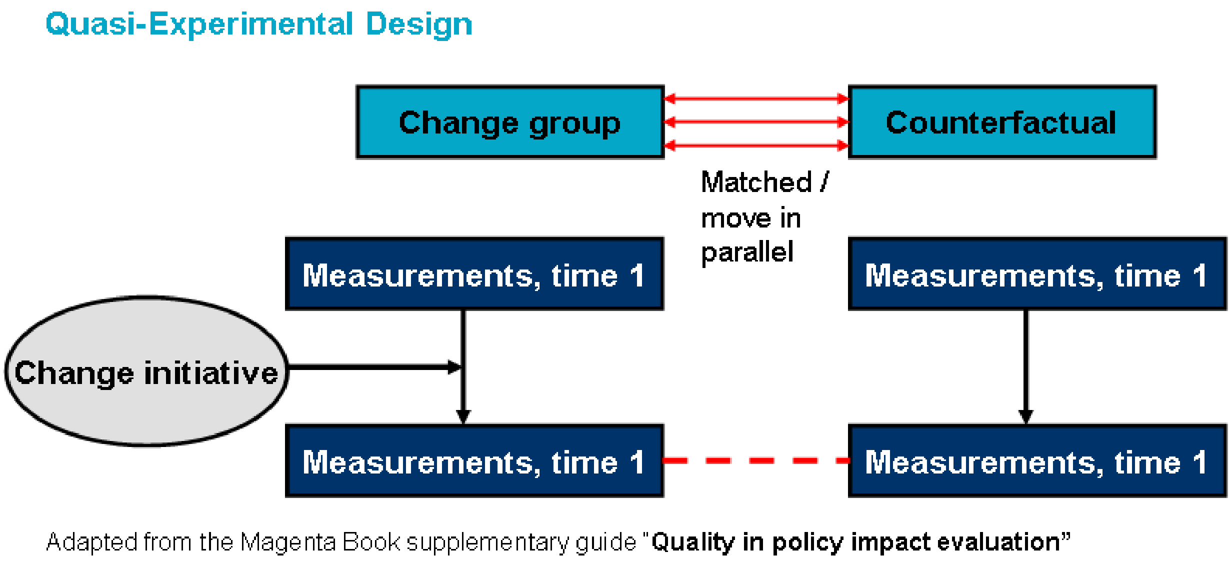
Intervention group versus well matched counterfactual: outcomes are compared between the intervention group and a comparison group directly matched to the intervention group on factors known to be relevant to the intervention outcome. If this is done well, then in principle the matched comparison group can be treated as though it was created randomly – hence the description 'quasi-experimental'. The statistical techniques such as Propensity Score Matching (PSM) to achieve this matching are covered in detail in the Magenta Book (pages 107 to 119). Further information on the use of PSM can be accessed on the Department for Work and Pensions research paper 'The use of propensity score matching in the evaluation of active labour market policies'.

Strong difference-in-difference methodology: there is no direct matching instead this approach involves a before and after study comparing two groups where there is strong evidence that outcomes for the two groups have historically moved in parallel over time. This is similar to the matched-comparison group design but there is no literal matching. Instead, the trends over time for the two groups are compared to provide an estimate of the overall impact of a policy.

The quality of these types of evaluations can vary, and is strongly linked to the quality of the data used with a substantial amount of data often being needed to do this well, and new primary data sometimes required. As with experimental designs, sufficient sample size is also important.

The diagram below illustrates the application process for quasi-experimental design.

Figure: 1.4



There are other methods of non-experimental nature that are generally regarded as weaker and riskier in terms of measuring attribution and open to bias. Further guidance on experimental and non-experimental methods are available:

- Cabinet Office, Behavioural Insights Team (2013) "Test, Learn, Adapt: Developing Public Policy with Randomised Control Trials"
- The Magenta Book and supplementary guidance "Quality in Policy Impact Evaluation"

Box 2: Example of an evaluation using Randomised Control Trial

Troubled Families Programme, Manchester City Council

The troubled families programme was designed to transform the family recovery services in the city. As part of the planning process a detailed cost benefit analysis was carried out raising a critical question:

- **Is the new delivery model more cost effective than business as usual? If so, how much more cost effective?**

The new delivery model adopted a key worker approach and incorporated a whole family plan. The key worker developed a plan with the family and had access to a range of public and additional commissioned services that could help the family address underlying issues with the objective of becoming independent.

'Business as usual' was less coordinated. Individuals in families were expected to navigate the system on their own and front line practitioners operated in different silos.

The new model was more expensive but it would be worth it if the outcomes were improved and the family became less dependent on benefits and other public services.

The evaluation took the form of a two year study. 240 families in two areas of the city were randomly selected to receive the new service and 240 families in the same areas continued to receive the usual service.

As well as indicators required to answer the key evaluation question, a raft of data was collected through assessments, plans and progress reports to explain which interventions were more successful and which types of families benefited most from the interventions.

The main challenge was to implement the intervention being tested with fidelity and to explain to front line staff the purpose of the study, the reason why so much data was required and the importance of avoiding contamination. Without the evidence from the evaluation, it would have been difficult to convince decision makers that the additional resource required represented a prudent investment.

The evaluation outcomes were incorporated into the cost benefit analysis, demonstrating that the new delivery model was more effective. In addition, the data was analysed to inform future decisions such as changing the Troubled Families criteria and training front line staff.

Enquiries to Sarah Henry, Intelligence & Performance lead: s.henry@manchester.gov.uk

2.5 Analysis and data interpretation

Having gathered together all the information required for evaluating a change initiative it is necessary to analyse and interpret the data.

Ideally, a plan for analysing data should be developed well before it is time to conduct analysis. The best time to develop an analysis plan is during the initial stages when key evaluation questions are being considered and appropriate data is being identified. It is important to match the approach to analysis to the type of information available and the kinds of evaluation questions to be answered.

Depending on the scope of an evaluation and available resources it may be necessary to organise the data before beginning the analysis. Seek advice from analytical experts if it is unclear how best to collect and organise data.

When handling data, particularly personal information and including the sharing and storage of data, it is important to comply with relevant legislation and guidance on data protection and ethical assurance.

Analysing quantitative data

Quantitative data will pre-dominantly be analysed using a range of statistical methods that can be used to understand and explain change. For example, close-ended survey questions will typically be coded into numbers so they can be analysed quantitatively. Some statistical analysis of quantitative information can be quite complex and expert advice might be required. However, in most instances descriptive statistics can be straight-forward to produce and will help summarize and identify key findings from which conclusions can be drawn. However, as described above, if the aim is to attribute change from the evaluation then data, its collection and its analysis must be capable of supporting such assessment using the experimental or quasi-experimental techniques described above.

Descriptive analysis is used to reduce raw data to an understandable level. Some common methods include:

- ***Frequency distribution***: tables or charts showing how many evaluation participants fall into various categories, for instance age, gender or ethnicity. Frequency distributions can generally be presented in simple graphs and charts that can easily be presented to a non-technical audience.

- ***Central tendency***: the number that best represents the 'typical score'. These include the mode, the number or category that appears most frequently; the median, the number in the exact middle of the data set; and the mean, the arithmetic average of the numbers. Each measure of central tendency has its advantages and disadvantages. Using the median, rather than the mean, for example, will give an answer that is unaffected by very large or small values, which can distort results.

- ***Variability***: which show the amount of variation or disagreement in the results. Common measures of variability include range, the difference between the highest and lowest scores, and standard deviation, which shows the degree of dispersion from the mean. Standard deviation can be useful for understanding if the characteristics of a population are fairly similar, for example if they are mainly clustered around the mean average, or if the nature of the population is more diverse.

More complex statistical techniques can also add value to evaluations by helping to understand and explain changes in data. For instance, calculating confidence intervals provides an assessment of the reliability of an estimate. This may be useful when attempting to measure improvements in performance following an intervention. For example, a survey may show that residents are more satisfied with services this year than they were last year – confidence intervals provide a view on the probability that the difference represents a real change rather than simply part of a trend or random event.

Following on from this, statistical inference testing includes a range of techniques that can distinguish between good and bad impacts. Comparing data from different areas, or from trials can give an insight into whether a new delivery model is actually delivering better results, or if a particular area is faring better than another. This involves more technical knowledge and may require analytical expertise.

Analysing qualitative data

On its own, or in combination with quantitative information, qualitative data can provide rich information about the impact of a project or programme. At the simplest level, qualitative analysis involves examining data to determine how it answers specific evaluation questions.

The first step when analysing qualitative information is to reduce or simplify the information. Important information may be interspersed throughout interviews or focus group proceedings. During this first stage of analysis it is important to decide which issues to focus on or exclude from the analysis. While it can be difficult to remove comments provided directly by informants, it is important to remain focused on the evaluation questions being interrogated and the relevance of the information to these questions.

When analysing qualitative data, look for trends or commonalities deeply-rooted in the results. Depending on the amount and type of data available, it may be helpful to assign codes to the responses to help categorise comments. When reporting findings such codes will help to identify the most prevalent themes that emerged. Key quotes are also a powerful way of illustrating important themes.

Always be careful when using potentially rich information from qualitative analysis and ensure that individuals are not inadvertently disclosed, for example, through the description of their or others family circumstances or service history.

Tips for analysing data

- Review and clean data to make sure everything is accurate, complete, and any inconsistencies have been resolved before beginning analysis.

- Leave enough time and money for analysis – it is easy to focus so much on data collection that not enough time is left to analyse the results.

- Identify the appropriate statistics for each key question – seek analytical support if needed.

- Keep the analysis simple.

Interpreting results and drawing conclusions

Both quantitative and qualitative analysis has its limitations. While the analysis can help to summarise and identify key findings, it is critical to interpret the results and draw reasonable conclusions. Drawing conclusions involves stepping back to consider what the results mean and to assess their implications. During this phase, consider the following types of questions:

- What patterns and themes emerge in the results?
- Are there any deviations from these patterns? If yes, are there any factors that might explain these deviations?
- Do the results make sense?
- Are there any findings that are surprising? If so, how can they be explained?
- Are the results significant from a clinical or statistical standpoint? Are they meaningful in a practical way?
- Do any interesting stories emerge from the responses?
- Do the results suggest any recommendations for improving the program?
- Do the results lead to additional questions about the program? Do they suggest that additional data may need to be collected?
- And, again, if the aim is to try and attribute change from the data, you need to be sure that the evaluation method enables you to do this.

Watch for, and resolve, inconsistencies

In some cases contradictory information can be gathered. For example, some stakeholders may describe important benefits of a redesigned service but these improvements do not appear in pre and post test comparisons. Various stakeholders may also disagree. For instance, staff may report improvements in participants that are not reported by the participants themselves. It can be challenging to determine which information is accurate, especially when comparing different peoples' viewpoints or perspectives. It is important to consider the validity of each information source, and to remember that various stakeholders can have valid viewpoints that vary based on their unique perspectives and experiences. To the extent that it is possible it is important to understand and resolve these discrepancies and report them to stakeholders.

2.6 Communicating the results

Reporting the findings of an evaluation is a key part of the evaluation cycle. It is worthwhile giving this some thought particularly in relation to presenting the findings, any lessons learnt and the intended audience. Most evaluations will be presented in the form of a written report which might typically include:

- **Introduction and approach**. An explanation of what is contained in the report and the process/approach used to gather the information. Explain the context of the intervention being evaluated, the nature and rationale for the intervention, the key evaluation question and methods.

- **Outputs and results**. Set out the information collected. It is usual to start with outputs; showing the scale of what was done. Then describe outcomes; the effects of a change or intervention.

- **Conclusions and recommendations**. Use this section to reflect on what the results mean: what lessons have been learnt by doing this work? What seems to be effective in addressing the problem originally identified? Does the original problem still exist? If so, how has it changed? What recommendations are relevant to others in considering the best ways of addressing these issues?

Other points to consider when reporting the evaluation findings include:

- What ***resources*** will be needed? Are sufficient resources available (money, staff and time) specifically to reporting – especially if disseminating the results to a much wider audience could benefit other local, regional or national policymakers.

- What will the reporting ***outputs*** be? A written report may not always be suitable. Other methods could include: presentations, website publication, events, press releases, newsletters.

- Is the ***timing*** of the report important? Are others undertaking similar changes? Are there any government papers or consultations that the evaluation could influence?

- Who is the intended ***audience***? Ensure the report will have a meaningful impact on the target audience. For example, investors or sponsors will want very concise messages so that they can make a decision, whereas practitioners will need more detailed information about how the work was done.

Part 3: Common pitfalls

Whatever approach taken to evaluation, there are some common pitfalls and challenges to be aware of; these are set out in the table below:

Pitfall	Mitigation
Collecting too much data Collecting a mass of data and then: not knowing what to do with it; and/or not being sure of its quality; and/or not knowing what it all means is one of the most common pitfalls of evaluation.	When planning an evaluation focus on: what is practical and possible to collect; what will be really informative; and what will be useful when analyzing information and reporting results. Base the evaluation approach and data collection on a robust logic model. This will focus the evaluation.
Waiting until after the project has started Evaluation should be a key part of service design. Waiting until the end before asking questions about the effect of service change is too late, by then those involved in an initiative may have moved on and the evaluation evidence will become reliant on peoples' impressions or memories.	Integrate evaluation planning into the project design. Ensure evaluation is linked to the outcomes the project is trying to achieve. This is best done as part of a business case – be clear what the project should achieve and how that can be credibly measured or assessed.
Not investing enough time or resource to evaluate There is no clear rule of thumb when considering the levels of resources to devote to monitoring and evaluation. It will vary according to whether the intervention or project is especially innovative, the level of risk and how scalable and sustainable it is.	Ensure information collection is part of everyday activity and be clear who is responsible for collecting this information. Make sure people with the right skills are engaged in the evaluation. It could be the deciding factor between a successful and unsuccessful evaluation.
Reliance on one source When collecting information and data over reliance on one source will undermine confidence in the analysis.	Try to use a range of sources and combine qualitative and quantitative information. Ensure that objective sources of information are used and not just surveys of those involved in the change.
Lack of ownership of the evaluation process Lack of direction or engagement from senior management and project or programme staff not bought into the importance of the evaluation process.	Involve senior management and project or programme staff early in the process and ensure they are clear about the importance of evaluation and its impact on the success of the project or programme.

Table: 5

Part 4: Externally commissioned evaluation

Monitoring and evaluation can be conducted in-house. The key issue is whether local staff have the relevant analytical skills or whether they can be developed. If this resource is not available it is often worth considering commissioning an evaluation from a third party.

There may at times be good justification to externally commission evaluation requirements, such as a need for an independent view which may carry more weight when trying to influence key partners and stakeholders. This section explains how you can get the best from using external suppliers. It considers:

- **The need for an external supplier**.
- **Hiring an external supplier**.
- **The evaluation budget**.
- **Managing the external evaluation**.
- **Learning the lessons**.

External evaluators can be brought in at any stage of a service transformation project. However, to ensure the best evaluation results, evaluation activities should be integrated into day to day work. Local providers should aim to build evaluation into project or programme plans rather than adding it on at the end. Building in evaluation at an early stage, even if the capability is not available locally will make it much easier for an external evaluator to build on existing plans.

4.1 The need for an external evaluation provider

It may be necessary to use an external supplier for the following reasons:

- Lack of internal resource or skills.
- Scale and complexity, such as the need to do sensitive work with troubled families.
- A need for specialist information collection skills or policy expertise.
- A need to obtain impartial feedback from service users or other stakeholders.
- A need for a fresh perspective and suggestions on ways to improve the project or programme.
- A need for an independent view.

4.2 Getting an external evaluation provider

It is best practice to competitively tender for an external evaluation. That may require an open procurement process or use of existing framework agreements. Regardless, it is important to prepare **an evaluation specification** to send to potential suppliers.

A poorly thought out or vague brief/specification is one of the key reasons why organisations fail to get the best from an external supplier. The table below sets out the key information a specification should cover.

An evaluation specification should cover:	
Context	▪ Why you need the evaluation ▪ What you need the evaluation supplier to do
Background	▪ What your organisation or project does ▪ Website details if you have one, so they can find out more
Aim	▪ The questions you need the evaluation to answer
The work	▪ The work you expect to be done ▪ Any specific stakeholders who need to be involved ▪ What you want them to produce such as a final report, progress updates, presentations ▪ Other requirements such as speaking at a launch event or helping with action planning on recommendations
Timetable	▪ Overall timeline for the evaluation ▪ The key evaluation milestones and any deadlines that need to be met
Managing the tender	▪ Who in your organisation will manage the contract with the supplier and whether there is a steering group or other relevant stakeholders to work with ▪ Any resources or data that you will make available
Supplier skills and knowledge	▪ What skills, knowledge, experience and values you expect the supplier to have. Ask them to tell you about all the people who will do the work, not just the person in charge
Bid evaluation criteria	▪ Be clear about the criteria to be used when assessing tenders from suppliers. This will help suppliers focus their tender on the more issues ▪ Set out what the bid evaluation process will be and any weighting to be applied to each criteria
Budget	▪ Under some circumstances it may be necessary to set out how much money you have to pay for the evaluation. Your procurement specialists will be able to advise. Provide a range and say if this includes expenses and VAT
Practical details	▪ What needs to be covered in the tender, for example: - An outline of how they would go about the work - A summary of their skills, knowledge and experience - Breakdown of costs

Table: 6

Circulate the specification with details of the short-listing and interview process. Interview potential suppliers even if only one tender is received to ensure they are right for the job. The interview should always involve the person who will manage the relationship with the supplier.

4.3 The Evaluation Budget

Before circulating the specification, an assessment of how much the work is likely to cost needs to be carried, especially if funds are limited. This is a useful check to see whether the requirements are realistic in terms of what can be done with the amount available. To estimate the evaluation budget, look at the specification and consider:

- What are the potential methods to be used? For example, will it involve major surveys or experimental methods, what is the minimum sample size? etc.

- How many days work could be involved? For example, two days desk research, days speaking to service users and so on. When the number of days involved, have been estimated, multiply this by a daily rate for a supplier. Most suppliers charge a fixed amount per day although they may charge less for longer term pieces of work. Pick an amount that reflects normal rates locally.
- What other costs might there be in carrying out the evaluation? For example, room hire for focus groups or steering group meetings, phone costs, administration or travel costs.
- Are there any other costs that need to be built in, such as staff or volunteer costs, report publication or dissemination costs?
- Decide whether to include the costs the supplier incurs during the tendering process, such as travel to interview.

If the draft budget exceeds the amount available it may be necessary to renegotiate the budget with the budget holder or cut down the brief by prioritising those activities that will add most value to the project and those that could be delivered by in-house staff.

4.4 Managing the external supplier

When an evaluator has been choose a contract should be put in place. This should set out the terms and conditions and may include issues such as who owns the material that is produced, confidentiality, conflicts of interest and termination.

A named individual should be made responsible for managing the contract and the relationship with the supplier to ensure they deliver value for money. It is critical that the supplier delivers what is required for the evaluation including their assessment of the strengths and weaknesses of the evaluation. The designated individual managing the supplier and the evaluation process should be responsible for:

- Communicating with the supplier.
- Making sure the supplier has access to the information required.
- Troubleshooting, where sensitive issues arise.
- Checking that deadlines are met, payments made and outputs delivered as required.

Building an effective relationship with the supplier will reduce risk, enabling unforeseen events to be dealt with promptly.

4.5 Learning the lessons

An external evaluation can present both opportunities and challenges. If the brief/specification is well written and appropriate, a good supplier is appointed and the contract well managed the findings should be relevant and there should be no great surprises. It is essential to build in time afterwards to think about how to use the findings to improve the project or programme and promote the lessons learnt. It might be possible to use the supplier to do this if budgets allow. It is also worth reflecting on the external evaluation process itself and what might be done differently in the future to improve it.

Part 5: Existing evaluation resources and guidance

5.1 Guidance to Support Self Evaluation

The **Magenta Book** is guidance on what to think about when designing an evaluation. It explains how results can be interpreted and presented, and what should be considered in this process:
https://www.gov.uk/government/publications/the-magenta-book

Joseph Rowntree - **guide to robust evaluation of social welfare policies**, including some good case studies:
http://www.jrf.org.uk/publications/making-most-it-economic-evaluation-social-welfare-field

ESRC Survey Question Bank – provides a stock of standard programme evaluation questions for you to copy or fit to your own purposes and also access to a wide range of secondary data including large-scale government surveys, international macrodata, business microdata, qualitative studies and census data from 1971 to 2011:
http://discover.ukdataservice.ac.uk/variables

5.2 Developing logic/theory of change models

University of Wisconsin, **Guide to developing logic models**:
http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html

Practical guidelines and diagrams on the theory of change from Nesta's "**Open Workshop guidance on planning for evaluation**":
https://openworkshop.nesta.org.uk/content/topic-1-planning-evaluation

5.3 General Evaluation resources

The NAO report "**Case study on integration: Measuring the costs and benefits of Whole-Place Community Budgets**":
http://www.nao.org.uk/wp-content/uploads/2013/03/10088-002_Whole-Place-Community-Budgets.pdf

Preskill, H. and Beer,T*.* *(2012)* ***Evaluating Social Innovation****,* Centre for Evaluation Innovation. Discusses the role of *developmental evaluation* to support innovation where outcomes methods are less determined:
http://www.evaluationinnovation.org/sites/default/files/EvaluatingSocialInnovation.pdf

Local Government Association – **Evaluation of the Customer-Led Transformation Programme**. A report evaluating the impact of projects using customer insight and social media tools to re-design and inform the way services are delivered:
http://www.local.gov.uk/documents/10180/11417/L13-840+CLT+evaluation+report+V5.pdf/0bbea5f9-0914-4690-b92d-8f48a7d75d12

Cabinet Office, Behavioural Insights Team (2013) ***Test, Learn, Adapt: Developing Public Policy with Randomised Control Trials***. Guidance on conducting Randomised Control Trials:

http://www.behaviouralinsights.co.uk/publications/test-learn-adapt-developing-public-policy-randomised-controlled-trials

DWP research paper "**The use of propensity score matching in the evaluation of active labour market policies**" provides helpful and practical use of matching techniques: http://webarchive.nationalarchives.gov.uk/20130314010347/http:/research.dwp.gov.uk/asd/asd5/WP4.pdf

HM Treasury - **The Green Book, Appraisal and Evaluation in Central Government** - a guide on economic appraisal and analysis, largely prior to implementation: http://greenbook.treasury.gov.uk/

Evaluation publications covering **integrated health and social care** from the Nuffield Trust: http://www.nuffieldtrust.org.uk/our-work/evaluation

Practical guidance from Nesta on **evaluating innovations**: https://openworkshop.nesta.org.uk/content/topic-1-planning-evaluation

UK Evaluation Society – **Guideline for good practice in evaluation**: https://www.evaluation.org.uk/assets/UKES%20Guidelines%20for%20Good%20Practice%20January%202013.pdf

Evaluation Support Scotland – A rich source of evaluation resources, guides and tools: http://www.evaluationsupportscotland.org.uk/

5.4 Information Sharing, Data Protection and Ethics

The **website with tools and resources to help any public organisations to share information** appropriately with partners and colleagues: http://informationsharing.co.uk/

Code of practice from the Information Commissioners Office: http://www.ico.org.uk/for_organisations/data_protection/topic_guides/data_sharing

Guide to data protection from the Information Commissioner: http://ico.org.uk/for_organisations/data_protection/the_guide

Information sharing: guidance for practitioners and managers: http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/00807-2008BKT-EN-March09.pdf

Information sharing flowchart and seven golden rules: http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/Info-sharing_poster.pdf

GSR Professional Guidance – **Ethical Assurance for Social Research in Government**: http://resources.civilservice.gov.uk/wp-content/uploads/2011/09/ethics_guidance_tcm6-5782.pdf

Social Research Association Ethics Guidelines: http://the-sra.org.uk/sra_resources/research-ethics/ethics-guidelines/

5.5 Cost Benefit Analysis Guidance and Unit Cost Database

Supporting Public Service Transformation: Cost Benefit Analysis Guidance for Local Partnerships - details a methodology for the cost-benefit analysis model originally developed by New Economy, local authorities and other public sector agencies across Greater Manchester. The guidance is based on HM Treasury's Green Book: https://www.gov.uk/government/publications/supporting-public-service-transformation-cost-benefit-analysis-guidance-for-local-partnerships

The CBA guidance is supported by a **CBA Excel model** which provides a structured and consistent approach for applying this methodology and a **Unit Cost Database** which includes a broad range of specific unit costs (currently over 600) in order to calculate both the costs of delivering proactive services and the potential savings in reactive costs that can be achieved. These resources are available at:
http://neweconomymanchester.com/stories/1828-new_economy

Annex A - Evaluation Planning Template

Evaluation questions (what do you want to know?)	How will you know it? (Indicators)	What information do you already have?	What additional information will you need?	What methods will you use to collect the additional information?	How much is it likely to cost?	Who will be responsible for collecting this information?	When does the information need to be collected by?