



Department  
for Transport

# TRANSPORT ANALYSIS GUIDANCE

## Unit E1: Evaluation

November 2022

Department for Transport

Transport Analysis Guidance (TAG)

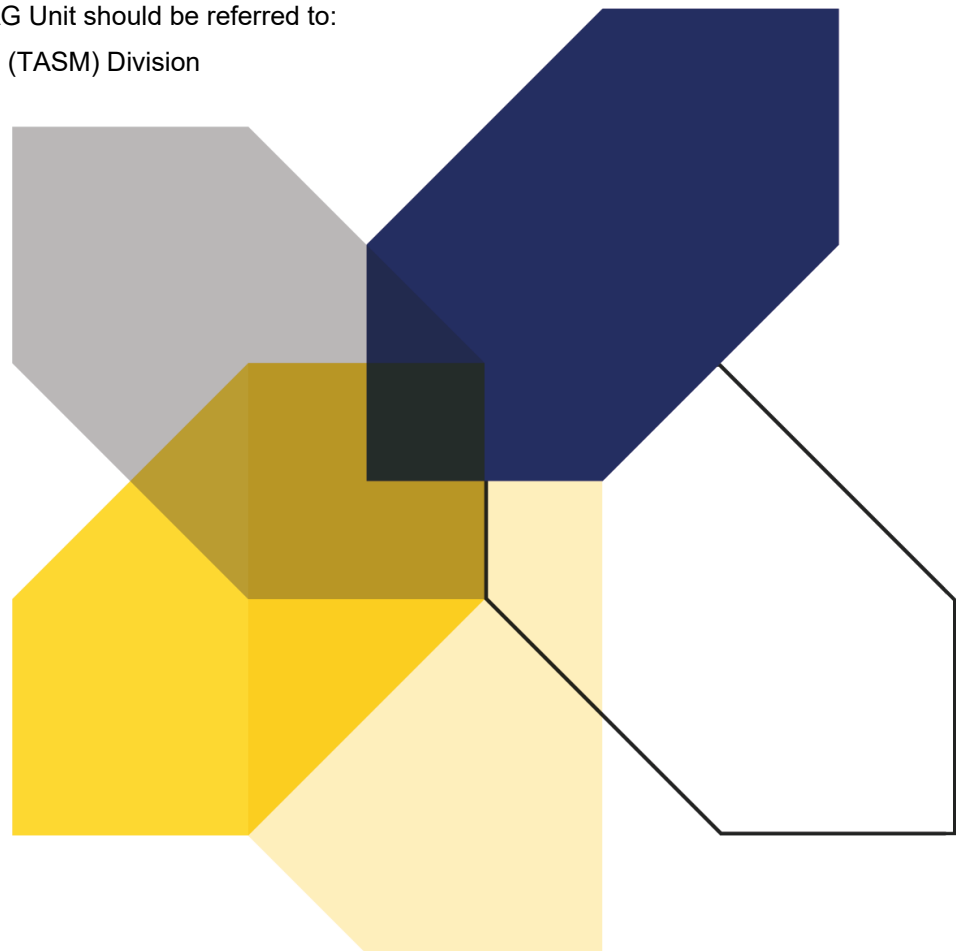
<https://www.gov.uk/transport-analysis-guidance-tag>

**This TAG Unit is guidance for the Evaluation Practitioner**

Technical queries and comments on this TAG Unit should be referred to:

Transport Appraisal and Strategic Modelling (TASM) Division

Department for Transport  
Zone 2/25 Great Minster House  
33 Horseferry Road  
London  
SW1P 4DR  
[tasm@DfT.gov.uk](mailto:tasm@DfT.gov.uk)



# Contents

1. The role of monitoring and evaluation	4
1.1 The purpose of this unit	4
1.2 What are monitoring and evaluation?	5
1.3 Evaluation of transport schemes	6
1.4 Relationship with benefits management	7
2. Evaluation planning	8
2.1 Embedding in delivery	8
2.2 Building from business case and appraisal	9
2.3 Piloting and testing	12
2.4 Budget and proportionality	13
2.5 Governance	14
3. Understanding the intervention	15
3.1 The theory of change	15
3.2 Logic mapping	17
4. Evaluation design	18
4.1 Evaluation questions	18
4.2 Process evaluation	19
4.3 Impact evaluation	21
4.4 Value-for-money evaluation	23
5. Focus on impact: The counterfactual	25
5.1 What is the counterfactual and why does it matter?	25
5.2 Common counterfactual and comparative approaches	26
5.3 Caveats	30
6. Measures of evaluation quality	31
7. Data requirements	33
7.1 Existing data	33
7.2 The baseline	34
7.3 Monitoring data	35
7.4 Other bespoke data	35
7.5 Data access	36

7.6	Data protection and ethics	36
8.	Reporting and dissemination	37
8.2	Interim reports and emerging findings	37
8.3	Feedback for appraisal	38
8.4	Publication	38
9.	Further resources	39
10.	Document provenance	40

# 1. The role of monitoring and evaluation

## 1.1 The purpose of this unit

- 1.1.1 This unit provides an overview of good practice in planning the evaluation of transport interventions to ensure robust evidence can be collected about the difference that they are making in practice and lessons that can be learnt for the future. It seeks to identify and complement existing guidance in this area (section 9) and thus support scheme promoters prepare business cases for publicly funded projects and programmes.
- 1.1.2 The unit is intended to support evaluation planning for a range of transport interventions in terms of mode, type of intervention (e.g., policy, package of measures, programme) and context (local, regional and national). The guidance does not provide specific advice for evaluating infrastructure programmes, although the general principles discussed below will be useful for these types of programmes as well. (See section 1.3 for more information.)
- 1.1.3 The guiding purpose of the unit is to help owners and promoters of transport interventions to plan and conduct proportionate evaluation for two main purposes:
- **Accountability** – to demonstrate to what extent an intervention’s objectives have been delivered successfully and public money has been spent well; and
  - **Learning** – to provide evidence of the effectiveness of the intervention, to manage risk and uncertainty and guide future development of the current intervention, as well as the design and delivery of future interventions of a similar type.
- 1.1.4 Throughout this unit the term ‘intervention’ is used to describe a range of transport activities which might require monitoring and evaluation: policies, programmes, schemes and packages of measures. In the field of transport, there is a wide range of interventions, such as policies to promote cycling and walking, smart ticketing programmes, public subsidies for bus services, repairs and enhancements to existing road and rail infrastructure and the construction of new roads, rail lines and airport capacity.
- 1.1.5 This guide cannot provide detailed instruction on how to monitor and evaluate such a wide range of interventions, and it does not attempt to do so. Its aim is to establish some common principles which can guide monitoring and evaluation activity for a broad range of transport interventions and modes. Owners and

promoters of transport interventions are strongly encouraged to involve evaluation experts in the early planning and ongoing implementation of monitoring and evaluation activities to ensure the best possible evidence is gathered.

## 1.2 What are monitoring and evaluation?

- 1.2.1 The elements of the policy development cycle are often described as: **R**ationale, **O**bjectives, **A**ppraisal, **M**onitoring, **E**valuation and **F**eedback, or ROAMEF for short, as set out in the Treasury's Green Book<sup>1</sup>. It is important to understand that this is not a linear process, but each element draws on and informs the others. To ensure robust defensible evidence will be available to demonstrate the impact and value for money of an intervention, it is essential to start planning monitoring and evaluation early, alongside planning and delivering the intervention itself.
- 1.2.2 **Monitoring** is a process for tracking progress in the implementation of an intervention by collecting data on its inputs, outputs and outcomes, with some evaluations continuing to monitor outcome measures after the intervention has ended. Monitoring should be planned before an intervention is delivered so that the data required to understand its progress is collected at the right time to inform future decisions for the intervention. This is particularly important for data that cannot be collected retrospectively, including baseline data, which should be collected before the intervention starts (see section 7.2).
- 1.2.3 **Evaluation** is a systematic process for understanding the relationships between an intervention's design, implementation and impact within the context in which it is delivered. It involves understanding how an intervention is being or has been implemented, what effects it has, for whom and why. It identifies what can be improved and estimates its overall impacts and cost-effectiveness.
- 1.2.4 Monitoring and evaluation are generally used in conjunction, with monitoring providing early evidence of outputs while evaluation builds on this to provide a fuller assessment of the outcomes and impacts of an intervention. Robust monitoring will typically be crucial in providing part of the necessary information base on which the evaluation will need to be founded. In the remainder of this guide, for the sake of simplicity, evaluation is used to refer to both monitoring and evaluation.

---

<sup>1</sup> The Green Book – Central Government Guidance on Appraisal and Evaluation, HM Treasury (2022).

## 1.3 Evaluation of transport schemes

- 1.3.1 This guidance is intended to be read in conjunction with the Treasury's Magenta Book<sup>2</sup>, which provides guidance on evaluation for government policy-makers and analysts, and the Green Book, which focuses on appraisal, with some advice on monitoring and evaluation. The Magenta Book sets out in more detail the role of evaluation in public policy, provides an overview of how to scope and design an evaluation, outlines key evaluation methods and provides advice on management of evaluations and dissemination of findings.
- 1.3.2 In the same way that other TAG modules seek to supplement the guidance in the Green Book on how to appraise government interventions with advice that is specific to transport interventions, so this document aims to supplement the Magenta Book with guidance on how to apply evaluation principles in the context of transport schemes.
- 1.3.3 It is important to note that not all publicly funded interventions fit well with the guidance for policy evaluation in the Magenta Book. This is particularly the case for large infrastructure programmes, which are difficult to adapt, reverse or replicate, and which typically have extended timeframes with outcomes building over long periods of time. It can be challenging to gather evidence to demonstrate the impact of, for example, a new road, rail line or airport because it can be difficult to know what would have occurred in the absence of the intervention. (Portfolio-level or meta-evaluations could help in these situations – see below.) Large infrastructure programmes are also distinctive in the tools they use alongside evaluation, including benefits management and some sophisticated modelling approaches. (See section 1.4 for a comparison of evaluation with benefits management. For further guidance, see the Green Book annex on [Valuing infrastructure spend](#) and the Infrastructure and Projects Authority's [Guide for effective benefits management in major projects](#).)
- 1.3.4 Other transport schemes fit more comfortably with the guidance for policy evaluation in the Magenta Book in being quicker to implement, incremental in nature, adaptable, reversible and potentially generalisable. Examples include smart ticketing schemes, policies encouraging cycling and walking, variable speed limits or initiatives supporting innovation. However, we will set out some good practice principles that can be followed in developing evaluation plans for all interventions. Where applicable, these will refer to the process of business case development for government projects. Some case studies of published evaluations will be provided to illustrate good approaches, and further resources will be recommended.
- 1.3.5 As many transport schemes have similar objectives, designs, benefits and costs to other schemes of the same transport mode, it is also useful to consider how

---

<sup>2</sup> The Magenta Book – Central Government Guidance on Evaluation, HM Treasury (2020)

evaluation evidence from them can be drawn together. Two helpful approaches are:

- **Portfolio evaluations.** This is where consistent evaluation approaches are used across a portfolio of schemes, to facilitate assessment of impacts at the portfolio level and comparison of impacts between different schemes.
- **Meta-evaluation.** This is where a consistent analytical framework is used to aggregate evaluation findings from schemes which had some common objectives but have not necessarily used consistent evaluation approaches.

## 1.4 Relationship with benefits management

- 1.4.1 For transport interventions, evaluation is often used in conjunction with a process called benefits management, especially for government major infrastructure projects. This section describes benefits management and its relationship with evaluation.
- 1.4.2 A benefit of an intervention is defined in benefits management as the measurable improvement resulting from an outcome perceived as an advantage by one or more stakeholders. The appraisal process outlined by TAG is the primary method by which benefits are quantified and valued for transport interventions.
- 1.4.3 Benefits management is the identification, definition, planning, tracking, realisation and optimisation of benefits. Its aim is to ensure organisations realise the planned benefits from their investments. Benefits management is described in APMG's *Managing Benefits* (2012). The Infrastructure and Projects Authority supports the use of benefits management on government major projects and has published a guide<sup>3</sup> to set expectations for this.
- 1.4.4 There is overlap between evaluation and benefits management and they are considered complementary disciplines. They are both concerned with understanding the outcomes of an intervention and how these relate to what was planned. They both use monitoring to track progress with delivering outputs and outcomes. They are both concerned with both accountability for investment expenditure and learning lessons for the future.
- 1.4.5 There are also key differences between the two approaches. Benefits management does not seek to establish how and why the outcomes came about and tends to mainly focus on directly measurable benefits of the intervention, and their maximisation, whereas evaluation can take a broader and longer-term look at impacts and considers to what extent benefits are attributable to the intervention. In terms of profession, benefits management practitioners tend to be project delivery professionals whereas evaluation

---

<sup>3</sup> Guide for Effective Benefits Management in Major Projects (Infrastructure and Projects Authority, 2017).

professionals tend to be analysts. However, these are not absolute or universal distinctions.

- 1.4.6 Given their commonalities, interventions tend to lead with one or the other approach. However, the two approaches are often used in conjunction, with different aims. For example, benefits management can be used to track the delivery of direct benefits of an intervention while evaluation can be used to address wider benefits which are less easily evidenced, such as the effects of the intervention on the wider transport system and the local economy.
- 1.4.7 The remainder of this guide will address evaluation but identify aspects of benefits management practice where they are relevant.

## 2. Evaluation planning

### 2.1 Embedding in delivery

- 2.1.1 It is sometimes the case that an intervention is fully implemented or is coming to an end before evaluation options are properly considered. By that stage, it is usually too late to implement the most rigorous evaluation the intervention could have benefitted from. Much of this guidance is relevant to retrospective evaluations as well. However, such evaluations will typically have significant limitations that could have been avoided by planning and implementing the evaluation alongside the intervention itself. Scheme promoters should be aware that evaluations designed and conducted when the project is coming to an end, or after it has ended, may not be able to provide robust, high-quality evidence of impact.
- 2.1.2 Evaluation is most effective when it is embedded in the delivery of an intervention (rather than being considered an optional 'add-on'). This means that:
- A commitment to conduct proportionate evaluation is established when the intervention is planned. Stakeholders for the intervention accept the obligation to support the evaluation so that it can provide them with the best possible evidence.
  - The body responsible for delivering the intervention also has responsibility for ensuring an evaluation takes place. (Where transport interventions are funded by central government, this requirement is usually a condition of funding.) For interventions that are part of a wider programme of work, having a programme-level framework that sets out quality expectations can



help ensure comprehensive comparable evidence is gathered and programme-level objectives are robustly evaluated.

- A proportionate budget and resourcing plan for the evaluation is included in the Management Case<sup>4</sup> so that the work is planned to proceed as the intervention is delivered. (See section 2.4 for more information on proportionality.) This could include a plan for procuring an external organisation to carry out the evaluation, if it is not going to be delivered in-house. Procurement processes can take several months, therefore early planning is essential.
- Evaluation experts are consulted from the early stages of intervention planning and continue to oversee evidence gathering throughout delivery and after completion.

2.1.3 Evaluation work should be designed to provide independent evidence about the effectiveness of the intervention to those responsible for its delivery and wider stakeholders. Two broad types of evidence can be distinguished:

- **Formative evidence**, which can guide decisions about future adaptations, design and delivery of the intervention. This will apply where interventions are flexible and can be modified 'in flight', or where a pilot intervention is planned to be rolled out in other locations. Such evidence may be provided in an interim evaluation report.
- **Summative evidence**, which can be used to provide accountability for expenditure on the intervention when it is complete (or after key milestones) and can guide the design of similar future interventions.

2.1.4 Whichever types of evidence are obtainable, they should be planned at an early stage so that the evaluation is designed to embed its findings into the process of delivery and decision-making for the intervention and other future interventions. If this is not done early, there is a risk that opportunities to learn from evaluation findings will be missed and the evidence gathered will have limitations.

## 2.2 Building from business case and appraisal

2.2.1 A business case provides justification for undertaking an intervention and an assessment of the expected benefits, costs and risks of the preferred option, including economic analysis in appraisal and modelling to identify its estimated value for money. The business case and appraisal are completed before the intervention is approved. During and after delivery of the intervention, evaluation assesses what benefits, costs and value for money were realised in practice,

---

<sup>4</sup> For guidance on the Five Case Model (strategic, economic, commercial, financial and management), please see Guide to Developing the Project Business Case (HM Treasury, 2018).

ideally providing quantitative evidence to set against the appraisal analysis. Any differences between the values estimated in appraisal and subsequent evaluation provide learning about the impacts of the intervention and may also suggest a need to alter the parameters used in future appraisal and modelling analysis. To maximise this learning, evaluation needs to build from the business case and appraisal.

- 2.2.2 The Management Case should set out what benefits management, monitoring and evaluation will be conducted to assess the effectiveness of the intervention and identify to what extent its intended objectives will have been realised. It is recommended that a proportionate 'Benefits Management & Evaluation Plan' be included as an annex to the Management Case.
- 2.2.3 Appraisal and evaluation require different types of expertise and skill sets. In order for the business case to be as informative and risk-proof as possible, it is recommended that evaluation experts are consulted during the preparation of the business case. Early consideration of the types of evidence that will be required to estimate the impact of the scheme will improve the quality and usefulness of both appraisal and evaluation. (Likewise, good evaluations will also seek to support and improve future appraisal – more on this below.)
- 2.2.4 Analysts supporting business case development should take steps to document and store outputs from appraisal and modelling for the Economic Case so that these can inform subsequent benefits management, monitoring and evaluation activity. This ensures that outturn benefits can be compared with forecasts so that any divergences can be learned from.
- 2.2.5 A good example is the National Highways POPE<sup>5</sup> methodology, which compares information collected before and after the opening of a major scheme against predictions made during the planning process. An appraisal summary table is included in the scheme proposal, with corresponding evaluation summary tables in one-year and five-year post-opening reports facilitating transparent comparisons. Using a consistent methodology for evaluating the impact of similar schemes enables overarching evaluations at portfolio level as well as periodic meta-analyses of previous evaluations to identify general trends.
- 2.2.6 To ensure consistency and continuity between appraisal and later evaluation stages, the Department recommends handover packs are prepared at the time of the appraisal to be used when planning benefits management, monitoring and evaluation activities. (For an example, please see Appendix D in the [TAG guidance for technical project managers](#).)
- 2.2.7 Similarly, the strategic objectives of the intervention, as set out in the Strategic Case, should be stored to inform benefits management, monitoring and

---

<sup>5</sup> Post Opening Project Evaluation

evaluation plans. As recommended by the [Transport Appraisal Process](#), these should include:

- **high-level or strategic outcomes**, typically expressing the desired end state and reflecting the aims and ambitions for the relevant area or population (including wider indirect impacts),
- **specific or intermediate objectives**, typically representing the intermediate effects of the intervention, including the direct and short-term objectives which need to be achieved for the high-level or strategic outcomes to be realised, and
- **operational objectives**, which normally represent the desirable outputs necessary for the intermediate objectives to be achieved

2.2.8 Monitoring and evaluation plans should build on these handover packs to ensure schemes are evaluated according to their intended objectives and outcomes as set out in the original business case and any significant updates to this. The evaluation design and methods must be able to show whether and to what extent the intended objectives and outcomes have been achieved, and provide credible evidence that the outcomes were caused by the intervention itself.

2.2.9 Rigorous evaluations will also explore how the effects of interventions vary across groups (e.g., different groups of users of types of places), reasons behind the extent to which outcomes have been achieved and any important lessons for the future. They will also account for external factors that have influenced results (e.g., the impact of the Covid pandemic on travel patterns) and any changes to the intervention objectives or delivery model over time.

2.2.10 Business cases for large transport infrastructure investments are often developed over a period of months or years. Moreover, the subsequent delivery plan may evolve from what was signed off when the investment was approved. For the purpose of evaluation, where business cases and delivery plans develop in this way, it would not be practical or helpful to examine all iterations when planning evaluation activities. Two iterations have particular relevance:

- The full business case that was signed off when the intervention was approved. This is important for accountability purposes, to show to what extent the assumptions underlying the investment decision proved to be correct.
- The most recent business case or delivery plan at the time when the evaluation is conducted. This is important for learning purposes, to show to what extent the assumptions about the impacts of the intervention at the time of its delivery proved to be correct.

2.2.11 Building evaluation from business case and appraisal in the ways set out above ensures a focus on the extent to which the intervention delivered the outcomes it was funded to deliver. However, good evaluation not only gathers data about the outcomes achieved, but also provides robust evidence that any impact identified can convincingly be attributed to the intervention itself (see section 5).

## 2.3 Piloting and testing

2.3.1 For new interventions or significant changes to existing programmes, it is recommended to pilot the intended delivery model on a small scale to assess its efficacy before deciding whether to roll it out more extensively. It is also useful to test specific delivery options or adaptations before finalising the delivery model for further roll out.

2.3.2 The most reliable way to test an intervention before introducing it at scale is through some form of experiment whereby it is somehow compared with a situation where the intervention does not occur (this is discussed in sections 4.3 and 5). Where these are not feasible, or in combination with these for stronger pilot designs, qualitative methods such as interviews with end users can provide essential information about the likely success (or otherwise) of the intervention in the format being tested.

2.3.3 Comparison areas are often used to test transport interventions. These should be of similar size and background characteristics to the intervention area, far enough away from it not to be directly affected by the intervention, and should have no other intervention occurring there which might invalidate comparisons. For example:

- Bus schemes such as real-time information systems and bus lanes can be tested in a single corridor into a city centre while other corridors retain standard systems so that they can be used as comparisons.
- The introduction of a workplace parking levy in one town centre could be evaluated by comparing what happens in another town centre in the same county which had comparable transport conditions before the intervention but did not change its parking policy.

2.3.4 So long as the intervention and comparison locations can be shown to have similar characteristics, comparisons of traffic data before and after the intervention's introduction in the different sites can produce meaningful evidence of impact. By choosing to introduce interventions in a way which enables comparison areas to be used, local authorities can learn more about the effectiveness of new schemes than if they introduce them across the whole authority at once. This is one of the reasons why evaluation should be planned in parallel with the intervention, from the earliest stages. (See section 5 for more information on counterfactual comparisons.)

- 2.3.5 Another form of testing is to vary the format of the intervention in different locations, with different groups of people or at different times. In such situations the comparison could be between two variants of an intervention as well as, or instead of, between an intervention and no intervention (it may be necessary to caveat such comparisons due to differences in context which limit their validity). Examples in a transport context could include testing variations in junction design, speed limit signage, bus priority schemes, smart ticket pricing and travel information displays. Building such tests into the development of an intervention can enable the outcomes and cost effectiveness of different approaches to be trialled on a systematic basis before a decision is made about which approach will best meet policy objectives.
- 2.3.6 Unless already validated for the intended use and context, data collection tools (e.g., questionnaires, interview/ focus group guides) should also be piloted to ensure they can gather the type and quality of data required for robust evaluation. Pilot data should be analysed using the techniques intended for the main evaluation analysis, and any required adaptations should be made before the main data collection starts.

## 2.4 Budget and proportionality

- 2.4.1 Monitoring and evaluation, as well as benefits management where this is used, should be regarded as core functions in the delivery of an intervention and should be budgeted for from an early stage of planning. Under the HM Treasury business case guidance, they should be developed as part of the management dimension of the business case so that full proposals, including evaluation, are ready by the time the intervention is approved.
- 2.4.2 In developing the plans and budget for evaluation, an assessment should be made about proportionality – in other words, what scale, detail and cost of evaluation will be proportionate to the intervention. An evaluation proposal should set out a rationale for the scale and cost of work that is proposed, which may be influenced by top-down and bottom-up considerations.
- 2.4.3 A common top-down approach to evaluation budgeting, that is most common in small or medium-sized interventions, is a ‘ready reckoner’ which sets the evaluation budget by reference to the overall cost of the intervention. Some authorities use allowances ranging between 0.5% to 5% of intervention budgets for evaluation. While these won’t work in all cases, particularly not for large schemes, the principle of relating evaluation cost to total intervention spend is generally a useful one.
- 2.4.4 Another useful principle is to consider the potential costs of delivery failures which evaluation evidence might help prevent, either for the present intervention or for future interventions of a similar type (see section 4.2, Process evaluation). The greater the opportunity to make better future decisions based on evaluation evidence, the stronger the case will be for setting a larger evaluation budget.

- 2.4.5 Such calculations should be balanced by bottom-up budgeting to assess what level of evaluation expenditure would be necessary to produce good evidence about the extent to which the interventions' objectives have been realised. Where good sources of data already exist, it may be possible to produce robust evidence for a modest budget. Whichever approach is used, it is helpful to keep records of evaluation costs for past interventions, to assist with setting the budget for future ones.
- 2.4.6 It is strongly recommended that evaluation experts are consulted at this early stage, who can advise on the types of evaluation activities that may need to be budgeted for. Depending on the complexity of the intervention, evaluation budgets may need to include bespoke data collection to complement routine monitoring, advanced data analysis methods and specific evaluation expertise that may require procurement. Sufficient time and budget allowances should be made for external procurement and unexpected events that may require changes to the evaluation approach (see section 5.3 for caveats).
- 2.4.7 Apart from the cost of the evaluation itself, factors that should be taken into account when deciding what evaluation design would be proportionate for a particular intervention include<sup>6</sup>:
- **Scale of overall investment:** Interventions benefiting from large amounts of public spending, in particular, require a robust evidence base to support decision making. Unless the causal mechanisms of an intervention are fully understood (section 3) and supported by credible evidence (sections 4, 5, 6), comprehensive evaluation designs will be required.
  - **Strategic importance:** Interventions that are central to the direction of government policy and organisational objectives may arguably justify more comprehensive evaluation than interventions that are less relevant to current policy priorities.
  - **Contribution to the evidence base:** Evaluations that can minimise risk and uncertainty (see [TAG Uncertainty Toolkit](#)) by providing learning to inform current project implementation and future investment decisions may also be prioritised. This consideration will be particularly relevant when the intervention is new or innovative.

## 2.5 Governance

- 2.5.1 Setting up governance structures for an evaluation project can help ensure its success. The following are recommended:
- **A programme board.** This should comprise senior decision-makers for the intervention and should help ensure that they own the evaluation project,

---

<sup>6</sup> Evaluation Strategy and Programme (DfT, 2022)

set its direction, provide it with necessary resources and ensure its findings are used to improve future policies or further roll out. The composition of the programme board should reflect the programme and organisational context.

- **An evaluation steering group.** This would be a more focused, working-level group composed of evaluators, other analysts, representatives of the intervention's programme team and external advisers. This group would meet regularly throughout the life of the project to steer progress, oversee quality assurance and address emerging issues.
- **External peer review.** It is recommended to enlist an independent technical specialist (or team) who can provide quality assurance at key stages such as the evaluation design and reporting.

## 3. Understanding the intervention

### 3.1 The theory of change

- 3.1.1** Evaluation planning requires a thorough understanding of the intervention, the outcomes it is expected to achieve and exactly **how** it is expected to produce these results. This may be referred to as the **theory of change** of the intervention.
- 3.1.2** Developing a theory of change is a critical first step in planning an evaluation as it provides a systematic way of setting the criteria against which the success of an intervention should be judged and the questions that need to be addressed in an evaluation. It also provides a guide to what evidence will need to be collected in the evaluation to enable this assessment.
- 3.1.3** The success of an intervention will also rest on expectations or assumptions in relation to how it is implemented and the wider context in which it is delivered. It is also important to make these broader assumptions explicit when planning an evaluation, as it will be important to test these assumptions when seeking to provide an explanation of why an intervention was successful (or not).
- 3.1.4** A starting point for considering the theory of change of an intervention is to review available documentation of its approach and objectives such as its business case and any subsequent implementation plans. Stakeholders responsible for delivering the intervention or who manage activities that will be affected by it should be consulted and, ideally, involved in developing the theory of change. It is helpful to collect a range of perspectives about how the intervention is expected to work. For this reason, evaluators often hold

workshops with key stakeholders to fully understand and co-develop the theory of change.

- 3.1.5 A theory of change can also reflect economic or social theory about mechanisms of change used by the intervention. It can also incorporate evidence from previous evaluation studies outlining how similar interventions have produced results in other contexts.
- 3.1.6 A variety of individual or mixed approaches can be used to map out how an intervention is expected to work, including:
- **Logic mapping.** This approach is a visual way to represent the theory of change of an intervention, set out as a logical chain running from inputs and activities to outputs, outcomes and impacts (typically depicted in a horizontal/ vertical chart or other formats such as tables). A completed map comprises multiple interacting causal chains. It may also show contextual factors and external influences which are expected to affect the intervention, key assumptions on which the theory of change rests and any identified risks to success. ‘Theory of change’, ‘logic map’ and ‘logic model’ are sometimes used interchangeably, though it is important to note these are not the same<sup>7</sup>. Essentially, a logic map or model depicts the *what* of an intervention (e.g., inputs going in, impacts coming out), while a theory of change goes further to explain *how* the inputs are expected to lead to impacts in a particular context.
  - **Benefits mapping.** This approach is used in benefits management (section 1.4) to identify the benefits that will be realised by an intervention and the organisational objectives they relate to. It works backwards from these end points to set out the stages towards benefits realisation, the drivers that create the need to intervene, the assumed enablers (inputs) and enabling changes that will lead to the expected benefits being realised. A benefits map is typically developed through a benefits discovery workshop which serves as the starting point for the benefits management process.
  - **Systems thinking.** This approach looks at the intervention’s setting as a system with interrelated elements that will influence the intervention and will in turn be influenced by it. For example, a new bus service could be viewed in the context of the local public transport system with other services, timetables and ticket prices. Where the relationships between different elements can be quantified, a systems map may form the basis for developing a model. This approach can support the design and evaluation of the intervention by exploring what potential impacts it might have on

---

<sup>7</sup> For a good overview of theories of change and their relationships to other visual and organisational tools, please see Dhillon, L. and Vaca, S. (2018), Refining Theories of Change, *Journal of Multidisciplinary Evaluation*, 14(30), 64-87, [https://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/496/444](https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/496/444)



existing elements in the system. It is complementary to the two approaches above which focus on the intervention itself.

3.1.7 These three approaches, which may be combined within the same project, have some common characteristics:

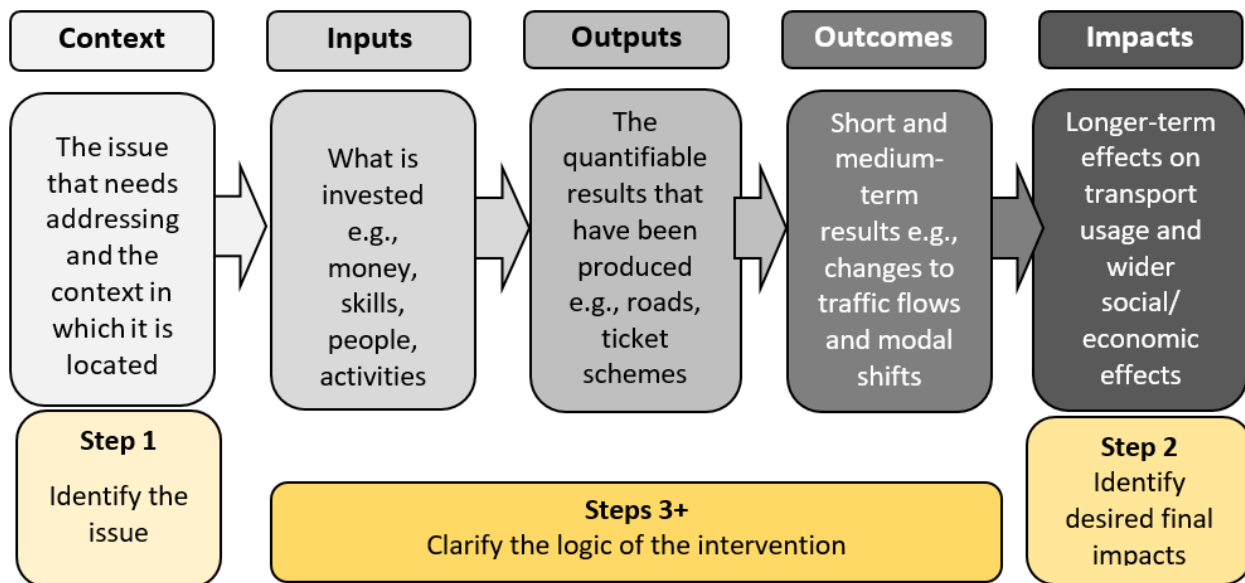
- They all use visual representations of the intervention and/ or the context in which it operates.
- They are informed by whatever evidence and theory exists about the type of intervention and the workings of the system into which it is introduced, which forms the starting point for the approach. (Case study evidence from previous interventions is key in this area.)
- They are typically developed interactively through engagement with stakeholders for the intervention, often in one or more workshops.
- They are typically used when an intervention is being developed but can be updated iteratively to reflect changing understanding of how the intervention and/or the system works.
- They can support both the development of an intervention and the development of a plan for how to evaluate its impacts or measure its benefits, by facilitating clarity about the outcomes of interest and relevant data sources.
- They can provide a basis for risk management by showing factors which may be barriers to success and counteract the realisation of benefits. They can support a scenario illustrating how an intervention might fail (which can also be described as a 'negative theory of change' or 'pre-mortem' in project delivery). This is very useful for pre-empting risks and negative unintended consequences.

## 3.2 Logic mapping

3.2.1 In evaluation, the most common representation of the intervention logic is the logic map or logic model. Further information on this topic can be found in 'Logic mapping: hints and tips' (DfT, 2010).

3.2.2 Figure 1 shows the typical components of a logic map, though there can be variations in how these are depicted (for example, the context is not always included in the visual representation, inputs and activities can be displayed separately, outcomes and impacts can be split into short, medium and long-term). Although a logic map is often depicted as progressing from left to right, the steps in developing it follow a different sequence, as the bottom of Figure 1 suggests. It is useful to start developing the model by considering the issue being addressed, then considering the impacts or change that the intervention

is intended to achieve before working backwards through the steps in the logic chain required to achieve this desired change.



**Figure 1 Main components of a logic map**

- 3.2.3 A clear theory of change and/ or logic map agreed and ideally co-developed with key stakeholders can be a valuable tool for deciding what metrics to monitor before, during and after the delivery of the intervention. Consistent data related to baselines (section 7.2), inputs, activities and outputs will play a critical role in evaluating the outcomes and impacts of the intervention.
- 3.2.4 By planning both monitoring and evaluation from the start in line with the theory of change, we can ensure that these are coherent, consistent and together can provide a comprehensive picture of how the intervention was delivered, what difference it has made (if any), and what useful lessons can be learnt for the future. (For more information on monitoring data, please see section 7.3).

## 4. Evaluation design

### 4.1 Evaluation questions

- 4.1.1 There are three main types of evaluation activity, each aiming to answer slightly different but complementary questions, often within the same evaluation:

- **Process evaluation:** What can be learnt from how the intervention was delivered?
  - Was the intervention delivered as intended?
  - What worked well, or less well, for whom and why?
  - What could be improved?
  - How has the context influenced delivery?
- **Impact evaluation:** *What difference has the intervention made?*
  - Did the intervention achieve the expected outcome? To what extent?
  - How exactly did the intervention cause the observed impact?
  - To what extent can the difference be attributed to the intervention?
  - What would have happened anyway (without the intervention)?
  - What unintended consequences did the intervention have (positive or negative)?
- **Value-for-money evaluation:** *Was this a good use of resources?*
  - How cost-effective was the intervention (compared to alternatives and compared to doing nothing)?
  - What was the value-for-money of the intervention?
  - How does the ratio of costs to benefits compare to that of alternative interventions?
  - How economic, efficient, effective and equitable was the intervention?

4.1.2 Evaluation questions should be chosen in line with the aims of the intervention and any priorities of the evaluation itself, such as addressing a specific evidence need. It is important that evaluation questions are referred to throughout the evaluation and used to guide the evaluation design, data collection tools, analysis techniques and report content. Any initial evaluation questions that were only partially answered or indeed remained unanswered due to data availability, quality or for other reasons, should be flagged as such in the reports. This can help identify areas where further evaluation evidence is required, including within the intervention theory of change.

4.1.3 This section outlines how these types of evaluation may be used for transport interventions. (Please see Magenta Book for additional generic guidance.)

## 4.2 Process evaluation

4.2.1 Process evaluation is particularly useful for transport interventions where there is a need to learn from how the intervention was delivered which is best met through evaluation research. This condition will not necessarily be met for all interventions. It is most likely to be met where an intervention is not fully developed and where the implementation approach is novel or may change considerably in future. It is less likely to be met if an intervention follows a tried

and tested process. For example, a local authority would not necessarily conduct process evaluation into every instance of constructing new cycle lanes if the same process were used for all lanes and contexts were sufficiently similar (whereas it may take steps to measure the impacts of introducing all lanes on levels of cycling and other traffic – see section 4.3).

4.2.2 Two examples of recent DfT process evaluations illustrate where evidence about delivery can be useful.

- Transforming Cities Fund. When a new ‘co-development’ process was used for the development of local authority business cases for this programme in 2019, the Department commissioned an independent contractor to carry out a process evaluation by reviewing documents and interviewing officials from DfT and authorities. The rationale for conducting this study was that this was a new method of developing programmes and so it was important to understand how the process was experienced by all parties, to guide whether and how this approach might be used for future programmes.
- Transport Open Data Projects. Between 2017 and 2020, DfT undertook three programmes which had the common aim of opening up data about transport to support new business applications. These were the Bus Open Data Service, Street Manager, which used data about roadworks, and a programme to open up local authority data about traffic management and parking. Since these projects were new and had some common features in terms of the nature of the technical and stakeholder communication challenges, it was decided to commission an independent process evaluation to learn from their implementation and capture lessons for these and future data projects.

4.2.3 In both these examples, new processes were being introduced and there was a need to learn about their implementation. This is a necessary but not a sufficient condition for process evaluation. It is also necessary to consider whether there is an alternative process which may facilitate learning from implementation. A further consideration is whether it is feasible to get the right level of co-operation from stakeholders to carry out a process evaluation.

4.2.4 In the case of major infrastructure projects, there are well-established processes for project [assurance reviews](#), such as those published by the Infrastructure and Projects Authority. Such reviews will receive strong engagement from project teams as an established part of the delivery process. In such instances, a process evaluation may seem to stakeholders to duplicate project reviews, which would in turn affect co-operation levels. This is not to say that process evaluations can never add value to project review processes, just that it will be necessary to articulate what value they can add. In the above examples, these project review mechanisms were not present so it was relatively straightforward to make the case for a process evaluation as an

exercise which could provide implementation lessons that could not be collected otherwise.

4.2.5 Where process evaluations are conducted, it will be helpful to start by considering the intervention's theory of change. A good theory of change should set out how the intervention is believed to work and process evaluation can provide independent evidence about the extent to which the mechanisms in the theory operate in the way expected. Process evaluation will also help identify any implementation issues which may need to be addressed for the intervention to operate more effectively. By examining an intervention's assumed causal mechanisms, process evaluation provides a useful complement to impact evaluation. It may also help explain why anticipated impacts were not in fact realised or why unintended consequences may have occurred. Process evaluation can also play an important role in establishing whether any targets were met (and, if not, why) and whether projects were delivered on the expected budget and scale.

4.2.6 Process evaluations commonly use a mix of quantitative and qualitative data collection. They may use monitoring data, reviews of project documents and surveys or qualitative interviews with delivery personnel and other stakeholders who use, benefit from, or oversee the intervention. Analysis of this information requires objectivity and distance from the project which is why process evaluation should ideally be conducted by independent researchers (external suppliers or in-house analysts not involved in the project).

### 4.3 Impact evaluation

4.3.1 It is always important to understand what difference a publicly funded intervention has made, and transport policies are no different. In particular, transport infrastructure schemes are expensive and cannot easily be adapted or reversed once they have been implemented. Significant effort is therefore expended in appraising what benefits the interventions will bring about, through a business case. Impact evaluation will help assess the degree to which we have been successful against the original objectives or success criteria outlined in the business case, and the extent to which the outcomes are demonstrably caused by the intervention.

4.3.2 While appraisals and well-considered business cases are essential, these are typically based on assumptions about how the intervention may work. It is only by collecting and analysing data before, during and after an intervention (wherever possible) that we can draw robust conclusions about the actual impact the intervention has had. Good impact evaluations can also provide reliable evidence of attribution (see section 5), which appraisal or benefits management cannot provide.

4.3.3 Good impact evaluations will also seek to show whether that impact varied across different areas or for different groups, whether specific factors influenced

the strength of the impact, and which aspects of the intervention are essential for obtaining the same impact in the future. Combined with process evaluation, good impact evaluation will also explain why the intervention, or some of its aspects, did not produce the results intended, so useful lessons can be learnt for future delivery and similar policies. Ideally, it will also explore whether variations or alternatives might have been more effective or would be likely to have provided better value for money.

- 4.3.4 With early planning, good-quality data and strong links between the appraisal and the evaluation stages (see section 2.2), impact evaluation can also contribute valuable data to help refine appraisal assumptions and the modelling of different scenarios when preparing future business cases.
- 4.3.5 The design of an impact evaluation for a transport intervention should begin with an understanding of its business case and assumptions made about benefits. To have accountability for the use of public funds, we must understand to what extent the planned and appraised benefits were realised in practice. The resulting evidence will then help strengthen future business cases for similar policies and contribute to a culture of learning and evidence-based decision making.
- 4.3.6 An impact evaluation should not confine its scope to addressing the benefits that have been planned and appraised. In the interest of deriving the right learning to inform the future development of this and other transport interventions, we also want to understand what other changes the intervention may have brought about (often referred to as unintended consequences or unforeseen/ unanticipated effects). These other effects could be benefits or disbenefits.
- 4.3.7 The implication for evaluation design is that it needs to be capable of picking up such effects. This entails monitoring a range of measures which may not necessarily be expected to change, but which are clearly relevant in the context of the intervention (see section 3.1.6). For example:
- When evaluating a change to a road junction that is designed to reduce accidents, as well as monitoring accidents, traffic flow should also be monitored to check that there is an impact on congestion.
  - For a clean air zone that is intended to improve air quality, the effect on traffic volumes and air quality in neighbouring areas should also be monitored to identify any spillover or traffic diversion effects.
- 4.3.8 Impact evaluation can be approached in two main ways (sometimes combined within the same project):
- **Quantitative counterfactual approaches:** These types of evaluation make comparisons between the beneficiaries of an intervention and a similar group

that did not receive the intervention. These approaches typically provide the strongest, most objective evidence of impact (see section 5, The counterfactual, including caveats).

- **Theory-based approaches:** In conjunction with these, or when these are not feasible, predominantly theory-based approaches can be used, such as qualitative comparative analysis, contribution analysis or process tracing (see Magenta Book for further information).

4.3.9 Either way, specific evaluation expertise is required to plan, design, implement and report good-quality impact evaluation. For this reason, it is essential that independent evaluation experts are consulted and involved from the intervention planning stage.

4.3.10 Examples of recent impact evaluations of transport interventions include:

- Driver 2020, a research programme trialling five interventions aiming to identify effective training and technology options in reducing the risk of collision for learner and novice drivers. The study uses a randomised encouragement design, where participants randomly assigned to the intervention or control groups choose whether or not to engage with their intervention, thus creating more realistic conditions to understand impact if rolled out more widely.
- The economic impact of new and improved rail lines study<sup>8</sup>. This used a quasi-experimental design under which time series data for a location served by a rail intervention were compared with data for a matched location which had no intervention, to produce an estimate of the intervention's effects.
- Using a theory-based approach to evaluate the impact of Roads Reform (changes made to the management and operation of the Strategic Roads Network). The multi-wave evaluation uses stakeholder interviews, telephone surveys of supply chain operators and extensive document reviews, among other data sources.

## 4.4 Value-for-money evaluation

4.4.1 Value-for-money (VfM) evaluation builds on and complements impact evaluation by assessing whether the intervention represents optimal use of public funds to achieve the intended outcomes. In a transport context, VfM is concerned with creating public value through **economic** (e.g., travel time, vehicle costs, tax revenues), **social** (e.g., health, safety, accessibility) and **environmental** (e.g., noise, air quality, landscape) impacts. (See DfT's [Value for Money Framework](#).)

---

<sup>8</sup> Economic impact of new and improved rail lines: case studies and method - GOV.UK ([www.gov.uk](http://www.gov.uk))

- 4.4.2 The National Audit Office<sup>9</sup> (NAO) uses three criteria to evaluate the VfM of public spending which, taken together, represent the cost-effectiveness of an intervention:
- **Economy**: minimising the cost of resources used or required to meet objectives (inputs) – *spending less*;
  - **Efficiency**: the relationship between the output from goods or services and the resources to produce them – *spending well*, and
  - **Effectiveness**: the relationship between the intended and actual results of public spending (outcomes) – *spending wisely*.
- (A fourth criterion is sometimes applied: **Equity**: the extent to which services are available to and reach all people that they are intended to – *spending fairly*.)
- 4.4.3 The Magenta Book highlights two common methods that focus on the social value and social cost of a public policy or programme:
- **Social cost effectiveness analysis**, which compares the overall costs and benefits of different ways of producing similar results, and
  - **Social cost-benefit analysis**, which quantifies and monetises overall costs and benefits, typically including broad financial, environmental and social impacts.
- 4.4.4 Evaluating the actual VfM of an intervention should take into account the methods and projections from appraisal stage, which may have included cost-benefit analysis, economic, environmental, social and distributional impacts (see [TAG Units A1-5](#)).
- 4.4.5 VfM assessment can provide a useful tool for benchmarking the intervention against potential comparator programmes, particularly those intended to produce similar outcomes.

---

<sup>9</sup> <https://www.nao.org.uk/successful-commissioning/general-principles/value-for-money/assessing-value-for-money/#>



## 5. Focus on impact: The counterfactual

### 5.1 What is the counterfactual and why does it matter?

- 5.1.1 This section provides further detail on how to assess the extent to which an intervention has made a difference in practice, and gather robust evidence that any improvements (or otherwise) can be reliably linked to the intervention itself and would not have been expected to occur without it. The section is therefore most relevant to impact evaluation, though the principles outlined here will be generally useful when planning most types of evaluation.
- 5.1.2 Outcomes and impacts that would have been expected to occur in the absence of the intervention ('do nothing', 'do minimum' or 'business as usual') are sometimes described as **deadweight**. Deducting deadweight from the overall estimated impact will produce the net additional impact of the intervention (referred to as **additionality**). Counterfactual analysis is a robust way to produce evidence of additionality.
- 5.1.3 Counterfactuals are credible estimates of what would have happened in the absence of the intervention and, as such, are critical to establishing **causality** and **attribution**. For instance, let's assume a local authority introduces free electric vehicle (EV) chargers in community car parks. An increase in EV sales is observed in the city six months later. Can we be confident that the free EV chargers are the cause of the sales increase? Can we assume that the higher EV sales can be attributed to the introduction of free EV chargers? Not unless we have a good indication of what would have happened without the free EV chargers – that is, a good counterfactual comparison.
- 5.1.4 Counterfactuals include control groups (created through random allocation), comparison groups (created through quasi-randomisation or matching on key characteristics) and other comparisons (e.g., before/after). More information on these approaches is provided in section 5.2.
- 5.1.5 When choosing the right counterfactual approach and data sources, it is important to consider whether the desired impact might also be felt beyond the target group/ area or whether any unintended consequences can be expected ('**spillovers**'). Using the example above, free EV chargers in one city may increase EV sales in areas beyond the city itself (positive spillover). Likewise, free EV chargers in community car parks may encourage more people to drive EVs instead of walking (negative spillover if the intention was to reduce transport costs or environmental emissions).
- 5.1.6 It is also important to consider whether an improved outcome for the target group or area may result in a corresponding deterioration somewhere else

(‘**displacement**’). For example, free EV charging in community car parks will transfer charging costs from drivers to local authorities or car park operators. Analysing what would have happened without the intervention must take account of such secondary outcomes in order to estimate its overall impact and value for money.

- 5.1.7 Other effects that should be taken into account when estimating the net impact of an intervention through counterfactual analysis include **agglomeration** (benefits that may arise through close location of businesses and/ or people), **multipliers** (e.g., further economic activity associated with additional local income), positive and negative **externalities** (unintended or indirect consequences)<sup>10</sup>.

## 5.2 Common counterfactual and comparative approaches

- 5.2.1 **Randomised controlled trials** (RCTs) are generally regarded as the most robust impact evaluation method. This approach consists of randomly allocating individuals, institutions or locations who could receive the intervention into two groups (typically), one of which receives the intervention (the ‘treatment group’) and one of which does not (the ‘control group’). Provided an adequate sample size is available, correct randomisation procedures are followed and all other influencing factors are carefully controlled, any difference in outcomes between the two groups can be attributed to the intervention. (That is, there should be no pre-intervention differences between the two groups because they have been allocated at random.) This is rarely feasible for transport interventions due to the challenges of randomisation, implementation fidelity and controlling all other factors in real life, as well as the practical and ethical issues of deciding ‘at random’ which individuals, institutions or locations receive a transport intervention and which ones do not.
- 5.2.2 A **staggered roll out, pipeline or stepped-wedge trial** occurs when an intervention is rolled out in waves or ‘steps’, or continuously over time, creating the opportunity to compare outcomes for a group that receives the intervention early with one receiving the intervention later. Because everyone eventually receives the intervention, findings are less likely to be biased by differences between those that do and do not receive the intervention. For example, a form of staggered roll out occurred naturally in the Bikeability cycle training programme for primary school pupils. Some schools opted to receive the programme when their pupils were in year 5, while others waited until year 6. Comparing changes in cycling behaviour of year 5 students across these two types of school will give a reasonable measure of the impact of the training programme (as the school’s choice of which year to receive training was unrelated to the prior cycling behaviour of pupils).

---

<sup>10</sup> See Green Book (2022) and Additionality Guide (Homes and Communities Agency, 2014).

- 5.2.3 **Regression discontinuity design (RDD)** is a robust quasi-experimental method that is useful when randomisation is not possible or feasible. It consists of assigning participants (or other units of analysis) to the intervention or control conditions based on a cut-off point on a particular measure of interest. For example, the impact of daylight savings time on the volume of road traffic collisions can be estimated by analysing collision data immediately before and after the clocks changing. Another example could be looking at the impact of travel infrastructure upgrades on shopping centre footfall, where centres with footfall above or below a pre-determined cut-off volume would be assigned to intervention or control. Provided a number of analytical assumptions are met, changes in footfall volumes between shopping centres just above and just below the cut-off can be attributed to the infrastructure upgrades.
- 5.2.4 **Difference-in-differences.** A difference-in-differences (DiD or diff-in-diff) approach requires two groups (or locations, institutions etc.) – an intervention group and a comparison – as well as measuring the outcomes of interest before and after the intervention. The before/ after difference in the intervention group is compared to the before/ after difference in the comparison group, with any *difference between the two differences* being attributed to the intervention effect. For example, a bus company wishing to redistribute passenger numbers more evenly throughout the day may test the effectiveness of on-board messaging on subsequent passenger flows. They may decide to implement the messaging in multiple towns in area A and select a set of otherwise similar towns in area B for comparison, where no such messaging will be implemented. They start by measuring passenger flows in both areas under similar conditions (e.g., same method, same time, same day of the week). Then they introduce the on-board messaging to the towns in area A only and, after a pre-determined period, measure passenger flows again in both areas. Assuming similar numbers of passengers would normally have been expected in both areas in response to various external influences, a change in passenger distribution in area A which cannot be observed in area B towns would represent the effect of the messaging intervention.
- 5.2.5 The assumption above is important. If we cannot be sure that the two areas would respond similarly to factors unrelated to the intervention (e.g., weather, policy changes, timing of data collection), it is likely the areas are not sufficiently similar for a DiD analysis. Such a counterfactual comparison would therefore be inappropriate. This can be avoided by making sure the area selected for comparison is similar to the intervention area in all key aspects expected to influence the outcome (or similar enough, with remaining key differences controlled statistically). One way to check similarity is to collect outcome data more than once before the intervention is introduced. If numbers in the two areas (or groups etc.) show a parallel trend before the intervention but follow a different trajectory after the intervention, the difference would normally represent the effect of the intervention.

- 5.2.6 A key advantage of the DiD approach is that it is robust to unobserved differences between groups that do not change with time (i.e., are fixed properties of the respective groups). This is important, as it will clearly not be feasible to collect data about any possible difference between the intervention and comparison groups, which may or may not be known.
- 5.2.7 Although before/ after differences between groups are sometimes presented as simple numerical or visual comparisons, it is important to note that such basic comparisons are not able to provide strong counterfactual evidence in themselves. DiD counterfactual analysis is a fairly sophisticated statistical procedure which should be conducted by expert evaluators or statisticians.
- 5.2.8 **Matched comparisons.** A similar approach, often used in conjunction with DiD, can be applied at a much granular level by zooming in on individual participants<sup>11</sup> within the intervention and comparison groups. Each participant in the intervention group, area etc. can sometimes be matched with an equivalent non-participant on specific variables that are expected to influence the outcome. (Note this is an advanced analytical method where the matching is executed using statistical software – not to be confused with comparative case studies or other qualitative comparisons.) Depending on the quality of the data and the accuracy of the matching, observing the outcome measure in the two groups/ areas can identify differences attributable to the intervention effect. For example, a local authority may decide to encourage shifting from driving to cycling using evidence-based methods. They decide to pilot a new cycling scheme in five towns, choosing five other towns for comparison (without the cycling scheme). They want to make sure that any increase in cycling in the participating towns can be reliably attributed to the intervention.
- 5.2.9 It is important to ensure that the town pairs that will be used for comparison are as similar as possible, except that one will receive the intervention and one not. The team will need to consider carefully the key factors that are likely to influence cycling and driving uptake: for example, the quality and coverage of cycling infrastructure, population density and affluence, climate, relevant local policies and so on. Each of the five participating towns will be matched with a comparison town that will be as similar as possible in these key respects. The outcomes of interest (e.g., automobile/ cycling traffic, estimated mileage travelled by car/ bicycle) will then be measured before and after the introduction of the new cycling scheme, as many times as feasible and proportionate. If the town pairs are correctly matched and no significant changes occur during the project, an increase in cyclists/ miles cycled in the intervention towns only will provide good evidence that the new cycling scheme is achieving its aims.
- 5.2.10 A limitation of matching is that it is only possible to match on the observed characteristics of areas, and it is obviously impossible to know if there are

---

<sup>11</sup> Depending on data availability and the level at which the intervention is aiming to make a difference, matching can be applied to individual people, clearly defined groups or other analytical units of interest.

unobserved differences. This limitation is often addressed by combining matching with DiD.

- 5.2.11 **Before/after comparisons.** For interventions where a suitable control or comparison group cannot be identified (e.g., national policies), a simple comparison of data collected before and after the intervention is sometimes used. For example, a new bridge is built between two areas of a town aiming to improve attendance at town hall meetings. A simple before/ after comparison would compare town hall meeting attendance before the bridge is built to meeting attendance after the bridge is built.
- 5.2.12 Unless the system is closed to external influences, single before and after measurements do not provide compelling evidence even if a difference is identified, given the multitude of factors that could have influenced the outcome of interest. In the example above, simply measuring attendance before and after building the bridge and attributing any change to the bridge itself disregards a whole range of different reasons why people may or may not attend town hall meetings (e.g., professional commitments, caring responsibilities, meeting time and duration and weather conditions).
- 5.2.13 There are several ways to strengthen the evidence in such a case. One way is by adding more data points. If an outcome measure of interest follows a stable trajectory for some time only to change after the intervention has been introduced, this may indicate an effect of the intervention on that specific outcome measure. The strength of the evidence can be further improved by including additional methods (e.g., exploring alternative explanations, re-examining the theory of change, using different data types and sources), until a credible defensible explanation is reached that can help with any future iterations of the intervention.
- 5.2.14 It is important to note that pre/post comparisons will not in themselves be able to account for wider changes in society and the economy, which can affect the success of transport schemes. For example, if an economic downturn leads to higher unemployment and a decrease in commuting, low 'after' figures on the use of a new bus route could lead us to conclude that the scheme has failed to meet its objectives, when in fact bus use might still be higher than it would have been without this intervention in place. Additional methods are required to ensure a fair, robust assessment of the actual difference that an intervention has made in practice.
- 5.2.15 **Interrupted time series** analysis is a type of before/ after comparison that is useful when a large amount of data can be collected consistently over a period of time to capture any fluctuations in the outcome of interest after the introduction (and the end) of an intervention. The approach is useful when evaluating national policies and when we wish to explore the effects of an intervention over time, though it should be noted this approach shares the

limitations of all before/ after designs. Examples include the effect of traffic calming measures or speed limits on the frequency of road incidents.

- 5.2.16 A type of interrupted time series called ‘ABA design’, which consists of introducing, discontinuing and reintroducing the intervention, can be a good way to test whether any fluctuations in the outcome of interest appear to be related to the intervention. Variable speed limits are a common transport intervention that lends itself easily to this type of analysis.
- 5.2.17 These and similar approaches are not mutually exclusive and can be used in combination within the same evaluation in order to strengthen the evidence gathered.

### 5.3 Caveats

- 5.3.1 With all counterfactual and comparative analyses, it is essential to consider any **significant changes** that may have influenced one side of the comparison but not the other, which will affect any impact claims in the absence of additional analysis and transparent reporting.
- 5.3.2 An example would be an evaluation testing the effectiveness of traffic restrictions on air quality in one region compared to air quality in a similar region where no traffic restrictions are introduced (control). An unrelated tax on highly polluting vehicles introduced to the control area during this evaluation will influence the air quality outcomes and affect the validity of the intended comparison. Similarly, opening a new industrial hub in either the intervention or the control area will influence the outcome in ways that may be virtually impossible to disentangle from the impact of the intervention being evaluated.
- 5.3.3 Similarly, before/ after analyses can be significantly affected by major changes occurring between relevant data collection points. Two recent examples include the 2008 recession and the COVID-19 pandemic, both of which caused significant changes in transport use and therefore affected the validity of analytical comparisons over the relevant periods.
- 5.3.4 These analytical limitations can be mitigated by carefully planning the evaluation design and data source selection, using additional statistical analyses that can recover some of the validity lost due to unexpected changes, and reporting the methodology and caveats transparently in any evaluation outputs.
- 5.3.5 It is equally important that any changes to the delivery model made while implementing the intervention are informed by robust interim findings. Mid-delivery changes informed by emerging data can usually be accounted for in the analysis without compromising the quality of the evaluation. One of the benefits of embedding evaluation in delivery is that it can facilitate evidence-based decisions about changes that might be necessary in order to give the

intervention the best chance of success. 'Tweaking' the intervention without properly considering implications on evaluation should be avoided.

- 5.3.6 When selecting counterfactuals, care should be taken to avoid **selection bias** (ensuring the comparison is fair and free from influences that cannot be accounted for in the analysis) and **reverse causality bias** (confusing cause and effect – e.g., using an earlier example, a local authority might introduce free EV chargers *as a result* of an increase in EV sales, rather than it being the cause of higher EV sales).

## 6. Measures of evaluation quality

- 6.1.1 In order for the evaluation itself to provide good value for money (i.e., good use of public funds), it is important that it is planned, resourced and executed in ways that meet established quality criteria. The Magenta Book describes a good evaluation as one that is useful, credible, robust and proportionate:

- **Useful:** A high-quality evaluation is designed to meet the needs of relevant stakeholders, including end-users, policy and decision makers, public accountability bodies, government and taxpayers. The outputs of a good evaluation are accessible, useful and timely. They communicate clearly the value of the project, lessons learnt and the limitations of the evaluation, to ensure the results are used responsibly.
- **Credible:** For an evaluation to be credible, it needs to be transparent, fair and objective. This relies on the project team making a deliberate effort to identify biases, assumptions and unrealistic expectations of what the project and the evaluation can achieve. Wherever possible and proportionate, the evaluation should be conducted by independent evaluators. If independent evaluators are not used, independent evaluation experts should at least guide and peer review the evaluation design and outputs.
- **Robust:** A good evaluation is well designed, well executed and well reported, with the results being proportionate in relation to the methods used. This includes evaluation questions aligned to project objectives, rigorous sampling strategies to ensure stakeholder views are adequately represented, analysis methods that maximise learning and avoiding unsupported claims. A robust evaluation also adheres to relevant ethical and data protection principles.
- **Proportionate:** The complexity of the evaluation needs to mirror the scope, budget, risk, policy priority and learning potential of the project or programme itself. For low-risk, low-priority interventions that already have a good evidence base, light-touch monitoring and evaluation may be sufficient. High-

risk, high-priority interventions breaking new ground and emphasising learning and improvement are likely to require comprehensive evaluations.

6.1.2 Well-established frameworks focussing specifically on the rigour of impact evaluation, often used by HM Treasury and the Cabinet Office to assess the quality of the evidence, include:

6.1.3 The **Nesta Standards of Evidence**<sup>12</sup>, which operate on a scale of 1 to 5, where Level 1 represents little more than a clear articulation of why the intervention is needed, and why it is better than the current situation, with Level 5 providing robust evidence that the intervention can be delivered at scale with demonstrable positive impact. The standards authors summarise the five levels as follows:

- Level 1: You can describe what you do and why it matters, logically, coherently and convincingly.
- Level 2: You capture data that shows positive change, but you cannot confirm you caused this.
- Level 3: You can demonstrate causality using a control or comparison group.
- Level 4: You have one or more independent replication evaluations that confirm these conclusions. [That is, the same evaluation design has led to similar results in different contexts.]
- Level 5: You have manuals, systems and procedures to ensure consistent replication and positive impact.

6.1.4 Similar classifications follow the **Maryland Scientific Methods Scale**, such as the version developed by the What Works Centre for Local Economic Growth, summarised below:

- Level 1: Either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group and with no control variables.
- Level 2: Use of adequate control variables and either (a) a cross-sectional comparison of treated groups with untreated groups, or (b) a before-and-after comparison of treated group, without an untreated comparison group.
- Level 3: Comparison of outcomes in treated group after an intervention, with outcomes in the treated group before the intervention, and a comparison group used to provide a counterfactual (e.g., difference in difference).

---

<sup>12</sup> [https://media.nesta.org.uk/documents/standards\\_of\\_evidence.pdf](https://media.nesta.org.uk/documents/standards_of_evidence.pdf)



- Level 4: Quasi-randomness in treatment is exploited (section 5.1.4), so that it can be credibly held that treatment and control groups differ only in their exposure to the random allocation of treatment.
- Level 5: Reserved for research designs that involve explicit randomisation into treatment and control groups, with Randomised Control Trials (RCTs) providing the definitive example.

6.1.5 All evaluations should aim for the highest standards of quality possible in the context of relevant practical constraints of public policies, including feasibility, ethics, data availability, proportionality and cost (see section 2.4).

## 7. Data requirements

### 7.1 Existing data

7.1.1 To complement any internal (unpublished) datasets, best use should be made of data already available either in project documentation or in the public domain. This is in order to maximise the value of the existing data for the taxpayer, but also to minimise the 'research burden' placed on busy project teams and wider stakeholder groups. In addition, using data that is publicly available facilitates transparency, potential replication of findings in other contexts or by other teams, and comparisons across similar projects. This will also generally be more reliable, although reporting lags are typically a significant limitation. However, with appropriate caveats, older data can still be relevant if no significant developments have taken place that are expected to have influenced the data in the meantime.

7.1.2 **Transport statistics** are available both in transport specific data releases and in more general sources. The Department maintains a searchable [database of transport data](#) which can be filtered by mode, location and time period among other options. This includes national statistics on electric vehicle [registrations](#) and [chargepoint provision](#), as well as comprehensive [transport statistics for Great Britain](#) and specific [statistical collections](#). Other publicly-available data includes [national rail statistics](#) and datasets curated by the [Office for Rail and Road](#). The national census provides [searchable data tables](#) about the population at national and local levels, including mobility, origin-destination flow data and car ownership. The [National Trip End Model](#) forecasts the growth in trip origin-destinations up to 2051 for use in transport modelling, including population, employment, car ownership and travel demand.

7.1.3 **Economic administrative data.** Evaluating the impact of transport interventions often requires estimating wider economic benefits such as

employment, economic activity and economic growth. Useful data sources include the Office for National Statistics (ONS) [Nomis](#) database, which provides up-to-date statistics on the UK labour market, the [Annual Population Survey](#), the [Business Register and Employment Survey](#) and various datasets on regional [gross value added](#). The [What Works Centre for Local Economic Growth](#) publishes regular briefings and evidence reviews, including a useful guide to [using data for local economic policy](#).

- 7.1.4 **Transport user surveys** can provide helpful context on user perceptions, attitudes, satisfaction and demand. Established relevant studies include the [National Travel Survey](#) and the [National Travel Attitudes Study](#). Transport Focus runs multiple tracker surveys of [passenger and road user attitudes](#), including buses, rail, strategic roads, trams, logistics and motorway services users. Data on travel behaviour and attitudes during the coronavirus pandemic can be found in the [All Change? Travel Tracker](#).
- 7.1.5 The Department has published a [transport survey question bank](#), including information on sources, response options and coding guidance. This can provide a useful starting point for developing bespoke surveys or adding transport-related questions to more general questionnaires.

## 7.2 The baseline

- 7.2.1 Good-quality baseline data on the outcome measures of interest, collected before the start of the intervention, is essential for robust evaluation, particularly when randomisation is not feasible. In order to estimate the impact of an intervention on a specific area, group or process, we need to know the situation at the start, so any gains or otherwise can be deducted.
- 7.2.2 Relying on stakeholders' memories of the situation at the start of the project or attempting to construct baselines retrospectively (especially if not using reliable secondary data) will inevitably have limitations. This is why it is essential to plan evaluation from the start, in parallel with planning the intervention itself. Following a structured approach to benefits management and evaluation, and ensuring this is set out early in the business case development process will ensure that baselines are adequately considered before project delivery begins.
- 7.2.3 Deciding what data to collect at baseline will require careful consideration of the intervention theory of change and/ or logic map. Wherever possible, consistent metrics should be collected at baseline, (during the intervention – if feasible) and after the end of the project. For example, if the intended outcome of a project is a reduction in traffic noise, the level of traffic noise should be measured before the intervention starts, during and after the intervention. Any significant events that are expected to influence the outcomes should also be documented and, where possible, accounted for in the analysis.

- 7.2.4 Baselines other than the outcome measures can be included in statistical analysis to improve the precision of impact estimates. Using the example of traffic noise, confidence in the impact estimate of a noise-reduction intervention could be improved by including factors such as population density or major road proximity. Analysis experts should be consulted from the early planning stages to ensure adequate baselines are collected at the right time.

## 7.3 Monitoring data

- 7.3.1 In addition to monitoring progress indicators as part of day-to-day delivery, project teams should ensure they are preparing the ground for robust evaluation at appropriate later stages. This includes implementing a data collection plan agreed at the start of the project, with data collected consistently in the same format at specific time points. Wherever possible and relevant, light-touch monitoring should continue after project completion to enable long-term impact tracking.
- 7.3.2 Ideally, monitoring data will include numerical data related to the outcome measures of interest and key factors expected to influence them, which will be essential for estimating the impact and value for money of the intervention at a later stage. These should be complemented by qualitative, contextual data which will facilitate process evaluation and lessons learnt about the implementation model. Deciding at the early planning stages what types of data will be required will facilitate more accurate costing (e.g., qualitative data collection and analysis is typically more expensive than collecting and analysing quantitative data).
- 7.3.3 Good monitoring data can facilitate adaptations and adjustments to maximise the effectiveness and value of the intervention before its completion. This is one of the key benefits of embedding formative evaluation in project delivery. However, care must be taken to document all changes and the data on which they were based, so these can be taken into account when estimating the overall effectiveness and impact of the intervention. Changes to the delivery model should only be made after careful consideration of any implications on evaluation and the quality of the evidence gathered (see section 5.3).

## 7.4 Other bespoke data

- 7.4.1 Good monitoring will enable good evaluation but, while monitoring can show what outputs have or have not been achieved, it cannot provide evidence of outcomes or impact. Additional data collection will often be required in order to estimate the impact and value for money of an intervention. To reduce bias and improve the reliability of the findings, it is best practice for such data collection and analysis to be conducted (or at least overseen) by evaluation experts not involved in the project itself.

- 7.4.2 Bespoke qualitative data will also be required to complement metrics and statistical analysis in answering critical questions about how the intervention was delivered, what factors might have influenced its outcomes, why results might have differed between settings and the extent to which the intervention worked as assumed in the theory of change. Interviews, focus groups, observations, in-depth case studies, (qualitative) surveys and documentary evidence are all examples of qualitative data sources that can enhance and explain quantitative findings, providing useful learning for other programmes or scale up.
- 7.4.3 Data collection methods and sources should be selected based on project needs and the evaluation questions requiring answers. The Magenta Book provides useful guidance on how to decide what data is required, how to select data sources and key analytical approaches. Helpful tips can also be found on the [Better Evaluation](#) website.
- 7.4.4 Decisions about the best data sources to be used in evaluation should be driven by the desired outputs, outcomes and impacts for each particular intervention, as well as its theory of change and/ or logic map. While many transport interventions will have similar desired outcomes (e.g., improved air quality, reduced noise levels, better connectivity), it is essential that the success of each intervention is estimated according to its own specific aims and objectives.

## 7.5 Data access

- 7.5.1 For existing administrative data, permissions and access should be organised well in advance of planned analysis. Most government datasets and other publicly-available data can only be accessed through the ONS [Secure Research Service](#). This is a secure online environment that provides access to pre-approved data for analysis purposes, with strict measures in place to prevent unauthorised access, downloading the data or identifying individual data subjects. Access to the secure service is permitted only for accredited researchers working in ONS safe rooms or in safe rooms approved for locations that have obtained assured organisational connectivity. The projects for which the data analysis is conducted also need to be approved in advance. Early planning is therefore essential to prevent significant delays.

## 7.6 Data protection and ethics

- 7.6.1 Routine monitoring data and other bespoke data required for evaluation must be collected, stored, processed and disposed of in accordance with all relevant data protection legislation and research ethics principles. This includes making sure that any stakeholders that are surveyed or interviewed for the purpose of evaluation are informed in advance of the lawful basis for data processing, what data will be collected, how it will be analysed, for what purposes, how long the data will be stored for, who it may be shared with and why, what measures will be taken to maintain participant confidentiality and, if relevant, anonymity.

- 7.6.2 The [Information Commissioner's Office](#) publishes detailed up-to-date guidance on relevant data protection legislation, with helpful checklists from ensuring [lawfulness, fairness and transparency](#) to what information should be [provided to participants in advance](#) and what to do if initial plans change.
- 7.6.3 Data protection safeguards are necessary to ensure monitoring and evaluation are ethical, but they are not in themselves sufficient. Project teams have a responsibility to ensure their data collection, analysis and dissemination do not pose any risk of harm to participants or research teams, do not unfairly exclude any relevant participants and have a clear public benefit, among other requirements. Careful consideration must also be given to commercial sensitivities and any other restrictions that may be specified in data sharing agreements.
- 7.6.4 The UK Statistics Authority provides a useful [ethics self-assessment tool](#) to help project teams review and mitigate ethical risks throughout the life of the project. Other helpful documents include the government guidance on [ethical assurance for social and behavioural research](#) and the UK Evaluation Society [guidelines for good practice in evaluation](#).

## 8. Reporting and dissemination

- 8.1 In order for evaluation to achieve its key objectives of improving accountability and learning, its results and methodology must be shared with relevant stakeholders in a timely and efficient manner.

### 8.2 Interim reports and emerging findings

- 8.2.1 One important benefit of planning and delivering evaluation alongside the intervention, particularly for new or developing programmes, is that ongoing data and emerging findings can be used to make adaptations, to give the intervention the best chance of success. Adequate sharing of data and emerging findings must be in place to ensure early learning can feed into refining the intervention and mitigating any unintended consequences, as well as benefiting other similar programmes or organisations.
- 8.2.2 When planning an intervention and its evaluation, a reporting and dissemination plan should also be agreed, including preliminary briefings, interim and final reports, to ensure that outputs can inform reviews at key delivery stages, major fiscal events and other relevant decision points.

- 8.2.3 To maintain transparency and evaluation quality, it is important that changes made during the intervention are discussed in advance with evaluation experts and documented clearly so they can be accounted for in subsequent analyses.

## 8.3 Feedback for appraisal

- 8.3.1 Just as it is critical for appraisal teams to consider the evaluation stage and take steps to enable it (for example, by consulting evaluation experts, preparing appraisal handover packs), so too evaluators should endeavour to facilitate better future appraisals.
- 8.3.2 Using robust evaluation to verify the assumptions made in appraisal can help refine such assumptions for future modelling and, in time, identify any areas of systematic mis-forecasting. (See the POPE methodology referred to above as a good example.) Wherever possible, project teams should ensure that such lessons learnt from evaluation are fed back to the appraisal teams to be incorporated in future work (see ROAMEF cycle, section 1.2.1).
- 8.3.3 In addition to recommendations for future evaluation (and, where relevant, implementation), final evaluation reports should also include recommendations for future appraisal of similar schemes, as well as, if feasible, data that can be used in future appraisals. Organisations should store the findings systematically so they can inform future appraisal guidance. For more information, please see [Strengthening the links between appraisal and evaluation](#).

## 8.4 Publication

- 8.4.1 Publication of evaluation findings is important for transparency, to ensure accountability for the use of public funds and so that lessons about the effectiveness of interventions can be shared with a wide range of stakeholders, including the public. This is in line with the Department for Transport's [Research publication policy](#), the Government Social Research Profession [Publication Protocol](#) and the [Code of Practice for Statistics](#) published by the Office for Statistics Regulation and the UK Statistics Authority.
- 8.4.2 When preparing the business case for a new or updated programme, careful consideration must be given to evidence from previous evaluations, to ensure that relevant lessons feed into improving the policy cycle and increasing the value of public spending.
- 8.4.3 Sufficient time and budget should be allowed in the overall project resourcing for thorough quality assurance and preparation for publication, including making sure the final outputs comply with the latest accessibility requirements. (For more information, please see the government's guidance on [publishing accessible documents](#) and the [Web Content Accessibility Guidelines](#).)

## 9. Further resources

### Other government guidance

[The Magenta Book - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[The Green Book: appraisal and evaluation in central government - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Public Value Framework and supplementary guidance - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Guide for effective benefits management in major projects - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Logic mapping: hints and tips guide - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Transport impact evaluations: achieving better attribution - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Local authority major schemes: monitoring and evaluation framework - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Transport schemes: links between appraisal and evaluation - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[TAG: transport appraisal process - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

### Selected examples of transport evaluations

[DfT evaluation strategy and programme, 2022](#)

[Local major schemes: meta evaluation 2007 to 2012 - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Evaluating low cost interventions to encourage the use of sustainable transport - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Inclusive Transport Strategy: evaluation baseline and technical reports - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Cycle City Ambition Programme Evaluation 2013-2018 - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Impact of the Local Sustainable Transport Fund: synthesis of evidence - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Evaluation of roads reform: wave 1 and wave 2 reports - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[HS1: first interim evaluation - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Thameslink Programme evaluation: baseline report - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[First of a Kind rail innovation programme: evaluation - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[Final report on the increased speed limit for heavy goods vehicles - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

[20 mph speed limits on roads - GOV.UK \(www.gov.uk\)](http://www.gov.uk)

# 10. Document provenance

This is a new TAG unit prepared in 2022.