March 2010

Guidance for transport impact evaluations

Choosing an evaluation approach to achieve better attribution

Developed by the Tavistock Institute in consultation with AECOM

Dr Dione Hills
Dr Kerstin Junge
The Tavistock Institute
30 Tabernacle Street
London EC2A 4UE

T +44 (0)20 7417 0407
F +44 (0)20 7457 0566
W tavinstitute.org

Although this report was commissioned by the Department for Transport (DfT), the findings and recommendations are those of the authors and do not necessarily represent the views of the DfT. While the DfT has made every effort to ensure the information in this document is accurate, DfT does not guarantee the accuracy, completeness or usefulness of that information; and it cannot accept liability for any loss or damages of any kind resulting from reliance on the information or guidance this document contains.

Foreword

The Department for Transport (DfT) is committed to maximising the role of evidence and continuous learning within policy making and delivery. HM Treasury's Green Book acknowledges the major roles rigorous and robust appraisal and evaluation evidence have in ensuring public services are delivered efficiently and effectively and are targeted to provide the greatest benefits. The impact evaluations of transport interventions seek to test whether the anticipated benefits have been generated when delivered in the real world, to identify any unintended effects, and allow the value for money assessment to be reviewed in light of the evidence post implementation. Therefore, impact evaluations have a key role in building the evidence base for future policy making and appraisal, by enhancing accountability and supporting organisational learning to inform future investment decisions.

One of the main challenges in undertaking impact evaluations of transport interventions is the ability to demonstrate that the observed outcomes and impacts have been caused by the intervention, confidently ruling out the influence of external factors. This guidance has primarily been designed to focus on the approaches which can attribute the outcomes and impacts to the intervention and provide some considerations to take when undertaking these evaluation approaches within a transport context.

However, it is recognised that there are many varied, and sometimes conflicting, aspects to consider when designing an impact evaluation and whilst the ability to attribute the findings back to the intervention is considered to be very important, it might not be the central requirement for some evaluations. Therefore, the guidance has been designed to provide a systematic process for considering the most suitable evaluation approach which is tailored to the type of intervention being delivered and the evidence requirements. However, due to the specific circumstances underlying each evaluation, this guidance is not intended to be prescriptive but provide a flexible framework for selecting the overarching evaluation approach.

As the primary objective of the guidance is to consider impact evaluation approaches which enable the attribution of the outcomes and impacts to the intervention, this document does not constitute comprehensive guidance on all aspects of transport evaluation (for example undertaking a process evaluation), neither does it provide detailed guidance on how to evaluate specific types of interventions. However, the earlier steps in the evaluation design will be relevant for all types of evaluation.

This guidance document has been developed to support policy makers, delivery bodies and analysts from the point at which they have decided to undertake a transport impact evaluation. Decisions about which types of interventions to evaluate or not needs to be considered at a strategic level taking account of the broader evidence requirements and the value it will provide policy makers and delivery bodies.

Social Research and Evaluation
Department for Transport

Acknowledgements

The Tavistock Institute and Department for Transport would like to thank a number of people who provided support and advice relating to this work.

In addition to Department for Transport colleagues who advised on the evaluation challenges in their particular sector, we would like to thank the following for their advice on wider evaluation debates and concerning specific examples of evaluations taking place:

Dr Avril Blamey, Avril Blamey & Associates, Planning and Evaluation Consultants
Charles Buckingham, Transport for London
Professor Ken Judge, School for Health, University of Bath
Professor Mike Kelly, National Institute for Health and Clinical Excellence (NICE)
Professor Alan Pearman, Leeds University Business School
Professor Mark Petticrew, School of Hygiene and Tropical Medicine, University of London
Professor Elliot Stern, Lancaster University
Louise Wootton, Highways Agency

Table of contents

Executive Summary 7
Introduction 9
1. Step 1: Clarifying the background to the evaluation 13
2. Step 2: Considering the nature of the intervention 15
3. Step 3: Mapping the intervention logic 23
4. Step 4: Defining the evaluation purpose and framing the questions 28
5. Step 5: Deciding on the best evaluation approach 33
5.1. When to choose an outcome approach 33
5.2. When to choose an experimental design approach 38
5.3. When to choose a theory-based approach 46
6. Step 6: Refining the choice 52
6.1. Experimental Evaluation Approach - Random assignment or quasi-experimental design? 52
6.1.1 Experimental designs 52
6.1.2 Quasi-experimental designs 53
6.2. Deciding which theory-based evaluation approach is appropriate 58
6.2.1 Is a Theory of Change approach most suitable? 58
6.2.2 Is a Realist Evaluation approach most appropriate? 61
6.3. Choosing a combined approach 66
6.3.1 Combining theory-based approaches with experimental designs 66
6.3.2 The extended intervention logic option 67
Summing up: six issues to consider for better attribution 69
Glossary 71
Key texts for further reading 74

Figures and tables

Figure 1: The ROAMEF cycle 10
Figure 2: Six steps to select an evaluation approach for better attribution 12
Figure 3: Components of an intervention logic 24
Figure 4: Intervention logic for the Walk in to Work Out programme 27
Figure 5: Flowchart for the selection of an outcome approach 37
Figure 6: Flowchart for choosing an experimental design 44
Figure 7: Flowchart for selecting a theory-based approach 51
Figure 8: Flowchart for deciding on experimental or quasi-experimental designs 57
Figure 9: Summary of selection criteria for evaluation approaches – based on the evaluation questions to be answered 70

Table 1: Questions to consider in creating the intervention logic 25
Table 2: Evaluation purpose, approach and type of evaluation questions to ask 31

EXECUTIVE SUMMARY

This guidance outlines a systematic approach to designing impact evaluations for transport interventions. It has been developed to highlight some of the available impact evaluation approaches with the particular focus on those approaches which enable the outcomes and impacts to be attributed to the intervention. Rather than prescribing an evaluation approach across all transport interventions, this guidance provides policy makers, scheme promoters, delivery bodies and analysts involved in the design, implementation and evaluation of transport interventions, with a step-by-step process to assessing their evidence needs and it will support them in selecting the most appropriate evaluation approach.

Evaluation can offer robust quantitative and qualitative data on the outcomes and impacts of a policy, programme, package or scheme: providing evidence not only for what changes it produced but also why and how these changes came about. Sound evidence has, in turn, a crucial role to play to improve the evidence base for future policy making.

However, a rigorous design is a precondition for evaluation to deliver sound evidence. Transport impact evaluations therefore need to be based on a suitable theoretical framework which is appropriate to the evaluation questions and the type of intervention to be investigated. The guidance is designed to help all those planning or preparing a transport impact evaluation to consider a number of key factors that will influence the design choices they need to make in order to create a sound evaluation. The process of selecting and refining these choices are often iterative.

Six steps need to be completed to select a robust transport impact evaluation approach:

First, it is important to be clear about the background to the evaluation: the objectives of the intervention, the users of the evaluation results and the financial and staff resources that are available for the work. All of these factors will influence the evaluation design as they will determine the scope of the evaluation work and play a key role in developing the evaluation questions in step four.

Second, a closer look needs to be taken at the kind of intervention that is to be evaluated: is it a policy, programme, package or scheme? Depending on intervention type, the evaluation purpose, questions, focus and timing will differ slightly. It is therefore useful to locate the intervention to be evaluated within one of these categories before progressing with the detailed evaluation design.

The design work for the evaluation begins in **step three** with mapping the intervention logic. This entails a process of systematically linking key components of an intervention so as to create links between activities, what they produced (outputs), short to medium-term results (outcomes) and long-term results (impacts). The intervention logic forms the basis for deciding on the focus of the evaluation and forming the main evaluation questions so that key knowledge requirements are met.

The **fourth step** is about finalising decisions about the evaluation purpose and developing the evaluation questions by drawing on the thinking in the previous steps. At the end of this step, the reader will have decided whether the main purpose of their impact evaluation is accountability or generating knowledge (or both) and what evaluation questions they are looking to answer.

Fifth, the most suitable overall approach to the impact evaluation needs to be chosen. Three broad approaches are, in principle, available: the outcome approach which compares the situation before an intervention with the situation after its introduction; the experimental approach which compares the outcome of an intervention with what would have happened in its absence by comparing two population groups (one taking part in the intervention, the other not); and the theory-based approach which articulates and tests the assumed connection between an intervention and anticipated impacts. Deciding which approach is most suitable means weighing up the strengths and weaknesses of each in light of the evaluation purpose, their ability to attribute the findings to the intervention, the key questions to be answered and the nature of the intervention.

In a **sixth** and final step, this choice is then refined. Depending on the outcome of step five and if a decision has been made that there is a need for evidence which is attributable to the scheme, then the next stage is to select an appropriate approach to achieve this. These include considering random or quasi experimental designs; theory-based approaches such as Theory of Change or Realist Evaluation. It also provides options for combining approaches depending on the specific evaluation need. For example using experimental evaluation methods within a theory-based approach to deliver evaluation findings which robustly tests the degree of change caused by the intervention and also provides an explanation for why the observed change occurred. Alternatively building in an extended intervention logic model to an outcome approach will seek to understand whether the anticipated change occurred and why. Guidance is provided on some of the main considerations to make when starting evaluations using these approaches.

INTRODUCTION

This guidance has been developed to support policy makers, scheme promoters, delivery bodies and analysts involved in the design, implementation and evaluation of transport interventions in selecting the most suitable approach to evaluating impacts.

Undertaking a robust evaluation to measure the extent to which an intervention has achieved its anticipated objectives is extremely valuable. It delivers sound evidence on success and value for money and also helps develop knowledge about the effectiveness and efficiency of an intervention which will build the evidence base for making future investment decisions.

Approaches to evaluation design should be determined by the purpose of the evaluation as well as the nature of, and circumstances around, the specific intervention. However, it is important that the evidence produced by an impact evaluation demonstrates, as far as possible, that the observed findings have been caused by the intervention. This is known as attribution. If evaluations are not designed to show that the intervention has caused or influenced change, then it is possible that the effects observed could have been caused by factors other than the intervention. This would reduce the confidence in the evaluation results.

This guidance highlights the main evaluation approaches which can be used to design evaluations for better attribution. In six consecutive steps it helps the reader choose the best approach for their planned evaluation.

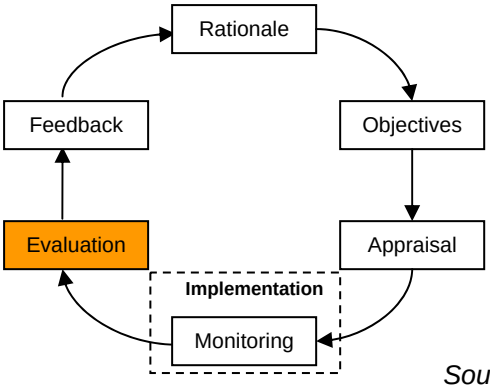
Why evaluate?

Impact evaluation is a tool that can offer transport analysts, policy makers, scheme promoters and other individuals responsible for assessing transport interventions[1] rich quantitative and qualitative data on its impact and the reasons for why and how this impact was achieved[2].

As the ROAMEF[3] cycle in Figure 1 illustrates, evaluation explores the impacts of the intervention once it has been implemented and builds on appraisal which assesses the proposed intervention prior to delivery. In reality, evaluations are most effective when planned at an early stage in the intervention design cycle as this ensures the appropriate data will be monitored during the intervention delivery period. In order to make best use of evaluation evidence it is important to maximise the learning which can be fed back into the development of future interventions or appraisal models.

[1] In the following sections the word 'intervention' is used as a collective noun to describe transport policies, programmes, packages, projects and schemes.
[2] Although the types of evidence produced varies with the impact evaluation approach undertaken. This is explained in detail in steps five and six.
[3] Rationale, Objectives, Appraisal, Monitoring, Evaluation, Feedback (ROAMEF)

Figure 1: The ROAMEF cycle



Source: The Green Book (modified)

Evaluation tools and techniques can be used to assess the relative success of any transport intervention: be this a well-researched intervention, one that is subject to appraisal or one about which little is yet known. Evaluation is therefore a flexible tool which can offer valuable learning, generating new knowledge or refining that which already exists about an intervention.

In order for an impact evaluation to deliver its full benefits it needs to be carefully designed. This means, first and foremost, the evaluation needs to be underpinned by an appropriate overarching framework that is suitable for the type of intervention to be investigated and the kind of questions that the evaluation seeks to address. Secondly, choosing an approach which will generate robust data on the linkages between an intervention and an observed outcome will give policy makers confidence in the evidence produced and whether this demonstrates that the intervention has been successful or not. This will inform decisions about whether to make further investments in the intervention or similar interventions in the future.

Purpose of the guidance

The following sections therefore offer a practical step-by-step guide for choosing the most appropriate approach on which to base an impact evaluation of transport interventions. They were developed following a review of evaluation literature and discussions with evaluation experts and users of evaluation, both from within the field of transport, and from other policy areas (public health, regeneration and international development which faces similar evaluation challenges).

The following guidance offers the following kinds of support:

- Help with identifying appropriate and relevant evaluation questions;
- Enabling the reader to review the respective strengths and weaknesses of different evaluation approaches and identifying the most appropriate evaluation approach for their intervention; and,
- Help with identifying a sound approach to evaluation which enables the impacts to be attributed to the intervention, particularly in the case of complex transport interventions.

The six steps therefore invite the reader to consider the context, rationale and focus of their impact evaluation and offer a systematic way of thinking through the best approach to take for their impact evaluation.

Step 1: *Clarifying the background to the evaluation* is a short preliminary stage which is designed to remind the reader of the delivery context in which the evaluation is being undertaken, in particular the intended use and users, the intervention objectives to be considered and the resources (both financial and non-financial) needed to execute the work successfully.

Step 2: *Considering the nature of the intervention* encourages the reader to think about the type of intervention they are planning to evaluate, whether it is a policy, programme, package or scheme, and how this will inform the evaluation approach.

Step 3: *Mapping the intervention logic* (i.e. the conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes) will enable the reader to reflect on the factors that are likely to shape the delivery of the intervention and therefore influence its success. Explicitly articulating these connections will help to highlight what evidence is required from the evaluation and in particular, indicate where key gaps in the existing evidence base might be.

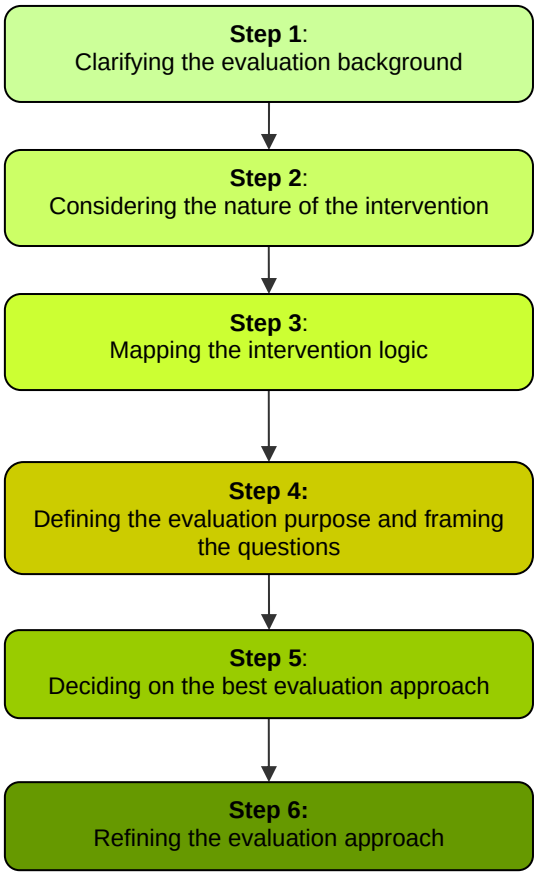
Step 4: *Defining the evaluation purpose and framing the evaluation questions* will build on the previous steps to ensure that the evidence produced by the evaluation meets the requirements of the stakeholders.

Step 5: *Choosing the most suitable overall approach to the impact evaluation*. Three broad approaches are, in principle, available: the outcome approach which compares the situation before an intervention with the situation after its introduction; the experimental approach which compares the outcome of an intervention with what would have happened in its absence by comparing two population groups (one taking part in the intervention, the other not); and the theory-based approach which articulates and tests the assumed connection between an intervention and anticipated impacts. Deciding which approach is most suitable means weighing up the strengths and weaknesses of each in light of the evaluation purpose, their ability to attribute the findings to the intervention, the key questions to be answered and the nature of the intervention.

Step 6: *Refining the choice*. Depending on the outcome of step five and if a decision has been made that there is a need for evidence which is attributable to the scheme, then the next stage is to select an appropriate approach to achieve this. These include considering random or quasi experimental designs; theory-based approaches such as Theory of Change or Realist Evaluation; whether the approach should combine elements from both; or, if an extended intervention logic model combined with an outcome approach best meets the needs of the evaluation. Guidance is provided on some of the main considerations to make when starting evaluations using these approaches.

Figure 2 illustrates how these steps fit together.

Figure 2: Six steps to select an evaluation approach for better attribution



After working through these six steps, readers will have chosen the most appropriate evaluation approach for their impact evaluation and will have started to think about the design for this evaluation (including making initial judgments on the appropriate methods for collecting the required data). Readers are also likely to have gained a deeper understanding of the benefits evaluation can offer them in their assessment of transport interventions, and how they might design an evaluation that delivers data that is useful to them.

1. STEP 1: CLARIFYING THE BACKGROUND TO THE EVALUATION

Step 1 is a short preliminary stage designed to ensure that the background circumstances for the planned evaluation have been fully considered. Before starting an impact evaluation it is important to be clear about what the objectives of the intervention are, the resources available for the evaluation, who the audience for the results is and how these will be used.

What are the objectives of the intervention?

It is important to clarify the aims of the intervention and the type of change it is anticipated to deliver. It can be helpful to differentiate between objectives which the intervention is expected to deliver over a short time period (1-3 years) and those which might take longer to be achieved. It can also be helpful to consider what the primary and secondary goals of the intervention are.

Who will use the results of the evaluation, and how?

To ensure the evaluation evidence delivers optimum benefits it is important that the requirements of the anticipated users of the findings are considered. The results might be used to implement recommendations, inform decision-making, clarify thinking, provide accountability to stakeholders and contribute to improved knowledge amongst those best able to take advantage of it[4]. Thus, when starting to design the evaluation, it is important to understand:

- Who the end-users of the evidence will be (e.g. analysts, DfT policy makers, other Government Departments, Local Authority scheme promoters, key stakeholders in transport decision-making such as Councillors, emergency services, local community groups, local transport operators etc.);
- What the different expectations for the results are; and
- What would allow them to make effective use of the evaluation findings. This includes, in particular, meeting different data requirements (but also mechanisms for dissemination and follow-up structures). For instance, transport analysts are likely to favour quantitative data measuring impact. DfT policy makers might be interested in issues such as understanding behaviour change, value for money, unintended outcomes and longer term impacts. Local Authority scheme promoters will possibly focus more on generating evidence which demonstrates the delivery of intended outcomes.

These considerations will influence the design of the evaluation. For instance, by understanding the range of requirements for the evaluation, the questions can be designed to reflect these and methods can be chosen that generate relevant evidence. Appreciating the requirements of stakeholders and involving them

[4] For a fuller discussion of this see also the European Commission's evaluation guidelines: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/designing_implementing/use_en.htm

throughout the evaluation cycle will build commitment to the evaluation and will help to ensure that results are viewed as valid and useful.

What resources are available for the evaluation?

Being clear about the resources available for the evaluation (both financial and personnel) will help with the selection of the most appropriate evaluation approach and the detailed methodological design work.

Financial resources: Budgets for evaluations should generally be proportional to the resources allocated to the intervention. Therefore, careful consideration needs to be given to how this may affect the evaluation design and outputs. Key stakeholders may need to be consulted on the implications of this, especially for smaller scale schemes.

Personnel resources: A consideration of the skills required to undertake an evaluation is important. There might be a requirement to build evaluation expertise by drawing upon the technical skills of evaluation consultants or developing the capabilities of in-house analysts. Due to the multi-disciplinary nature of evaluations it is important to consider the range of analytical specialists who might need to contribute to the evaluation. These decisions will be informed by the scale of the intervention, the capacity required to deliver the evaluation, the key evidence requirements and the need for independent evaluators.

After a consideration of the background to the impact evaluation, it is now important to be clear about the nature of the intervention. This is discussed in Step 2 below, which looks at the different types of intervention (policy, programme, package or scheme) and how this may influence the focus and purpose of the evaluation and the type of evaluation questions addressed.

2. STEP 2: CONSIDERING THE NATURE OF THE INTERVENTION

Step 2 focuses on the nature of the intervention and, building on the considerations made in Step 1, will highlight how this will influence the evaluation requirements. The focus of the evaluation questions will be informed by the type of intervention being evaluated. Policies, programmes, packages or schemes have very different objectives, reach and timeframes so the aim and focus of an evaluation will differ accordingly. This step will highlight to the reader some of the general characteristics of their type of intervention and the implications for evaluation design by discussing its key features. Whilst there are some similarities between the different types of interventions, this section discusses them separately to highlight the (sometimes subtle) differences between them. The reader is therefore advised to go directly to the type of intervention they are looking to evaluate.

Evaluation of policy

Policy making is the process by which governments translate their political vision into programmes and actions to deliver 'outcomes' – desired changes in the real world[5]. For example:

- The sustainable long-term strategy for the development of air travel in the UK 'The Future of Air Transport White Paper' (2003) sets out a strategic framework for the development of airport capacity in the United Kingdom over the next 30 years, against the wider context of the air transport sector; and
- The Government's strategy for improving road safety during the period 2000 – 2010 'Tomorrow's roads: safer for everyone' (2001).

The purpose of policy evaluation tends to be focused on building knowledge about the extent to which the policy has been successful in achieving its objectives, who has specifically benefited from the policy and why the impacts occurred.

Policy evaluation questions can include the following examples[6]:

- Have the anticipated outcomes and impacts been achieved?
  - To what extent are the observed changes additional to what would have happened in the absence of the policy?
  - Were there any unanticipated impacts / displacement effects?
  - How are the impacts distributed?
- Has the policy been successful, why / why not?
  - Which elements of the policy were particularly influential in achieving the overall goals?
  - Which target groups was it most effective for?
  - What lessons can be learnt for future policy development?
- What is the contribution of the policy to DfT's strategic goals?

[5] 'Modernising Government" 1999 www.archive.cabinetoffice.gov.uk/moderngov/download/modgov.pdf

[6] This is not an exhaustive list. Step four provides more detail about framing evaluation questions.

- What was the contribution to cross-government commitment?
- To what extent did the anticipated costs and benefits match the actual outcome?

Key characteristics of policies and implications for evaluation design

1. The policy scope
Policies tend to aim to deliver change across a broad range of issues, and the evaluation needs to be designed in a way which identifies the extent to which the various strands of the policy have contributed to the observed change(s) and how these inter-relate. Additionally, the design stage needs to take into account the geographical scale of the intervention. For example, some DfT policies are targeted nationally and this means that it might not be possible to compare situations with and without the intervention as it will be difficult to find an accurate comparison area. If this is felt to be important to policy makers, then it might be worth considering whether it is feasible to pilot the policy and fully evaluate the pilot before implementing it across the whole target population.

2. The policy context
To deliver cross-cutting objectives policy-making can involve a joined-up approach (for example, between Government Departments). This requirement will need to be reflected in the design of the evaluation to ensure that all relevant stakeholders are involved and all the important impacts are measured. Equally, the evaluation approach will need to consider the wider context in which the intervention is being delivered, as they are often integrated with other policies, and the wider strategic agenda. This has implications for ensuring the relevant stakeholders are involved in the evaluation and also that the methods chosen will generate evidence that is useful to their needs. It also means that the influence of the wider context on the impact of the intervention or the observed change directly needs to be understood.

3. The mode for implementing the policy
Policy delivery tends to be via a range of mechanisms. For example, changes to infrastructure, legislation, funding, directives, fiscal incentives or the delivery of awareness-raising campaigns etc. In these cases, the evaluation design needs to be able to identify the relative contribution of these different delivery mechanisms on the observed impact. It is also important to consider who is responsible for delivering policy as this might influence the evaluation findings. For instance, in policies which are decentralised the approach to delivery is likely to depend on locally defined need. In order to draw conclusions about the impact of the policy, the evaluation design might therefore need to reflect localised variations in the way the policy was delivered.

4. The implementation timeframe
Policies are strategically designed to have a longer term and sustainable influence on behaviours. They tend to take a number of years to fully implement. Ideally the impacts of the policy should be observed over a number of years to allow for the full benefits to be realised once the policy is implemented. However, there is always a danger, over longer time periods that other external factors, such as new government policies, will also influence any observed changes in impacts. The evaluation approach will therefore need to be able to identify whether the policy caused the effect, or whether the effect was due to a combination of influences. Where the evaluation is undertaken over a shorter timeframe, the shorter term outcomes can be used as an indication that the policy is on course to deliver longer term impacts (logic mapping - see step 3 - is particularly helpful here), and used to inform forecasts of longer term benefits as in appraisal.

5. The nature of anticipated outcomes
Policy objectives often provide the strategic framework for delivering initiatives which reflect localised needs and priorities. This requires the policy objectives to be fairly broad rather than prescriptive. This makes it important to consider at an early stage of the evaluation design how the objectives will be measured.

Evaluation of programmes

Programmes are a series of similar schemes introduced in a in a coherent and co-ordinated way in a defined area (or areas) over a defined period of time (likely to be several years) in order to achieve a particular overarching policy objective. A programme could be defined as a temporary, flexible organisation created to co-ordinate, direct and oversee the implementation of a set of related projects and activities in order to deliver outcomes and benefits related to the organisation's strategic objectives. Programmes implement public policy. For instance, cycling programmes such as Cycling Demonstration Towns (2005) and Cycling Cities and Towns (2008) are pilot programmes through which the increased cycling objective is implemented. Numerous localities participate, each implementing their own packages (see below) in order to achieve the overall programme's objectives.

The purpose of evaluating programmes is to justify the investment of delivering a co-ordinated programme to achieve larger-scale impacts than an investment in smaller, single schemes would achieve (i.e. whether it achieved its objective as well as contribution to government policy objectives, and why). It is also valuable to learn about how the components of the programme have led to the observed change and what their relative contributions are. As part of a programme evaluation it is also possible to evaluate the success of individual packages or schemes, if this is required to meet the objectives, purpose and questions of the evaluation. Where the programme is being delivered across a number of sites, it also creates the opportunity for an evaluation to be undertaken in a representative sample of sites, rather than seeking to evaluate all parts of the programme in depth.

Programme evaluation questions can include the following examples[7]:

- Have the anticipated outcomes and impacts been achieved?
  - To what extent are the observed changes additional to what would have happened in the absence of the programme?
  - Were there any unanticipated impacts/ displacement effects?
  - How are the impacts distributed?
  - How did the different schemes within the programme interact to produce the observed impacts?
- Has the programme been successful, why / why not?
  - Which schemes (or packages of schemes) have been particularly successful, and why?
  - To what extent did the co-ordinated approach to implementing the intervention influence the observed impacts?

[7] This is not an exhaustive list. Step four provides more detail about framing evaluation questions.

    - What is the relative contribution of individual interventions to the observed impacts?
- What lessons can be learnt for future programme development?
- Which are the most useful measures to include in future programme design?
- What is the contribution of the programme to DfT's strategic goals?
- To what extent did the anticipated costs and benefits match the actual outcome?

Key characteristics of programmes and implications for evaluation design

1. The programme scope
Programmes are designed to deliver a combination of initiatives in order to address an overarching objective. Therefore, as far as possible, the evaluation needs to be designed in a way which identifies the extent to which the various components of the programme have contributed to the observed change(s), and how individual components inter-relate.

2. The context of the programme
The delivery of programmes will tend to require a joined-up approach between a number of partners. For example, cycling programmes can require the involvement of a variety of organisations including Government Departments, Cycling England, Local Authorities, transport operators, employers and schools. The evaluation design will need to take into account the range of stakeholders, including those involved in the delivery of a programme, to ensure that their interests are reflected.

3. The mode for implementing the programme
Programmes tend to be delivered in a way that is best suited to localised need. Their content might therefore vary by area of delivery. In order to draw conclusions about the impact of a programme, the evaluation design would need to be able to take into account such localised variations in programme delivery and specific aims and objectives.

4. The implementation timeframe
Programmes are often designed to be implemented over a number of years. This means that there is a risk that external factors other than the programme could influence any observed changes since the baseline position was measured. The evaluation approach will therefore need to be able to identify whether the programme caused the effect or whether the effect was due to the combination of other influences. If the evaluation is undertaken over a shorter time frame, but the impacts of the programme are anticipated to continue for several years[8], then the results of the evaluation can also be used to inform forecasts of the anticipated longer term impacts, using similar principles to those used in appraisals.

5. The nature of anticipated outcomes
Programmes often co-ordinate the delivery of measures which reflect localised needs and priorities. This requires their objectives to be fairly broad rather than prescriptive. This makes it important to consider at an early stage of the evaluation design how the objectives will be measured. However, programmes are designed to achieve broader policy objectives (e.g. DfT strategic goals) so the extent to which changes in these can be attributed to a programme will need to also be considered in the design of the evaluation.

[8] See Step 3 logic mapping for more information

Evaluation of packages

Packages are a combination of measures introduced to address common or shared objectives. These measures are not normally part of a co-ordinated set of activities at a national level (unlike a programme), but may still be used to work towards achieving one (national) goal. For instance sustainable travel packages of measures may include any of the following: car clubs, car sharing, teleworking, travel plans (schools, residential, work based), cycle schemes, pedestrian facilities, and travel information.

The purpose of evaluating packages tends to be generating knowledge about the impact of the package as a whole (i.e. whether it achieved its objective as well as contribution to government policy objectives, and why). However, especially for those running packages locally, it may also be helpful to understand what contribution individual interventions have made to the whole package and whether implementing a package of measures provides additional benefits. This information will, for instance, be very helpful for informing the design of future packages.

Package evaluation questions can include the following examples[9]:

- Have the anticipated outcomes and impacts been achieved?
  - To what extent are the observed changes additional to what would have happened in the absence of the package?
  - Were there any unanticipated impacts/ displacement effects?
  - How are the impacts distributed?
  - How did the different elements within the package interact to produce the observed impacts?
- Has the package been successful, why / why not?
  - Which schemes have been particularly successful, and why?
  - To what extent did the co-ordinated approach to implementing the intervention influence the observed impacts?
  - What is the relative contribution of individual schemes to the observed impacts?
- What lessons can be learnt for the future development of packages?
- Which are the most useful measures to include in future package design?
- What is the contribution of the package to DfT's strategic goals?
- To what extent did the anticipated costs and benefits match the actual outcome?

[9] This is not an exhaustive list. Step four provides more detail about framing evaluation questions.

Key characteristics of packages and implications for evaluation design

1. The package scope
Packages are likely to involve not only one overarching objective, but also a number of additional subsidiary objectives, linked to the individual activities incorporated into the package. The evaluation design will therefore need to be able to show how each individual intervention and its objectives complements, or provides additional support, to the overarching aim of the package.

2. The context of the package
The composition (in terms of number and types of actions taken) and implementation (in terms of stakeholders involved) of a package is likely to be highly context dependent. This means it will vary according to particular local situation and needs. An impact evaluation would therefore need to be designed in a way that allows it to say something about what contextual factors are important for packages to achieve their anticipated outcomes.

3. The mode for implementing the package
Packages may be implemented at different levels. Some schemes of the package may be delivered at Local Authority level, others at a sub-local level (for instance at school level or within a specific neighbourhood). The number of organisations involved in either taking decisions or delivering actions on the ground may also be broad and varied. Involving these stakeholders in the evaluation will be important both to hear their views and to gain access to any (monitoring) data collected locally.

4. The implementation timeframe
Packages might be less coordinated than programmes and consequently different elements of the package may be implemented over different timescales. The evaluation design will therefore need to be able to account for this and what it means for the impact of the package. If the evaluation is undertaken over a shorter time frame, but the impacts of the programme are anticipated to continue for several years[10], then the results of the evaluation can also be used to inform forecasts of the anticipated longer term impacts. These longer term impacts might be assessed using the same principles as are used in appraisal, although it will be particularly important to note the sustainability of the intervention in the longer term, particularly if implemented using short term funding.

5. The nature of anticipated outcomes
Evaluating packages means looking at the cumulative impact of the individual schemes to the objective of the package. The combination of different activities under one package will often aim to achieve broader policy objectives (e.g. DfT strategic goals) so the extent to which changes can be attributed to the package will need to be investigated by the evaluation.

[10] See Step 3 logic mapping for more information

Evaluation of schemes

A scheme is a single endeavour undertaken to create a unique result or change in response to a specific problem. Schemes therefore include a range of undertakings including: infrastructure measures; alternative transport schemes (such as community bus services, car sharing initiatives, sustainable travel modes); and others. When a number of schemes are combined in a single area they can form a package of measures (see above).

The purpose of evaluating schemes is to justify the investment into a particular endeavour (e.g. a major scheme) and therefore to assess whether the anticipated benefits have been achieved. It is also valuable to test the underlying appraisal assumptions and to learn more about how impacts were achieved.

Scheme evaluation questions can include the following examples[11]:

- Have the anticipated outcomes and impacts been achieved?
    - To what extent are the observed changes additional to what would have happened in the absence of the scheme?
    - Were there any unanticipated impacts / displacement effects?
    - How are the impacts distributed?
- Has the scheme been successful, why / why not?
    - Which target groups was it most effective for?
    - What lessons can be learnt for future scheme development?
- What is the contribution of the package to DfT's strategic goals?
- To what extent did the anticipated costs and benefits match the actual outcome?

Key characteristics of schemes and implications for evaluation design

1. The scheme scope
Schemes are individual measures designed to deliver unique results. This means that the scheme's end result is different to the results of other projects or interventions which might be taking place alongside it (as, for example, where the scheme is part of a wide package or measures, or one element in a larger programme of action. The evaluation will therefore need to be designed so that the unique contribution of the scheme to an issue can be established.

2. The context of the scheme
Schemes can vary in size from a site specific measure to an area-wide treatment or campaign. The impact evaluation will therefore need to be designed so that approaches and methods are commensurate with the scheme's scale.

[11] This is not an exhaustive list. Step four provides more detail about framing evaluation questions.

3. The mode for implementing the scheme
The overall conceptual, financial and operational responsibility for a transport scheme will most often be with one organisation. However, often this organisation will rely on third parties to implement the scheme, e.g. bus operators or charities running community bus schemes, engineering consortia building roads and managing the process thereof. Often, therefore, the effect of implementation mechanism on the overall impact of the scheme is likely to be of interest to the evaluation. Involving these stakeholders in the evaluation will be important both to hear their views and to gain access to any (monitoring) data collected locally.

4. The implementation timeframe
Schemes are often implemented within a limited time period, though their effect can be permanent (e.g. some infrastructure schemes). The longer a completed scheme has been in operation, the more important it will be for the evaluation approach to be able to identify what factors other than the scheme could be responsible for the impact observed. It will also be particularly important to make some kind of assessment, or forecast, of likely longer term benefits, as well as assessing their immediate impact.

5. The nature of anticipated outcomes
Schemes will often have specific objectives and relatively clearly defined predicted impacts (which may have been quantified through the appraisal process) therefore the evaluation should measure the extent to which these have been achieved. Additionally, capturing unintended outcomes and impacts, as well as delivering evidence for how observed impacts were achieved, can also provide valuable lessons for the future.

3. STEP 3: MAPPING THE INTERVENTION LOGIC

Step 3 is about clarifying the main components that are required to enable the intervention to deliver its intended impacts and to articulate how these are connected. The aim of this step is to help the reader build a map of the intervention logic to diagrammatically illustrate the relationship between the intervention and the intended impacts. This is an essential preparatory activity for refining your evaluation questions (Step 4) and deciding on the best approach to attribution (Steps 5 and 6).

Step 3 is designed to help the reader map the intervention logic of the scheme, package, programme or policy which is to be evaluated. This will help articulate what needs to happen in each stage of the intervention in order for the anticipated impacts to be achieved (this is referred to as causal pathways from here onwards). Based on this it will be possible to identify where the knowledge and data gaps are and hence where evaluation effort should be focused.

Intervention logic can be defined as "the conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes."[12] The terms "programme logic", "intervention logic" and "programme theory" are sometimes used to mean more or less the same thing.

Intervention logic is a method of systematically linking key components of an intervention so as to produce a causal pathway across the:

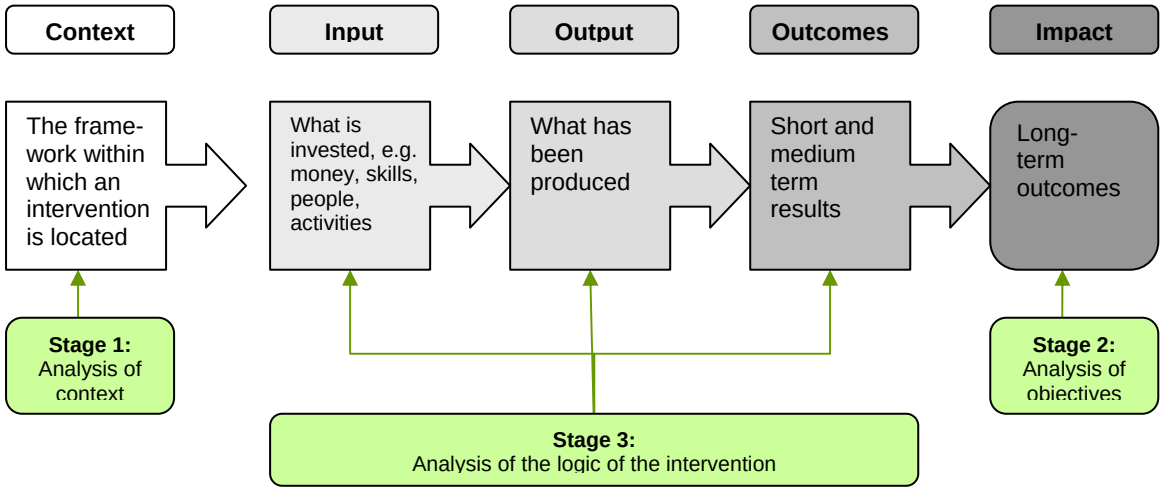
- Inputs (i.e. what is being invested in terms of resources and activities);
- Outputs (e.g. target groups reached, roads built, products developed);
- Outcomes (i.e. short and medium-term results, such as changes in traffic flow levels and modal shifts); and
- Impacts (i.e. long-term results such as better quality of life, improved health, environmental benefits etc).

Causality is central to an intervention logic as they order events in such a way that the presence of one event or action leads to, or causes, a subsequent event or action[13]. Figure 3 below illustrates the main components of an intervention logic.

[12] UK Evaluation Society, www.evaluation.org.uk/Pub_library/Glossary.htm

[13] For a more detailed description and use of intervention logic in evaluations you may wish to visit the European Commission's online resource for socio-economic evaluation (Evalsed): www.ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/planning_structuring/logic_models/description_en.htm

Figure 3: Components of an intervention logic



The process of drawing up the intervention logic ensures that the decision about ***what to evaluate*** and even ***how to evaluate*** (in terms of the approach to be selected) is based on a sound analysis and explicit articulation of the anticipated scope and scale of the intervention in terms of input, output, outcomes and impacts.

Transport analysts will often have already considered the causal pathways during the appraisal process. Therefore, these models will be helpful sources of information to draw on when compiling an intervention logic, as well as reviewing:

- Any supplementary / updated information that may have emerged since the appraisal model was developed;
- Any links between the individual components based on up-to-date information on the intervention;
- Any relevant information that may not have been included in the appraisal (on context, input, output, outcome, impact).

Drawing on the appraisal evidence and any other articulation of the intervention logic you should consider the following:

Table 1: Questions to consider in creating the intervention logic

	Issues to consider for the logic mapping
Context of the intervention	What **national** transport policies does the intervention originate from and support?
	What **regional** issues and priorities does the intervention originate from and support (transport, regional development, economic development, social inclusion objectives, health…)?
	What **local** issues and priorities does the intervention originate from and support (transport, local economic development, social inclusion objectives health, environment…)?
	What **sub-local issues** and priorities does the intervention originate from and support (transport, social inclusion objectives health, environment…)? What **other contextual factors** may influence the ability of the intervention to achieve its outcomes and impacts
Input	What financial resources are being invested in implementing the intervention?
	What other resources are being invested? E.g. people and (partner) organisations, skills, equipment, technology (e.g. electronic road signs), research or appraisal, etc.
Output	What is the intervention looking to 'produce'? This can include: ▪ What **activities** will directly result from the intervention? E.g.: building new road or rail infrastructure; street furniture; delivering training; information or awareness campaigns; passing regulation; provision of public transport priority facilities; walking and cycling facilities; parking controls; or, travel plans introduced. ▪ What **participation** will directly result from the intervention (who will be reached)? E.g.: types of transport users, partners, agencies, decision-makers, groups in society, areas of a specific town/city.
Outcomes	What is the intervention looking to achieve in the **short to medium term**? For instance: less congestion, raised awareness, partnership working, better skills, and change of attitude and / or behaviour.
Impact	What is the intervention looking to achieve in the **long term**? For instance: support the UK economy; contribution to climate change objectives; improved safety, security and health of the population; improved quality of life, or greater equality of opportunity.

It will be helpful to draw on a range of data, for instance existing evaluation evidence, primary research, policy / programme / project documents, strategy reports and appraisal work. These are highly likely to not only include information relating to each of the above categories, but also to indicate how components link together. It is also worthwhile consulting with those stakeholders who were involved in conceptualising the intervention so that their insight can inform understanding of the intervention logic. Involving wider stakeholder groups, such as intended target groups or relevant agencies, non-governmental or voluntary organisations can also be helpful in order to capture a broader range of perspectives about the anticipated effects of the intervention.

The intervention logic is a visualisation of the information collected on the relevant questions under each category, and on the links between them. It will initially display the connections made explicitly, or even implicitly, in documents or by stakeholders and will require the evaluator to analyse this information so as to extract details about individual components as well as links between them. There are three main benefits to mapping the intervention logic at an early stage of designing the evaluation.

1. Modelling the intervention logic will firstly illustrate where any knowledge gaps or limitations in the evidence base exist. This will allow the evaluation effort (in terms of evaluation questions, methodologies and resources) to be directed on generating evidence that fills these gaps. Once the intervention logic has been drawn up the knowledge gaps it highlights will need to be considered carefully

when phrasing the initial evaluation questions. Step 4 is designed to help the reader achieve this.

2. The intervention logic helps highlight the relative distance of anticipated impacts from the inputs and activities and the extent to which these can already be quantified. Outcomes are those 'first order' impacts (achieved in the short or medium term) that result immediately from the outputs of an intervention (e.g. reduced congestion). Impacts are those longer term results of transport interventions that often relate to the social, health or environmental domain (as, for instance, expressed in DfT's strategic goals for delivering a sustainable transport system, in relation to climate change, competitiveness and productivity, safety, security and health, quality of life and equality of opportunity[14]). Being clear about the intended outcomes and impacts of the intervention will be helpful in two respects: Firstly, in identifying what the evaluation should measure; and secondly, to inform considerations about which evaluation approach will enable these outcomes and impacts to be attributed to the intervention. This will be explored in more detail in Step 5 and Step 6 below.

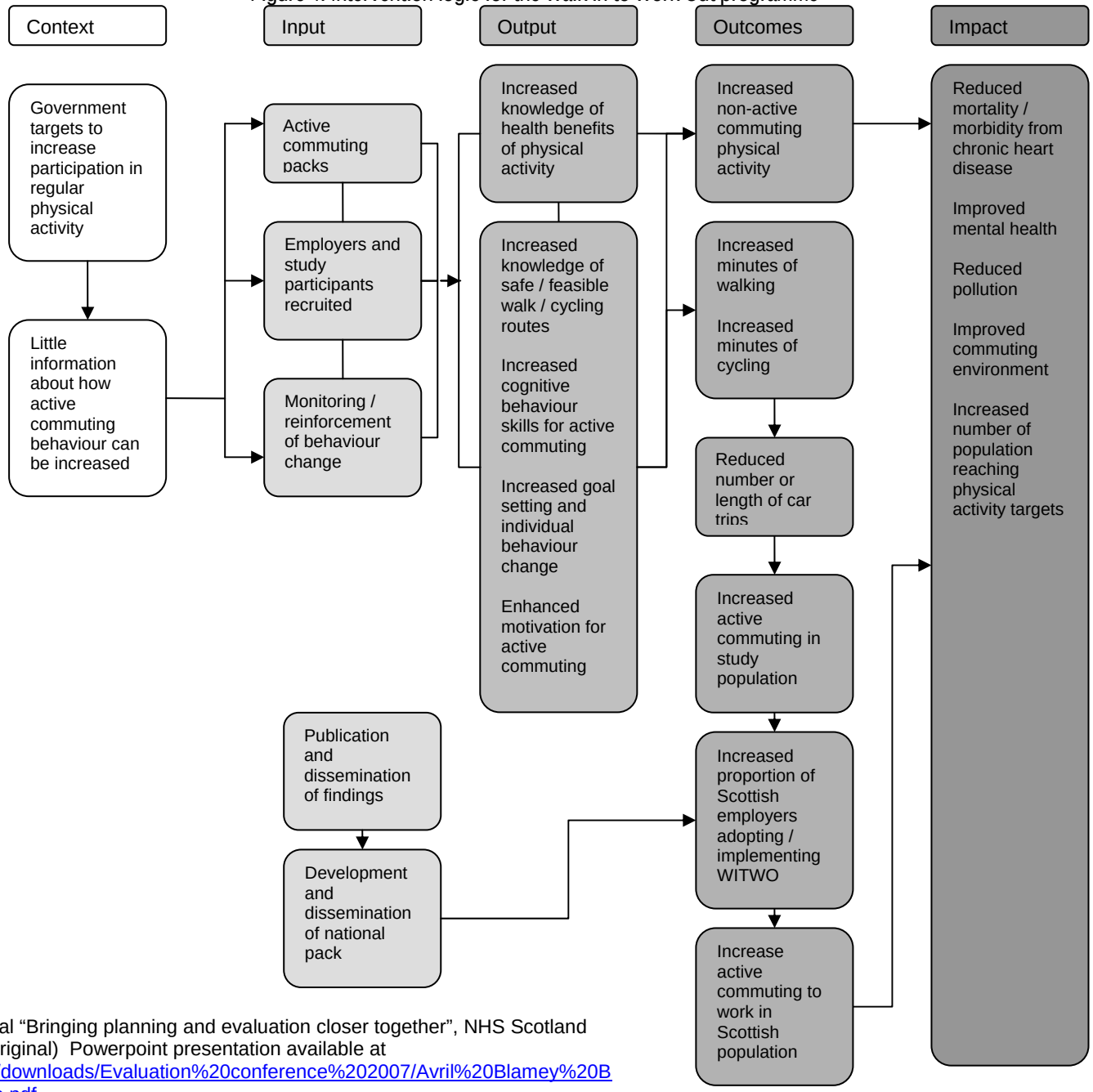
3. The logic map may highlight where links between input, output, outcomes and impacts are unclear. It may also raise new questions relating to any one of the dimensions. Mapping the intervention logic at an early stage of the evaluation design will inevitably mean there will be gaps as not all components will be identified, and links are likely to be missing or merely be speculated about. Similarly the more complex the intervention the more complex the intervention logic map is likely to be, reflecting the larger number of activities, outcomes and impacts. However, if resources for the evaluation are limited, focusing on knowledge gaps will help reduce the number and scope of the questions asked to a manageable level. Often this process of refining and amending will be part of a negotiation process between the different parties involved in the evaluation (commissioner, evaluator, those delivering the intervention).

Figure 4 overleaf provides an example of a simplified intervention logic[15] for the "Walk in to Work Out" programme, which was an active commuting intervention run in the Glasgow area. The rationale for the study was to encourage workplaces to highlight existing facilities for active commuting and the evaluation approach was based on a randomised controlled trial (see section 6.1 for more information about randomised control trials) to establish if a self help intervention could increase active community behaviour.

[14] Further information about DfT's strategy "Delivering a Sustainable Transport System" is available from the DfT website www.dft.gov.uk/about/strategy/transportstrategy/

[15] This provides an illustration of a logic model and is not intended as a generic template. When developing the logic model the evaluation designer will need to find an appropriate model structure which reflects the complexity of the intervention.

Figure 4: Intervention logic for the Walk in to Work Out programme



Source: Blamey, A et al "Bringing planning and evaluation closer together", NHS Scotland 2004 (modified from original) Powerpoint presentation available at www.bhfactive.org.uk/downloads/Evaluation%20conference%202007/Avril%20Blamey%20BHFNC%20Conference.pdf

4. STEP 4: DEFINING THE EVALUATION PURPOSE AND FRAMING THE QUESTIONS

Step 4 encourages the reader to apply their thinking from the previous three steps to their planned evaluation. It offers the opportunity to finalise two important issues: defining the exact purpose of the impact evaluation and refining the evaluation questions by drawing on the intervention logic developed in the previous Step 3. Both are important for choosing an approach that best meets the needs for the evaluation.

Having worked through the first three steps, the reader will have achieved the following:

- Reminded themselves of the evaluation context (Step 1);
- Clarified the nature of the intervention they are looking to evaluate (Step 2); and
- Drawn up, or revisited, the intervention logic for the intervention under investigation (Step 3).

This Step 4 is designed to help pool the knowledge gained from working through Steps 1 through to 3 so that it can be applied to the planned evaluation. This Step 4 therefore provides an opportunity in particular to consider the evaluation questions and to check that they are, firstly, suitable to the type of intervention and evaluation planned and that they are, secondly, focused on those areas where the intervention logic shows the most important knowledge gaps. Following Step 4 the reader will therefore be able to start considering the most effective approach for their evaluation requirements and for demonstrating attribution.

Defining the evaluation purpose

Impact evaluations often have multiple purposes which will reflect the questions and evidence requirements across a range of stakeholders. Generally, being able to attribute the change in impacts to the intervention will be particularly relevant for evaluations which will provide evidence for accountability purposes and /or are designed to generate new knowledge about an intervention.

The reason for clarifying the purpose(s) of the evaluation is it will inform decisions about the types of questions the evaluation will be designed to answer. This is particularly important if evaluations are multi-purpose as the design will have to accommodate the range of requirements. Below this is illustrated further using the examples of two types of impact evaluation purpose.

Knowledge based evaluations tend to be undertaken to increase understanding of what interventions work in which circumstances and why. They provide evidence which can be used to inform the development of future interventions. Therefore, the generation of transferable lessons is important. Undertaking a knowledge based evaluation may, for instance, be particularly useful for interventions where there is little prior evidence to demonstrate that the intervention has caused the desired

impact (e.g. because it is a new policy) or where shifts in policy require an updated investigation of an established intervention. One of the reasons for carrying out a knowledge based evaluation is to provide evidence on an intervention to inform future transport appraisal assumptions (e.g. by testing systematically some of the appraisal assumptions or by offering additional evidence on top of those factors conventionally included in appraisals). In this case, integrating evaluation findings into appraisal work will make a contribution in future option generation.

The case study below offers an example of a recent knowledge based evaluation in the field of road safety.

Case study 1: The evaluation of local road user safety policy and practice – a knowledge based evaluation

In 2007, the Department for Transport commissioned an evaluation of local road user safety policy and practice.

Road user safety policy is being implemented locally, and there remain knowledge gaps in the areas of scheme, policy design and implementation, especially in relation to how these affect the impact of local road user safety policy. The evaluation therefore needs to provide evidence on ***what has worked and not worked, for whom, in what circumstances, and why*** to help DfT and Local Authorities achieve road safety targets by learning from good practice. A Theory of Change approach (see Step 5) is being used to meet these requirements because it allows the evaluators to systematically analyse the causal chains through which road safety policies achieve impacts locally.

Further information: http://www.dft.gov.uk/rmd/project.asp?intProjectID=12528

Because of their focus on uncovering 'what works and why', knowledge based evaluations tend to ask questions which:

- Are aimed at understanding cause and effect relationships (causal);
- Test assumptions behind an intervention to support change (critical); or
- Identify the main issues relating to an intervention where little prior knowledge exists, for instance what factors contribute positively or negatively to its impact (exploratory).

Understanding why change was produced is important in knowledge based evaluations. Therefore, their design must be sufficiently sophisticated to ensure relevant information is collected in a methodologically rigorous way.

Accountability evaluations tend to be undertaken where there is a need to demonstrate that the investment into the intervention has delivered the impacts anticipated at the appraisal stage. When planning an accountability evaluation it is important to demonstrate how far an intervention has achieved its objectives and how effectively resources were used. Evaluation questions tend to focus on asking whether the predicted outcome was achieved, with less of a focus on understanding how and why. They are more likely to provide evidence which observes and measures change in the impacts. Case study 2 provides an example of an evaluation with a strong accountability focus.

Case study 2: Post Opening Project Evaluations (POPE) – an accountability evaluation

POPE is a methodology used by the Highways Agency (HA) for the evaluation of all its major road schemes and Local Network Management Schemes (LNMS). A key requirement of major schemes evaluation for HA policy makers is to show that schemes deliver predicted benefits. A POPE study therefore "will look at how much a scheme actually cost (outturn cost) compared to predicted costs. These costs include: construction costs, land, preparation and supervision costs. The study also calculates the benefits of the scheme and compares this to the benefits which were forecast."[16] There is therefore a strong accountability dimension to POPE evaluations, probing whether the investment was sound (though learning for future appraisals is also an important dimension).
Source: http://www.highways.gov.uk/evaluation and Highways Agency interview conducted for preparing these guidelines

Although their focus is typically on the question of whether the scheme has worked or not, accountability evaluations can still benefit from a more in-depth investigation of the relationship between an intervention and the measured impact. For example, to answer questions such as why and how an intervention achieved an impact (and not just that it did), or what factors were most important for achieving it. Thus, local or national stakeholders may decide that an evaluation that is purely geared towards showing what was achieved is not sufficient and that it is also important to additionally understand why an outcome was achieved and to gather transferable learning for future interventions. If the accountability evaluation is undertaken for interventions that are delivered in a contested or politically sensitive area, it is more likely to require an in-depth understanding of the effects (positive or negative) the intervention caused and sound evidence that the outcome justifies the investment.

Clearly, all of these considerations suggest that there will be circumstances where evaluations will be designed to answer a combination of knowledge based and accountability questions.

Table 2 below is designed to help the development of evaluation questions based on the purpose of the evaluation and the nature of the intervention. This table allows the reader to check that a comprehensive and clear set of evaluation questions have been framed in relation to the evaluation purpose. This is particularly important if the impact evaluation is required to address multiple purposes. Undertaking this exercise should provide the reader with a better understanding of the links between evaluation purpose and the respective requirements for demonstrating attribution.

[16] See also: http://www.highways.gov.uk/evaluation

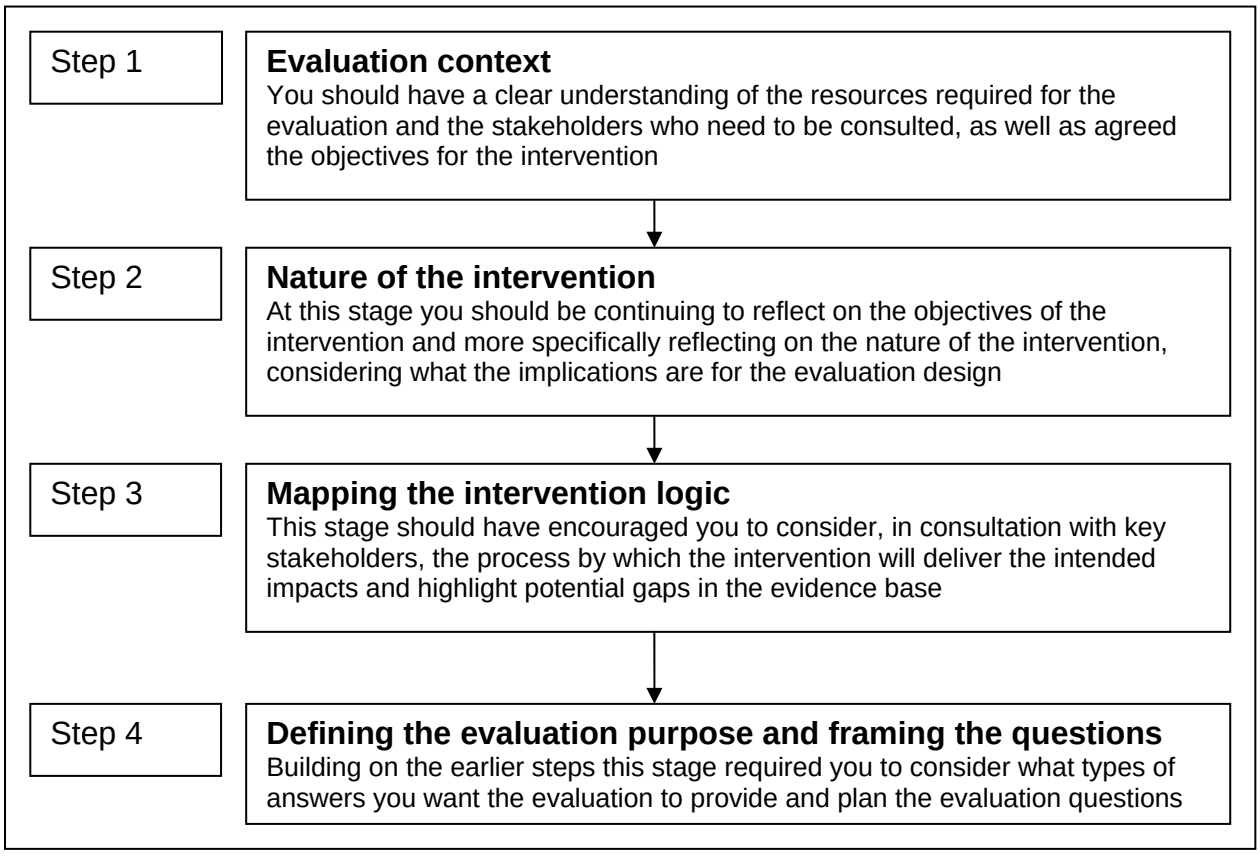
Table 2: Evaluation purpose, approach and type of evaluation questions to ask

	Focus	Typical evaluation questions[17]	Transport relevant examples	Your evaluation questions
Knowledge based evaluation	Explaining what aspects of the intervention have caused the observed impacts by investigating the links between: the investment in the intervention - its outputs – short term outcomes – longer term impacts; alongside understanding the influence of wider contextual factors. Delivery of the intervention may also be considered.	*Causal*: questions which strive to understand and assess how and to what extent that which occurred is attributable to the intervention	*What impact has the intervention had on changes in travel times, traffic flows, land use patterns, the socio-economic make-up and environmental improvements?*	
		Critical: questions which are intended to explore the underlying rationale for the intervention and confirm whether this has been successful or where improvements can be made	*Were the original assumptions made on the impact of the intervention correct?* *Did it have any unanticipated impacts?* *How might the intervention have had a bigger impact on travel to the area?*	
		Explanatory: questions aiming to identify the main factors with which an intervention achieves its impact (or those that are counter-productive)	*What are the main factors or mechanisms through which the intervention has achieved its impact?*	
Accountability evaluation	Aims to test whether the intervention has delivered the intended outcome based on the targets predicted before the intervention was implemented	*Descriptive*: questions intended to observe and describe changes in outcome/ impact measures	*Did the implementation result in the anticipated change in travel times, traffic flows, land use patterns, socio-economic developments and the environment?*	
		Normative: questions which aim to interpret the results in relation to anticipated targets or goals (are the results and impacts satisfactory in relation to targets, goals, etc?)	*To what extent has the change in travel patterns anticipated from the intervention been achieved?*	

[17] Source: http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/designing_implementing/designing_planning/questions_en.htm

Summary

Being clear about the evaluation focus and the kind of questions that need asking ensures that the purpose of the evaluation is met and, therefore, the evaluation generates the required evidence. By regularly reflecting on the evaluation purpose, the questions will provide a steer towards the kind of evaluation approaches that can be undertaken to answer them. At a later stage of the evaluation design, these precise evaluation questions will also support the decision on the kind of data that need to be collected through the evaluation, the suitability of existing data sources and the primary data collection methods which can be used to fill any gaps in the evidence base.



5. STEP 5: DECIDING ON THE BEST EVALUATION APPROACH

Step 5 helps the reader choose the most appropriate overall approach for their impact evaluation by inviting them to consider whether the key conditions for using one of three approaches is met: outcome study, experimental design or theory-based approaches.

After mapping the intervention logic and framing the evaluation questions, it is now possible to consider the most appropriate overall approach for the evaluation. This is a complex, but important, stage in the evaluation design which requires careful consideration of both the key features of the intervention to be evaluated and the relative strengths and weaknesses of the main possible approaches for transport evaluations.

This section aims to support this decision-making process by taking the reader through the three evaluation approaches (outcome studies, experimental approaches and theory-based approaches) that are available for impact evaluations in the transport field. This step will summarise the main features of each approach, identifying their relative methodological strengths and limitations and consider the extent to which the evidence on impacts they produce allows conclusions to be drawn on whether impacts identified can be attributed back to the intervention. Working through this step the reader must systematically review the approaches against their evaluation requirements in order to identify the approach which will most effectively produce the type of evidence required.

5.1. When to choose an outcome approach

Outcome studies (sometimes referred to as summative evaluation) compare the situation prior to the introduction of an intervention with that after its introduction. Any observed changes which reflect anticipated effects are assumed to have resulted from the intervention. Outcome studies therefore answer the question *'to what extent have predicted outcomes and impacts been achieved?'* They are most suited to evaluations for accountability purposes.

At the appraisal stage, assumptions on the outcomes and impacts of the intervention are likely to have been made, drawing on information gained through extensive research and modelling. The focus of an outcome evaluation is likely to be the collection of numerical data measuring the extent to which these predicted outcomes have been achieved. Under the right circumstances, outcome studies can be effective at demonstrating attribution. However, they can also be inappropriate in many cases. These circumstances are outlined below.

Strengths of an outcome study approach

A purely outcome study design can be useful where the only purpose of the impact evaluation is to demonstrate accountability. This approach is often suitable for answering questions which measure the extent to which the predicted outcomes have been achieved. It can be suitable for interventions for which there is an established body of knowledge about impacts and effectiveness. Examples include traffic calming measures (such as, speed bumps, speed cameras, road narrowing measures) and some well established infrastructure interventions. In particular, when suitable monitoring data is already available, this approach can provide evaluation evidence with minimum resource investment and might therefore be suitable for relatively smaller scale or lower cost interventions. In larger scale interventions, the outcome approach is greatly enhanced if earlier appraisal work has provided a robust calculation of anticipated impacts compared to the 'do minimum' alternative (counterfactual). Where there is strong emphasis on cost benefit analysis, the outcome approach will also be strengthened by consideration of issues such as deadweight, displacement and substitution effects, as advised by the Green book[18].

Limitations of an outcome study approach

However, a key limitation of the outcome approach is that it does not offer proof of a causal relationship between an intervention and an observed outcome where there are a lot of external factors that may have an influence on the results. When used in such conditions, there can be less confidence in the evaluation results.

Outcome evaluations also do not provide evidence about why the intervention was successful (or not). Therefore, it is not suitable for evaluations where the key stakeholders need to understand how or why (i.e. the causal pathways) the outcomes and impacts were achieved or to learn lessons about the effectiveness of the intervention to inform the design of future initiatives.

The nature of this approach to evaluation means that the causal relationship between the intervention and the outcomes and impacts are predicted but never fully tested. This is particularly problematic in circumstances where other factors might influence the intended impacts, for example, other interventions which are initiated during the evaluation period or external economic conditions. In larger projects, some of these other factors may have been addressed through initial modelling, but these will not take into account unanticipated changes in the wider environment.

A key risk for this approach is if the evaluation's findings identify a significant deviation between actual outcomes and those that were originally anticipated, because the evaluation itself will give little indication of whether the gap between anticipated and actual outcomes is because initial assumptions on which the intervention was planned being incorrect, the plan was never fully implemented, or because of unanticipated factors intervening.

[18] www.hm-treasury.gov.uk/data_greenbook_guidance.htm. Chapter 7.

There is also a danger that the evaluation findings will not provide sufficient data concerning any unintended consequences, i.e. unexpected outcomes or impacts which were not related to specific objectives and targets. Outcome evaluations tend to focus their resources mainly on gathering data related to initial objectives therefore, data relating to wider impacts might not be collected or be overlooked.

Circumstances which support the use of an outcome approach

Although this approach is useful for comparing actual effects against those predicted its value rests on being carried out in a methodologically rigorous way. In order to ensure the reliability of the evaluation and increase the ability to attribute the observed changes to the scheme, outcome studies should only be considered in the following circumstances:

1. **Scope of intervention:** The objectives of the intervention have a clear primary purpose and should not be aiming to respond to a number of issues.
2. **Context of intervention:** The intervention is delivered at a local or sub-local level, and in an environment which will remain relatively stable during the time in which the outcomes and impacts will be delivered. For example, caution should be taken if it is being considered for use in circumstances when additional interventions are likely planned to be introduced during the evaluation period which could also influence the observed results.
3. **Mode of implementation:** The intervention is implemented consistently and delivery is directly controlled by the organisation responsible for the intervention even if it is delivered through other organisations.
4. **Intervention timescale:** Anticipated outcomes and impacts are expected to occur over a relatively short time period (between 1-3 years post implementation), even though they might be sustained for much longer time periods.
5. **Nature of anticipated outcomes:**

- The interest of the evaluation is to measure short or medium-term outcomes which are expected to be directly caused by the intervention (for instance evaluating the effect of speed bumps) rather than longer-term and wider impacts.
- These outcomes can be measured using robust monitoring data (including administrative data), ideally directly measured using a continuous time series data commencing several months or years before the intervention took place. This would provide evidence concerning the situation prior to and after the intervention without the need for using proxy measures.
- In addition, the initial appraisal model and / or logic mapping should have indicated that the intervention will generate a large change and that there are likely to be few alternative explanations for the change(s) taking place.

If these conditions are met, then there are <u>two further considerations</u> which will inform the decision to undertake an outcome evaluation:

- The intervention is relatively low-cost, indicating that the scale of investment required for alternative evaluation approaches would be out of proportion to the cost of the intervention; and,
- Appraisal modelling has taken place which gives clear indications of the anticipated results in the situation without the intervention taking place (i.e. a ‘do

minimum' scenario or counterfactual), against which actual outcomes can be compared.

The logic mapping undertaken in Step 3 offers a framework for systematically considering the circumstances for the evaluation and hence whether the outcome approach is suitable for the evaluation of the intervention. The intervention logic can also offer a framework for the collection of data on the effectiveness of the implementation (did this go according to plan?), and concerning any other changes taking place in the wider context that would add understanding to the interpretation of the evaluation results. This will help to ensure that evaluation data is available to account for any significant deviation between the actual and the anticipated outcomes.

If outcome studies are used to evaluate the impact of interventions delivered in different circumstances to those outlined above, then there is a real danger that wrong conclusions would be drawn from the data gathered. Two kinds of mistakes might then be made:

- Outcome studies may produce **false negatives** where little or no change is observed and this is wrongly attributed to the intervention not being effective. For example, the building of a by-pass apparently leads to little improvement in traffic congestion, but the benefits of the by-pass has been counteracted by an unanticipated growth in traffic generated by the arrival of a new business in the area.
- They may produce **false positives** where change is falsely attributed to the effectiveness of a particular intervention. For example, in Australia a fall in cycling casualties was linked with – and attributed to – the wearing of cycle helmets but could equally have been attributed to the fact that the law requiring their use had also led to many children giving up cycling.

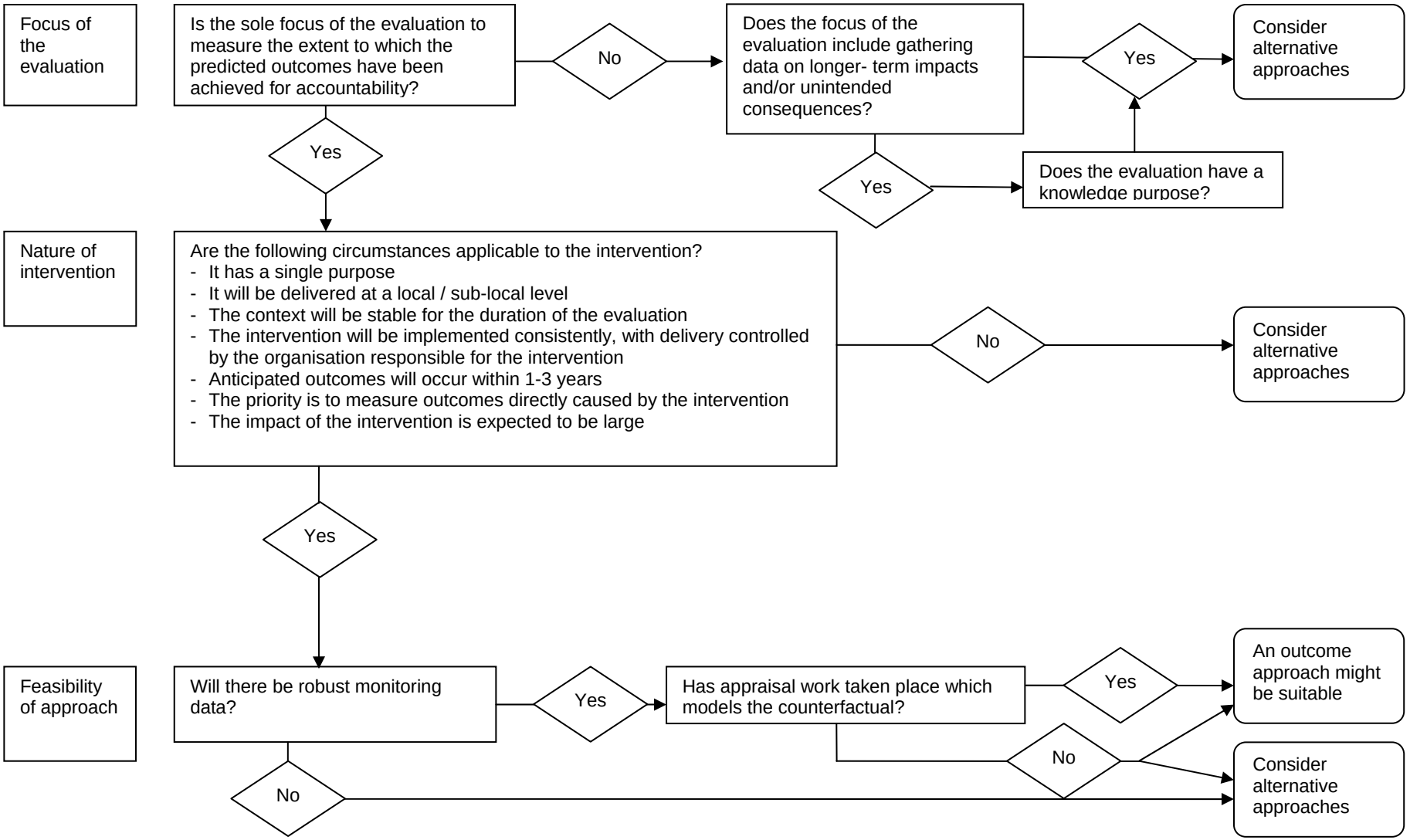
Additionally questions might be raised about the robustness of the data. Evaluations might then be seen as offering limited value compared to those using alternative approaches, in terms of providing an evidence base for demonstrating the success of the intervention and developing new policy or future appraisal activities.

As has been highlighted above, an impact evaluation can have numerous purposes. It is therefore possible that on balance an outcome approach, if used correctly, answers most of the evaluation questions asked for accountability purposes. However, the evaluation may have an additional knowledge dimension which the outcome approach is not able to address. In this case, it is advisable to choose a mixed approach. An option for this is discussed in section 6.3.

Flowchart for choosing an outcome approach

Figure 5 below offers a flowchart to help you identify whether the circumstances of the intervention are most suited to an outcome approach in delivering attribution led transport impact evaluations.

Figure 5: Flowchart for the selection of an outcome approach



5.2. When to choose an experimental design approach

The **experimental approach** is designed specifically to test causal relationships and answers questions about whether the intervention has had the anticipated results. It offers scientific proof of a cause-effect relationship between an intervention and an outcome so that conclusions from an evaluation drawing on experimental approaches offer a robust evidence-base for future decision-making.

The aim of the approach is to compare the impact of the intervention with what would have happened anyway without the intervention. It analyses two populations, one in receipt of the intervention and one without the intervention in order to observe what happens in the absence of the intervention. This evaluation approach provides evidence that the intervention has been successful if the group receiving the intervention (known as the treatment, programme or experimental group) shows significant changes in the impact measure compared with the group not receiving the intervention (referred to as the control or comparison group). Experimental approaches are therefore particularly suitable for evaluations that have an accountability purpose.

The focus of experimental designs is to produce robust quantitative data about the impact of an intervention on the intended population by showing the change produced. Therefore, experimental designs are particularly useful for evaluations which need to deliver rigorous data on the extent to which an intervention has led to the intended change in impacts, over and above what would have happened in the absence of the intervention.

Strengths of an experimental design approach

Experimental evaluation designs can be a robust approach for assessing impact which can be confidently attributed to an intervention. When used in the right circumstances, this approach can provide robust evidence to answer evaluation questions that ask about whether the intervention has produced the observed change. More specifically, experimental designs are highly effective in answering evaluation questions which are geared towards testing or validating the model underlying a service intervention, or comparing the effectiveness of the intervention under different conditions[19]. Experimental designs provide robust quantitative data which can be integrated into appraisals.

This approach is therefore suitable in answering the following types of evaluation questions:

- Questions that are geared towards demonstrating that an intervention is responsible for a particular change;

[19] Cullen, J. and Hills, D. (1996), *The Role of RCTs in Assessing Services Effectiveness: a Critical Review*, EDRU Occasional Paper, London: The Tavistock Institute

- Where the primary interest is in testing or validating the model underlying the intervention, or comparing the effectiveness of the intervention under different conditions;
- Where the intention is to measure outcomes (i.e. short to medium-term impacts);
- Where the intention is to measure a single impact, rather than on assessing the impact of the interaction of several outcomes.

Limitations of an experimental design approach

Experimental designs do not tend to be suitable for all types of intervention and have particular financial and ethical implications which are described later in this section. Some other limitations of this approach are highlighted below.

This approach does not provide evidence to explain ***why*** the intervention was successful (or not) in meeting its objectives or why some people receiving the intervention responded differently to others. Both issues might be of interest to those responsible for the implementation or delivery of future initiatives. Additionally, the evidence will not provide understanding about any unintended impacts or displacement effects caused by the intervention, for example, traffic moving to another route to avoid a particular road intervention.

Further, it is not always appropriate to use the findings from a single experimental evaluation to infer what would happen if the intervention was extended to a wider population. This is because the evaluation findings are specific to the particular setting, context or time period and might not be replicated in other situations. However, it is possible to make generalisations from evidence gathered from a number of evaluations of the interventions in different settings.

It is possible that the experimental nature of the evaluation can have a direct influence on the outcome of the intervention. The Government Social Research (GSR) Magenta Book explains that "where individuals or institutions are knowingly assigned to either programme or control group, the fact that units are aware of their control (or experimental) status may alter their behaviour"[20]. If this occurs it can become difficult to disentangle the effect of the evaluation approach from that of the intervention. The most robust approach to overcoming this possible influence is to avoid disclosing who is in which group (known as blinding), but in reality for transport interventions this is rarely possible.

Circumstances which support the use of experimental design approaches

Experimental approaches should be considered in the following circumstances:

1. **Scope of the intervention:** The intervention is designed to achieve a single goal, it can be delivered in isolation from other activities which may also influence the goal and the intervention can be allocated to parts of the population whilst being withheld from others. Case Study 3 below provides an example of this.

[20] GSR Magenta Book, 2005, Paper 7: Why do social experiments?

Case study 3: The Glasgow Walk in to Work Out initiative[21]

This scheme has a single goal: to test whether written interactive material could increase active commuting behaviour in work places. These materials were produced and distributed to some people and not others in order to establish treatment and control groups. The scheme was not co-ordinated with other activities, but if, for example, a second initiative had been launched on the same population (e.g. to reduce car parking provision for congestion benefits) then this might also have influenced the observed numbers walking to work and it would have been more difficult to discern what the 'Walk in to Work Out' scheme achieved.

2. **The intervention context:** If the target population for the intervention is area-based (rather than targeting individuals or institutions), it will need to be implemented at a local (within a county) or sub-local level. Firstly, this is so that comparator groups within the same areas can be identified, something that will not be feasible if the intervention is delivered nationally. Secondly, this smaller geographic scale also makes it easier to minimise variation between the treatment and comparator areas. Comparing across counties or regions is not recommended as there are often inherent differences between the areas (such as local policies or administrative systems) which can effect the interpretation of the findings as it is not possible to distinguish the extent to which observed change can be attributed to the intervention or is caused by differences between the two groups.

   Further, the intervention should not be expected to have differential impacts caused by the setting it is implemented within. Setting refers to the neighbourhood or area in which the intervention is implemented, that is the socio-economic, ecological, physical, political or other environment in which an intervention is located. So for example, if it is anticipated that the intervention will affect disadvantaged groups differently from the rest of the target population or if an objective of the evaluation is to understand why particular social groups respond differently to the intervention, then experimental designs are inappropriate approaches. They are designed to provide evidence for whether an intervention has had an impact, but not how or why.

   Finally, the evaluation will observe impacts during a time when the political environment will remain relatively stable. For example, it should not be used in circumstances when additional interventions are planned to be introduced during the evaluation period that could influence the observed results.

3. **The mode of implementation:** Once the intervention has been implemented it will not be changed during the evaluation period. This means that the intervention is unlikely to be modified, and that no other major initiative will be implemented that will have an impact on the initiative or the outcomes. Similarly, if the intervention is implemented over multiple sites it has to be consistently delivered. Otherwise, the changes which are caused if the intervention does not remain constant might influence its impacts.

[21] Mutrie, N et al (2002) "'Walk in to Work Out': a randomised controlled trial of a self help intervention to promote active commuting" *Journal of Epidemiology and Community Health*, Vol 56, pp. 407-412s

4. **Intervention timescale:** The results of the intervention will take place in a relative short timescale post implementation (between six and 18 months). The longer the gap between intervention and outcome the more likely it will be that these will be influenced by changes in the population, new and related interventions, changes in economic circumstances etc.

5. **The nature of anticipated outcomes: there should be little or no uncertainty about the outcome of an intervention**. The experimental design requires some prior knowledge about the possible range of outcomes an intervention is able to achieve. This is necessary to allow the evaluator to develop hypotheses grounded in research or experience with previous interventions which can be tested via the experiment. Without this prior knowledge it is not possible to interpret the findings correctly. This suggests that experimental designs will be particularly beneficial for the evaluation of interventions which have an established evidence base (but are being implemented in new circumstances).

   The interest of the evaluation is to measure the outcomes which are expected to be directly caused by the intervention. The Walk in to Work Out evaluation, for instance, tested the increase in walking and cycling to work as a result of the material produced. Another experimental evaluation (the evaluation of the National Child Pedestrian Training Pilot projects, 'Kerbcraft'[22]) tested whether practical training would develop and improve children's safety skills as pedestrians. Thus, experimental evaluations cannot be used to measure longer-term and wider impacts. This is in order to minimise the risk that any other factors will have an influence on the observed impacts. If the impacts are going to take a number of years to be achieved (such as, for example, long term health impacts) then the probability that observed impacts have been influenced by factors other than the intervention is increased.

   It must also be possible to clearly define and measure the expected outcomes of an intervention for an experimental design to be successful (these should be agreed across stakeholders).

   Finally, **it is not necessary for the observed effect to be very large in order to conduct an experiment**. Rather, experimental design is particularly useful if anticipated observed (e.g. through ongoing monitoring) effects are smaller, as with very large direct affects a simple before–and-after study is likely to suffice.

As you are considering these circumstances it is worth noting that transport interventions are often delivered within relatively long timescales (three years and longer) and often occur as part of packages. Therefore, care must be taken when deciding on an experimental approach.

Is the experimental design appropriate?

Once it has been decided that the intervention lends itself in principle to the use of an experimental approach, and that the evaluation questions are suitable, then the practical feasibility of applying it to the intervention should be explored. The following points need to be considered:

- The ability to control who receives the intervention and who does not and to track the behaviours of these groups over time. Therefore, consideration needs to be made about ways of minimising the number of people who may wish to stop participating in the intervention or evaluation.

[22] See www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme1/childpedestrianprojects/networkchildpedestrianhtml

- A final consideration on the appropriateness of experimental designs is the resources available for the evaluation. Experiments can be highly expensive, so it is important to assess early on whether the budget available for an evaluation permits an experimental design.

Is the experimental design ethical?

Before deciding on whether to choose an experimental design, an ethical assessment should be undertaken. For this, the following questions will need to be asked[23]:

- ***Will members of the control or experimental groups be exposed to known risks or known harmful outcomes?***

  If there is a known risk attached to participation in an experiment, the approach should not be chosen.

- ***Will the experiment mean members of the control or comparator group are denied services which are known to be beneficial on the basis of existing evidence?***

  For instance, a planned evaluation might seek to assess the effectiveness of a road safety education measure in a particular area, but robust research evidence from elsewhere already indicates that this kind of activity is highly effective in reducing casualties. This would create an ethical dilemma as withholding the intervention from some study participants would be likely to put them at greater risk. The evaluators would have to find a way to address this, either through informed consent or the study design (e.g. by delaying participation in the intervention rather than withholding it).

- ***Will members of the control group be denied access to services to which they have an historical entitlement?***

  Take, for instance, the hypothetical example of an evaluation which seeks to test the extent to which the travel behaviour of older people is influenced by benefits such as free bus passes. An experimental or quasi-experimental design would involve withholding this benefit from a group of pensioners for the duration of the study. In cases like these, an experiment would be unethical and the evaluator would need to make provision for this, either through ensuring informed consent by participants (see below) or an alternative research design.

In case of the last two questions, asking participants to provide informed consent to the experiment might be one way of addressing any ethical concerns. Individuals have the experiment described to them in detail and are asked to provide written consent to be randomly allocated.

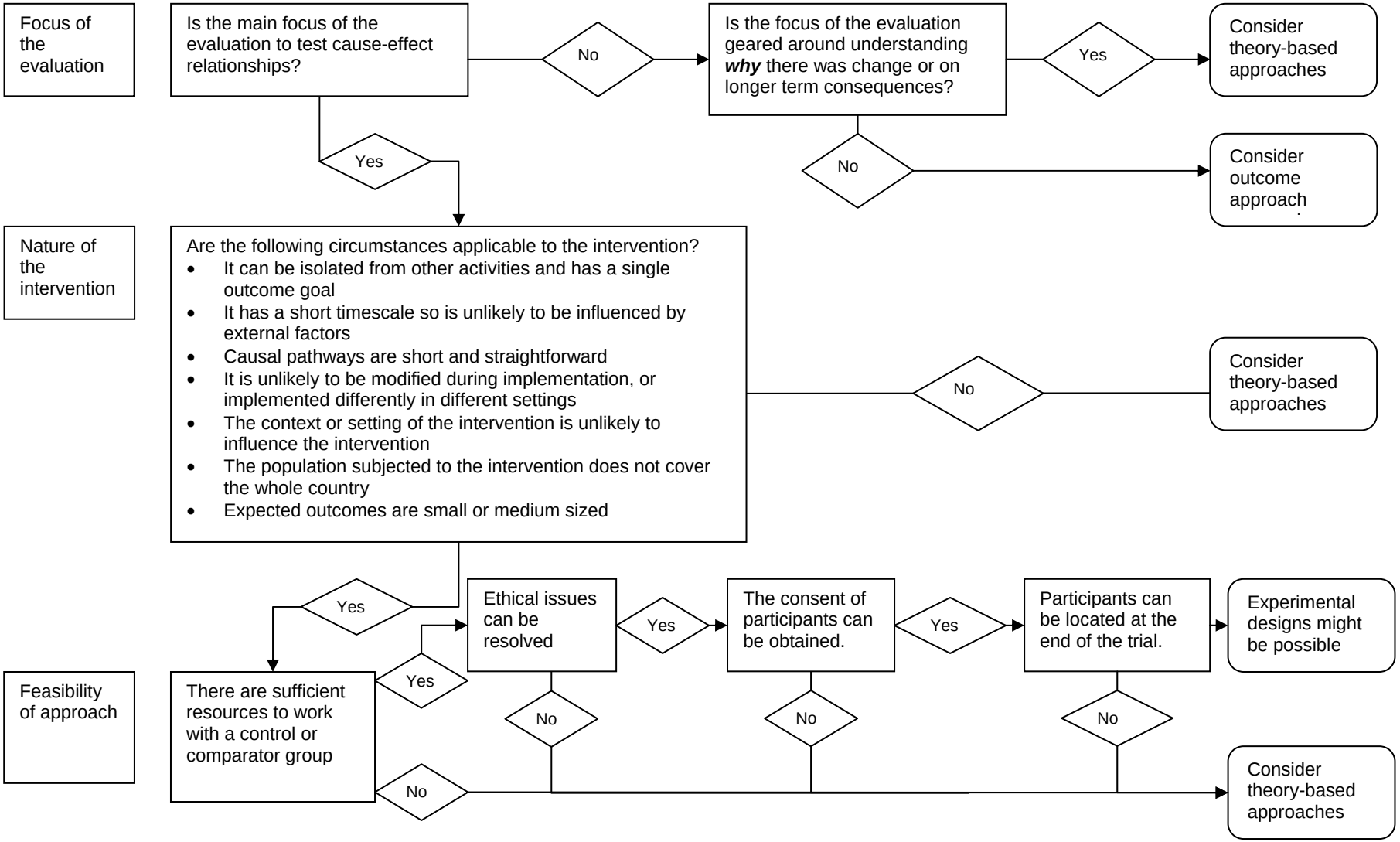
However, unless the benefits of the experimental design can be balanced with the ethical (and possibly legal) implications[24], a different approach should be considered.

[23] See the GSR Magenta Book for more details
[24] See the GSR Magenta Book

Flowchart for choosing an experimental design

Figure 6 below summaries the key circumstances that should be met to make an experimental design methodologically sound for attribution led transport impact evaluations. If they are met, the reader is advised to continue by working through Step 6, see section 6.1. If the checklist below reveals that one or several of these key conditions are not met, the reader is advised to continue with this current Step 5.

Figure 6: Flowchart for choosing an experimental design



As was the case with the outcome approach, this discussion presents the suitable circumstances but these are not automatically sufficient reasons for choosing an experimental design for your impact evaluation. This is because the purpose of your evaluation may require you to find answers to questions this approach cannot offer data for. Furthermore, if you are looking to evaluate a policy, programme or package you may find the approach could provide evidence on the individual schemes but will not sufficiently evaluate the intervention as a whole. If this is the case, you may wish to consider adding an experimental design to a theory-based evaluation framework as part of a multi-methodological evaluation design. You may, for instance, be interested in using an experimental design to test the outcome of a particular scheme as part of a wider evaluation. Such a combined approach is described in section 6.3.

Nevertheless, if on the basis of the discussion above you want to choose an experimental design for your evaluation, you may wish to continue by reading section 6.1 to help you further refine your approach.

5.3. When to choose a theory-based approach

Theory-based approaches to impact evaluation allow for a systematic articulation and testing of the assumed connection (i.e. the theory) between an intervention and the anticipated impacts. The focus of theory-based evaluations is not only on understanding whether an intervention has worked but on *why* and *under what conditions* change has been observed. These issues can be explored using a wide range of research methods (both qualitative and quantitative), and data obtained from different sources will often inform the evidence base (known as triangulation) to strengthen confidence in the conclusions. Theory-based approaches are therefore particularly suitable for evaluations that have a knowledge focus and seek to generate learning for future interventions.

Theory-based evaluations provide evidence on the outcomes and impacts achieved by the intervention (including unintended ones), the combination of factors that contributed to achieving them (this may also include contextual and implementation issues) as well as how outcomes and impacts were achieved.

Undertaking a theory-based evaluation therefore means understanding, systematically testing and refining the assumptions (or theories) behind an intervention. One result of theory-based evaluations can be an elaborate map linking the different elements of an intervention in a matrix highlighting the pathways of an intervention.

Strengths of a theory-based approach

Theory-based approaches are particularly strong for evaluations of very diverse and long term interventions that display a mix of activities, target groups, delivery mechanisms and settings (i.e. interventions that are very complex). That is, they tend to be used in situations where the other approaches discussed in this document would not deliver reliable results because impact pathways are not very straightforward, a number of variables may influence the results and because the research questions are geared towards explaining reasons for change. Theory-based evaluation can also be useful for the evaluation of interventions where little prior knowledge about causality exists as a tool to build up that knowledge. It allows the complexity behind interventions to be articulated, and through the evaluation process the influence of different (combination of) factors on outcomes and impacts to be tested. Therefore, the theory-based approach allows evaluators to consider issues such as the context of an intervention, the way it has been implemented and other social, environmental, political and geographic variables which might also influence the impact of the intervention. Theory-based approaches are also the only methods currently in use that allow inferences to be made about the possible long term impacts of an intervention.

Limitations of a theory-based approach

However, in terms of delivering objective proof that an intervention has worked, theory-based evaluation is not as strong as experimental designs. This is,

essentially, because theory-based evaluation does not involve the construction of a counterfactual (unlike experimental and outcome approaches) through which change can be measured and attributed to the intervention. In part, this is because the evaluation questions that theory-based approaches address are different (rather than asking whether an intervention has worked they ask how and why it has worked); in part this is because it tends to be used to evaluate interventions where experimental and outcome approaches would not deliver meaningful results.

Thus, rather than working with a counterfactual, attribution in theory-based evaluations is achieved by involving key stakeholders in the development of the intervention theory and ensuring that, if the evaluation results confirm this theory, they are satisfied that the intervention has had the desired outcomes and impacts[25]. However, this does not mean that a counterfactual cannot be incorporated into a theory-based approach. For instance, where the evaluation investigates a programme or package, individual schemes may lend themselves to an experimental approach[26]. The evaluator could also investigate whether and how any initial appraisal work might be used for this purpose.

In theory-based approaches, the construction of an initial set of hypotheses of how an intervention will bring about change is the starting point for the evaluation. The range of stakeholders involved in theory articulation may influence interpretation of attribution at the beginning and at later stages of the evaluation as various groups will bring their different backgrounds and hence a diverse range of interpretations to this task. Selecting a balanced group of people to input into this task is therefore very important.

However these initial theories are constructed (be this, for instance, from stakeholder interviews or literature or a combination of both), there is perhaps a risk that the investigation will focus on intended outcomes (both evaluator and informants may develop 'tunnel vision' and only look out for factors they expected to see) and ignore important unintended impacts. Equally, variables which represent important causal factors might be left out or go undetected. Evaluators therefore have to take care to keep an open mind and make use of the opportunities the approach offers to capture the full range of impacts generated by an intervention.

Care should also be taken when considering the scale at which theory-based approaches are applied for an attribution-focused impact evaluation[27]. The more complex an intervention, and the larger the scale at which it is being implemented (e.g. national level), the more factors, mechanisms, groups, sub-groups and circumstances will need to be considered to explain how an intervention has led to the observed impact. In very large evaluations (be this, for instance, because the intervention is implemented at a national scale or in a lot of sites) the theory-based evaluation will need to be designed carefully to manage the resulting complexity (e.g. by choosing a case study approach in multi-site interventions).

---

[25] More explanation on this is given in Step 6.

[26] See also section 6.2.1 on *Choosing a combined approach*.

[27] Mackenzie, M et al (2006) "Using and Generating Evidence: Policy makers' Reflections on Commissioning and Learning from the Scottish Health Demonstration Projects", *Evaluation*, Vol. 2 No. 2, pp. 211-226

Finally, theory-based evaluations may lead to the objectivity of the evaluator being questioned. Theory-based evaluations require comparatively close collaboration between the evaluator and those designing and implementing an intervention in order to draw up and develop the theory of how the intervention will create its impact. It may therefore be felt that the evaluator may lose their position of 'objective outsider'. However, the relatively close relationship with stakeholders will potentially deliver a deeper insight than could be achieved through a more distant stance. Triangulation of evidence from different sources is another technique to generate robust conclusions.

Circumstances which support the use of a theory-based approach

For interventions which are not suitable for experimental or outcome evaluation designs, a theory-based approach could be considered. Theory-based approaches are likely to be suitable for interventions in the following circumstances:

1. **The scope of the intervention:** If the intervention is responding to a fairly new issue, using innovative approaches, and/ or responding to complex policy challenges then the evidence provided by theory-based approaches on why the intervention was successful or not will offer a strong evidence base for future investment decisions. The risk for using alternative approaches in these circumstances is that the limited evidence base available for predicting the impacts of the intervention will mean that the generation of hypotheses or modelling of impacts is less likely to be accurate.

   However, theory-based approaches can also be useful for evaluating interventions where there is an existing knowledge base about possible impacts. In particular, this includes the evaluation of major schemes where comprehensive transport modelling informs the appraisal analysis. Theory-based approaches still may be applicable in these circumstances, even in cases where the intervention is based on well-established initiatives, but where the evaluation evidence is required to answer new questions. For instance, the premise of the evaluation focus may question transport model forecasts or it might be looking to explore a new dimension of a mature policy such as secondary impacts on health, the environment or other dimensions relating to key strategic goals.

   If the intervention seeks to address a number of issues or delivers a number of schemes, then this approach offers a systematic way of measuring the combined influence of the intervention and the relative effects of the schemes on the impacts. Alternative evaluation approaches are not designed to investigate the interaction of multiple schemes on the impacts because they do not explore why the intervention worked.

2. **The context of the intervention:** If the intervention is likely to be implemented in an environment where it is anticipated that other factors will influence its impacts, theory–based approaches would be looking to explore and account for wider contextual as well as unintended impacts. Where an intervention is being implemented nationwide (a situation that makes it impossible to use experimental designs as no comparator areas or population groups will be available), theory-

based approaches still allow the evaluator to investigate impacts and how and why they can be attributed to the intervention[28].

The extent to which the evaluator can be confident that the intervention caused the observed changes can become limited if the **context** in which the intervention is taking place is likely to have a considerable influence on the impact of the intervention. Context in this case could mean the economic, social, political, cultural or environmental circumstances of a particular geographic area (this could be the community level, the local level, the regional or even the national level).

3. **The mode of implementation:** Theory-based approaches are appropriate for interventions that may not remain static but **vary in implementation or change in focus over time**. For instance, where the delivery of the intervention involves multiple actors, especially for interventions which are being implemented in response to localised requirements, differences in the interpretation of desired objectives and variations in implementation are more likely to occur. The implementation of road safety policy, for instance, delivered locally through partnerships may mean that schemes fundamentally differ in their design, target populations and resources. This will have an effect on how the schemes contribute to the intervention's impacts.

4. **The timescale of the intervention:** where an intervention is implemented over a very long timeframe (say, three years and over) theory-based evaluation approaches are also appropriate. Long-term interventions are often subject to change (be this the management structure of the intervention, how it is implemented, possibly even in terms of the range of activities carried out) as learning from implementation feeds back into the system or demands on the intervention shift as the policy situation changes. Impact is also more likely to be influenced by new initiatives emerging as implementation progresses. A theory-based impact evaluation is able to not only capture these changes but to investigate their effects on the causal chain that is being established.

5. **The nature of anticipated outcomes:** In circumstances where the intervention is only one part of a wider set of interventions the observed changes might be the consequence of the interaction of multiple factors. The intervention itself therefore represents a 'contribution' to an overall outcome, rather than the major causal factor. One example for this is the THINK! publicity campaign which aims to make a contribution to the Government's aim of reducing road deaths and serious injuries but interacts with a whole range of national and international safety measures. Theory-based evaluation would be a useful approach in this case, because it would allow the evaluator to identify the processes of interaction with other interventions and thus to pinpoint how THINK! reinforces and adds value to them.

   Theory-based approaches are also suitable for the evaluation of interventions which do not have strong pre-defined expectations of impacts (especially for interventions where appraisals have not been conducted). This may be because

[28] Though see above for possible limitations of the approach as a result of overly complex and large interventions.

work has not been done on creating a systematic map linking inputs to anticipated impacts, perhaps because an intervention is highly speculative or because there are divergent views across stakeholders about the impacts of the intervention

Is a theory-based approach appropriate?

If the nature of the intervention is suitable for a theory-based approach, it should be considered whether the approach will produce the right kind of data to answer the defined evaluation questions.

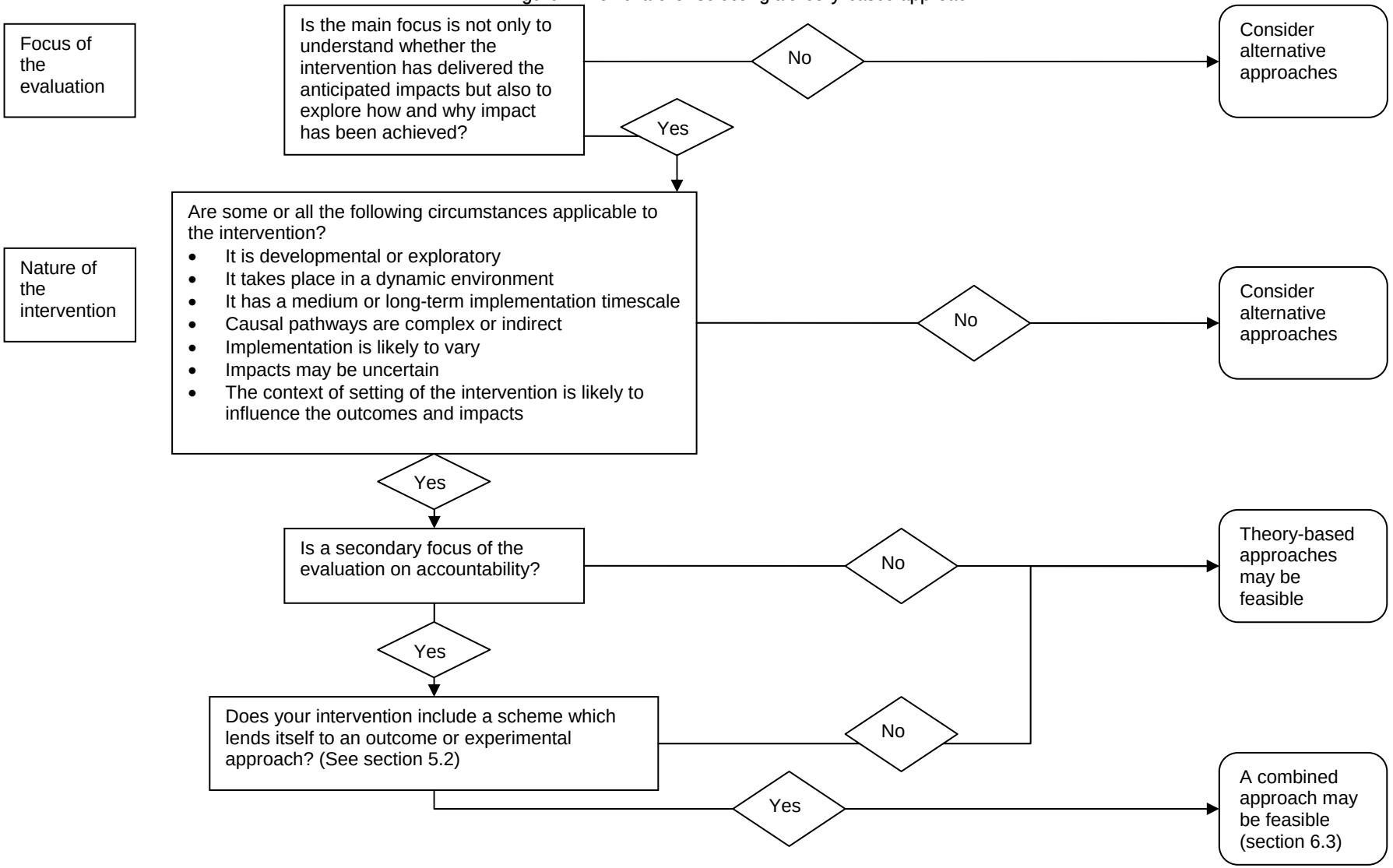
Theory-based approaches generate data on why the intervention delivered the observed impact, to understand what the enablers of success are and to highlight unintended outcomes and their origin. For example, they will provide answers to these types of questions: in which circumstances (administrative, geographic, social or economic settings) was the intervention successful and why; did the target population respond differently to particular components of the intervention, why and which aspect had the most success? The theory-based approach acknowledges the importance or relevance of intermediate stages of delivery, and the context of implementation, as much as measuring the end-result. Investigating how an impact was generated will therefore deliver important learning that can be used for the development of an intervention.

Flowchart for selecting of theory-based approaches

Figure 7 below offers a flowchart to support decision-making on the use of a theory-based approach to attribution led transport impact evaluations. It is not necessary for the intervention to meet all the circumstances; rather, it is sufficient if one or two of them apply. You may wish to continue by reading section 6.2 to help you further refine your approach.

However, where many of the conditions named below do not apply, it could be worth considering the use of a combined approach which connects a theory-based approach with an experimental design if the intervention allows for a control or comparator group to be established. This option is outlined further in section 6.3.

Figure 7: Flowchart for selecting a theory-based approach



6. STEP 6: REFINING THE CHOICE

The final step of these guidelines is designed to help the reader fine-tune their selected evaluation approach so that the best approach is chosen for demonstrating that the observed outcomes can be attributed to the intervention. Therefore, it focuses on experimental (section 6.1) and theory-based (section 6.2) approaches and provides more detail about possible evaluation designs using these approaches or combining these approaches (section 6.3). For readers considering an outcome approach, section 6.3 develops the approach to provide greater attribution and understanding about why the observed change has occurred, by extending the role of the intervention logic model within the evaluation. By the end of this step, you will therefore have made a final decision on which evaluation approach to select and will start to consider the appropriate methods for collecting evaluation evidence.

It is recommended that readers focus on the section relevant to the approach selected in Step 5.

6.1. Experimental Evaluation Approach - Random assignment or quasi-experimental design?[29]

If in Step 5 it has become clear that the intervention meets the key conditions for an experimental approach, the next question that needs to be asked is whether a full experiment using random assignment is possible or whether a quasi-experimental design is more suitable. This section is designed to support this decision-making process.

Designs for undertaking experimental approaches compare two groups with the same characteristics. The only difference between them is that one group receives the intervention and the other group does not. The evaluation will then observe the behaviour of the two groups over time, and the differences in behaviour change will be considered to be caused by the intervention.

6.1.1 Experimental designs

These compare two population groups that have been *randomly assigned* to either participate in (or receive) the intervention (this is called the *experimental group*) or not to participate in (or receive) the intervention (this is called the *control group*). By randomly assigning study participants to an experimental and a control group statistical probability rules are used to ensure that other factors (not the intervention) which might explain an outcome are equally likely to be present in both groups. This means that any observed change in the experimental group can only be due to the intervention. By using randomisation the evaluator can therefore be confident that the outcomes that are being measured can only be attributed to the intervention. One example where a randomised experimental design was applied is the Walk in to Work Out initiative introduced in the previous section (step 5).

[29] This section draws heavily on: Cullen, J. and Hills, D. (1996), *The Role of RCTs in Assessing Services Effectiveness: a Critical Review*, EDRU Occasional Paper, London: The Tavistock Institute

Case study 4: Randomisation in the "Walk in to Work Out" initiative

The 'Walk in to Work Out' evaluation aimed to test whether a 'self-help intervention' delivered to individuals via written materials could increase active commuting behaviour as there had been little prior investigation into ways in which active commuting could be encouraged. Study participants who had been identified as thinking about, or doing some irregular, walking or cycling to work were selected from three Glasgow workplaces which were in the same area of the city and served by a range of public transport links and marked cycle routes. Those volunteering to participate in the trial were randomly split into experimental and control groups. The experimental group received the 'Walk in to Work Out' pack immediately; whereas the control group was told the pack would be forwarded six months later (but did, in fact, not receive it within the lifetime of the study). Follow-up questionnaires used to measure outcomes were sent to both groups after six and twelve months to measure the effectiveness of the intervention. The experimental design offered statistical proof that those individuals who received the pack were twice as likely as those who had not to increase walking to work.
Source: Mutrie, N et al (2002) "'Walk in to Work Out': a randomised controlled trial of a self help intervention to promote active commuting" *Journal of Epidemiology and Community Health*, Vol 56, pp. 407-412s

6.1.2 Quasi-experimental designs

This design is also based on the principle of comparing outcomes from a defined group who receives the intervention with a group not receiving the intervention. However, it is used in circumstances when the allocation to the experimental and control groups cannot be randomised. An example would be situations where an intervention has been designed to cover the whole eligible population, (e.g. if all Key Stage 2 pupils in a secondary comprehensive school received cycle training at the same time). In this case it is not possible to divide these pupils randomly into an experimental and control group because they all would have received the training. Other instances where a random assignment of people into experimental and control groups is not possible is where the issue to be investigated cannot be randomly assigned (e.g. where an evaluator is looking to investigate the effects of gender or age on the outcome of an intervention). In these conditions, quasi-experimental designs can offer a viable alternative to full experiments. In quasi-experimental designs the group not receiving an intervention is called the *comparator group* and is created not by randomly assigning study participants but by selecting the group to be as similar as possible to the treatment group. Using statistical techniques and pre-test measures means that, even without randomisation, it is possible to draw reliable conclusions on the contribution of an intervention to an outcome using quasi-experimental designs.

There are two main methods: the non-equivalent group design and the time series design. A *non-equivalent group design* compares two groups who are not randomly assigned and hence are different (i.e. non-equivalent). Going back to the cycle training example: to measure the outcome of the training on the cycling behaviour of pupils against a counterfactual (i.e. the status quo or non-training situation) the comparator group would have to be Key Stage 2 pupils from another school and therefore differences between the two groups (e.g. in terms of setting of the school and cycling proficiency of pupils) would have to be accounted for. A multiple *time series design* involves periodic measurement of the treatment and comparator group which starts before the intervention is implemented and continues throughout the duration of the 'treatment'. By repeatedly measuring changes in the study groups, the time series design helps the evaluator rule out other explanations for change.

This is because the data build up a trend before, during and after the intervention so that irregularities (e.g. induced by an external event) can be detected.

Ethical reasons might also be behind a decision not to randomise. For instance, offering safety training to one half of the children in one year and not the other on a random basis to test effectiveness could be seen as putting the control group (those who do not receive the intervention) in greater risk of accidents and injuries. On this basis, evaluators might reject the option of randomisation.

One example for the use of a quasi-experimental design is the Kerbcraft evaluation.

Case study 5: The quasi-experimental design of the Kerbcraft evaluation

The aims of the evaluation of the National Child Pedestrian Training Pilot projects (Kerbcraft) were to assess whether the Kerbcraft training package to develop and improve children's (aged between 5–7 years) safety skills as pedestrians had the intended impact. A quasi-experimental approach was chosen because too many children within a school had already completed the training when the evaluation started to make a randomised experiment an option. A further reason was ethical: there were concerns about the safety implications of some of the remaining children in participating schools receiving training and others not. The quasi-experimental design involved matching schools running a Kerbcraft intervention as part of the pilot with schools in the same Local Authority that were not included in the pilot and randomly selecting children from both to participate in the study. The evaluation generated quantitative data which showed that the impact of road safety training interventions was best seen in changes to behaviour and not the result of casualties/fatalities.

Quasi-experimental designs are therefore an alternative where a full experiment cannot be conducted but an assessment of outcomes against a counterfactual situation is still required. However, generally the conclusions on attribution the evaluator can draw from quasi-experimental research are less robust than in experimental designs. This is because quasi-experimental designs are used when it is not possible to create the perfect counterfactual situation achieved by experimental designs through randomisation (for the reasons outlined above). It therefore becomes necessary to create a scenario that resembles as closely as possible the conditions without the intervention. The degree of confidence in the conclusions of a quasi-experimental design therefore depends on how good this simulation is which means much care needs to be taken in the design of a quasi-experimental study.

In Step 5 above a decision will already have been made that the objective of the planned evaluation is to investigate whether the intervention has produced an anticipated outcome and that it meets the conditions necessary for an experimental approach. Therefore, the decision now is whether it is more appropriate to use an experimental design or a quasi-experimental design. As randomisation is key to a successful experimental design, the main question to ask is whether a random allocation of the study population to an experimental and control group can be achieved, sustained and is ethical. To inform this decision there are a number of elements to take into account.

Who or what is the target of the intervention?

Is the **size of your target population** large enough for a full experiment? Applying a full experimental design with randomisation successfully means ensuring that the size of the target population is sufficiently large to make this a viable option. This

means, the study population should be large enough to ensure that your experimental and control group both represent key population characteristics (i.e. that they are statistically equivalent). A very small target population could mean that randomly selecting participants could lead to a biased sample, i.e. that experimental and control groups are not identical in representing certain population characteristics. This will make it difficult to conclude whether any of the observed changes are due to the intervention or the differences in the types of participants included in the two groups. These difficulties also occur if people included in a very small study population leave the experiment before it is finished (also referred to as attrition), perhaps because they lose interest in participating or their circumstances change. If you lose participants from a small experimental group, you might, once again, have a biased sample and the confidence with which you can conclude that the intervention has led to the observed changes is reduced. Therefore, where the target population is very small a quasi-experimental approach would be more suitable, using, for instance, a time series design on the experimental and a non-equivalent comparator group. The repeated measuring – carried out before, during and after the duration of the experiment – would identify any irregularities resulting from a loss of study participants as results would (suddenly) change.

Does the target group receiving the treatment exhibit **atypical or extreme characteristics** from the rest of the population? This makes randomisation difficult as participants in the control group are unlikely to have equivalent characteristics if randomly assigned from the whole population. A study using experimental designs under these conditions would therefore not produce sound conclusions on the impact of an intervention as it will not be able to exclude alternative explanations. A quasi-experimental non-equivalent group design, which matches study participants to be as similar as possible, might therefore offer a more reliable comparator and thus produce better results.

To what extent is there certainty that there will be **no interaction** between the experimental and the control group for the duration of the experiment? Partial exposure to the intervention by control group participants would corrupt the random allocation. The results of the experiment would be biased as a result as the study would not deliver a true comparison between those 'receiving' the intervention and those who do not and thus which observed changes were due to the intervention. For instance, in the case of travel plans, which are often implemented over small geographical areas (whether community, work or school) there might be considerable interaction between intervention and non-intervention groups. Learning might be shared between participants and non-participants at work or while collecting children from school, the behaviour of control group members modified during the study period and therefore the results of the experiment distorted. A time series design could be suitable in these situations as data patterns built up through repeated measuring especially when combined with a non-equivalent group design. Changes resulting from the interaction of the two groups (e.g. learning gains in the comparator group) would be reflected in the data collected and can therefore be factored into the outcome calculations.

This means there should not be any additional people joining the experimental or control group for the duration of the study. As above, this would jeopardise the random assignment of people to experimental and control groups and introduce bias

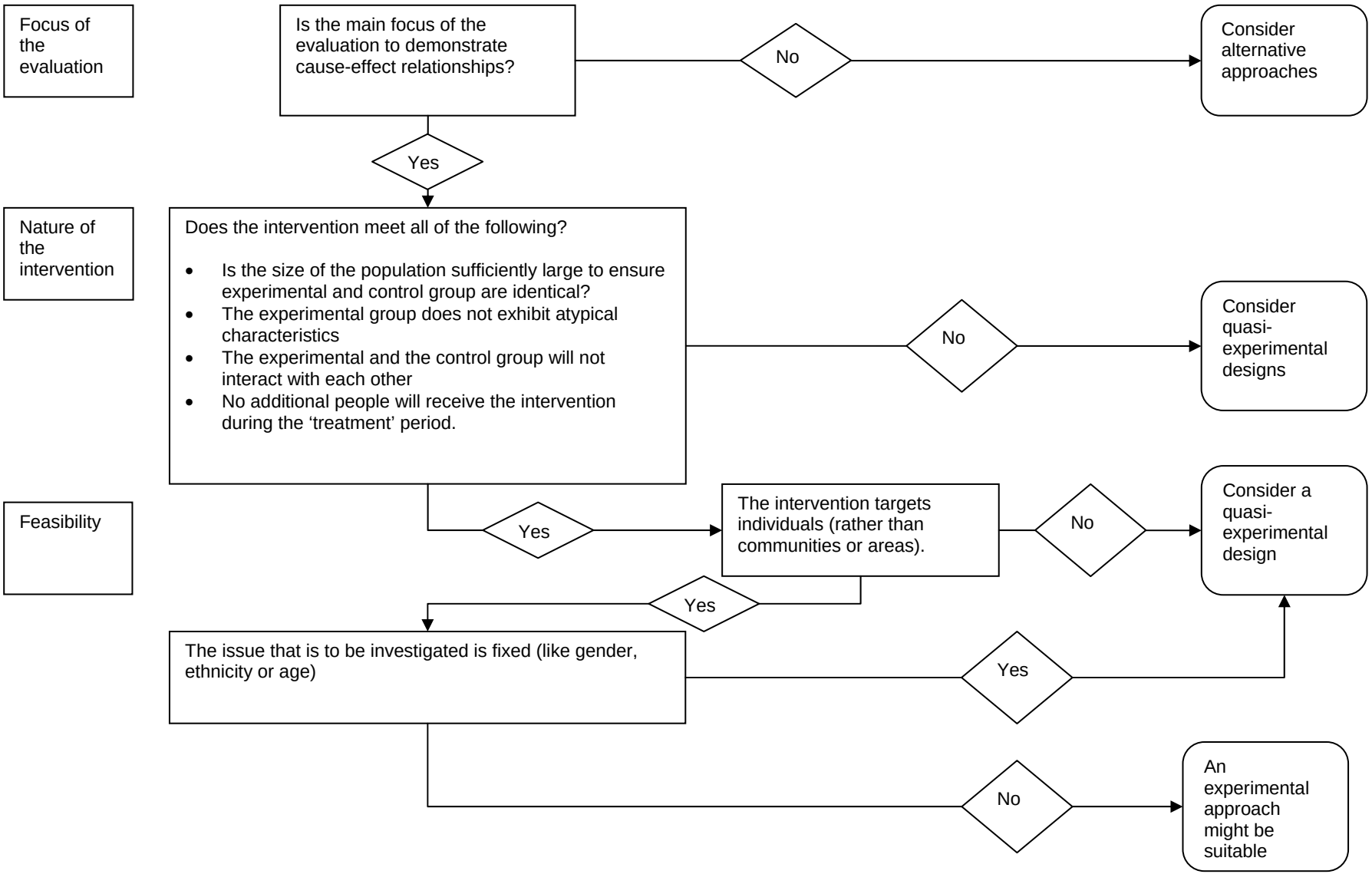
into the study which would affect the validity of the study results. As above, a quasi-experimental design combining a non-equivalent group and time series design would offer the evaluator useful data to control for the influence of these factors onto the outcome.

Is it possible to conclude that environmental factors are not responsible for the outcome? The validity of experimental designs relies on the ability of the evaluator to control environmental factors that may have an influence on the outcomes of an intervention. For this reason, it will often be more difficult to use experimental designs for interventions that target communities or areas (be this post codes, wards or even Local Authorities and administrative areas above this level). In these cases, a random allocation of these units into an experimental or control group is unlikely to offer sound results as it may lead to a comparison of units of analysis that are vastly different socially, demographically, politically, economically or geographically. In short: an experiment would not compare like with like so conclusions could not be drawn confidently that it was the intervention, and not other factors, that are responsible for any change observed. In this case, therefore, a quasi-experimental design would be more appropriate. This allows the evaluator to purposefully select study areas to be as similar as possible in terms of their key attributes and therefore to rule out alternative explanations for the observed changes. The question of whether or not an intervention has made a difference can therefore be answered with greater confidence.

Does the issue you want to investigate allow a full experiment?

Sometimes a full experiment is not possible because the issue that is of interest cannot be subjected to an experiment (manipulated). For instance, if we wanted to investigate whether people's gender influences their travel behaviour, an experimental design would not be possible: you cannot randomly assign people to be either male or female (the same is of course the case with other demographic factors, including age, ethnicity or intelligence.) In instances where it is not possible to assign individuals to the issue you want to investigate, a quasi-experimental design would be a more appropriate design choice. Experimental and comparator group would be as similar as possible with gender as the only discernable difference between them. The results of such as study would then allow the evaluator to draw conclusions on the impact of gender on travel behaviour. This example is also a good illustration of the difference in results produced by experimental and quasi-experimental research. Because quasi-experimental research does not draw on randomisation and statistics to eliminate alternative explanations, the evaluator would not be able to say that gender determines travel behaviour. They would merely be able to say that gender is an indicator for people's travel behaviour.

Figure 8: Flowchart for deciding on experimental or quasi-experimental designs



6.2. Deciding which theory-based evaluation approach is appropriate

If a theory-based approach has been identified as appropriate, a choice needs to be made between a Theory of Change approach and a Realist Evaluation approach. Unlike in the case of an experimental design, where this choice is effectively one of methodological rigour and ethical viability, the choice between Theory of Change and Realist Evaluation should be made in relation to the primary focus of the evaluation. If a theory-based approach is considered the most suitable approach but neither Theory of Change nor Realist Evaluation are practically feasible (for instance, because of financial or time resources) then an alternative option is an extended intervention logic evaluation (see section 6.3 below).

6.2.1 Is a Theory of Change approach most suitable?

Theory of Change is a systematic and cumulative study of the links between activities, outcomes, and context of an intervention. It involves the specification of an explicit theory of how and why an intervention might cause an effect which is used to guide the evaluation. It does this by investigating the causal relationships between context-input-output-outcomes-impact in order to understand the combination of factors that has led to the intended or unintended outcomes and impacts. Theory of Change therefore tests, and normally develops the implementation theory of an intervention and allows this to be modified or refined through the evaluation process. A range of research methods, often both quantitative and qualitative, can be used in order to gather data that contribute to this task. The evaluation often leads to a map showing which factors at which levels have combined to produce the observed outcomes. This will build on and develop the intervention logic developed in Step 3 (see Figure 3) and therefore will include all the components displayed in this map.

The Theory of Change approach is particularly useful for evaluations that are interested in testing whether and how a planned intervention will deliver change. Theory of Change can also be used for the evaluation of interventions where little is yet known about the impacts it might have, perhaps because the intervention is highly innovative or very complex (e.g. the impact of a package of measures on cycling behaviour and ultimately health impacts). The approach requires the involvement of those stakeholders responsible for designing and implementing the intervention from the start of the evaluation planning and the theory-building process. This allows for a pooling of knowledge and develops a more complete explanation of how the intervention will deliver the intended impacts (the causal pathways). The aim is to seek consensus among key stakeholders on causal pathways to help articulate how the intervention will deliver the intended impacts, to identify any alternative impacts it might have or consideration of alternative explanations for the observed outcome. It is also through this process of including stakeholders and building consensus about how outcomes and impacts have been achieved that the Theory of Change approach is able to attribute the observed change to the intervention: if the evaluation data confirm the implementation theory, as agreed

upon by the key stakeholders, this is evidence for a causal relationship between intervention and outcomes and impacts[30].

A Theory of Change approach is particularly suitable when the evaluation:

- Seeks to test the **implementation theory** behind an intervention, i.e. the connections between the inputs, activities, outputs, outcomes and impacts that are (explicitly or implicitly) assumed prior to implementation. For instance, the success of an intervention to promote cycling might be based on the number and kinds of organisations that need to be involved, the level of investment required, the number and type of engineering and education measures that needs to be implemented, the number of people and locations that need to be involved in order to achieve the anticipated outcome and impact targets. Testing the implementation theory through the impact evaluation would mean collecting data on each element (as set out in the map in Step 3). The achievement of observed outcomes and impacts is then explained with (and attributed to) the extent to which inputs / activities and outputs were achieved as planned.
- Theory of Change should also be used if the intervention has multiple and/or diverse stakeholders, as the evaluation seeks to establish a consensus over the implementation theory and it is important for stakeholders to buy into the theory.
- Investigates a highly **complex intervention** that addresses multiple issues, has a broad focus (for instance transport and health) and / or consists of different components (e.g. the different schemes forming a package or a programme) or is implemented in different locations.
- Makes it necessary to consider how **contextual factors** (e.g. people, organisations or socio-economic circumstances) influence the design and implementation of an intervention and what this means for the outcomes and impacts achieved. There should therefore be an interest in the evaluation generating new knowledge that can be used to inform future interventions.
- Examines an intervention implemented over a **long timescale**. This method would factor in any changes (within the intervention or wider social / economic context) into the evaluation as explanation for why the observed impact occurred.
- Is interested in identifying both **anticipated and unintended outcomes and impacts** of an intervention and how they have been achieved (including how interactions between outputs and outcomes are responsible for change). Theory of Change can therefore be used to produce **new theories**, using the final data, about how and why an intervention produces change.

Box 1 below presents an overview of the process for designing a Theory of Change impact evaluation.

[30] See also: Blamey, A and Mackenzie, M (2007) "Theories of Change and Realistic Evalaution. Pease in a Pod or Apples and Oranges?" *Evaluation*, Vol. 13 No 4, pp. 439-455

Box 1: Designing a Theory of Change evaluation

Getting started with designing an evaluation using a Theory of Change approach

1. When planning a Theory of Change evaluation, it is best to plan for the evaluation at the intervention design stage or before implementation commences. This will enable the evaluator to start capturing hypotheses on the Theory of Change in real time and therefore ensure their accuracy (as opposed to reconstructing them retrospectively).

   Begin by making a list of the stakeholders who are involved in the planning and implementation process. Use opportunities of joint meetings to work collaboratively on developing the intervention theory (see Step 3), or if this is not practical, speak to individuals separately. It is important to note any areas where perceptions of the intervention theory vary between the stakeholders and try to achieve consensus. This could, for instance, be done through a workshop that brings all stakeholders together for the purpose of exploring the nature of these variations and try to resolve them.

   It is recommended to develop your intervention theory "back to front", starting with the expected impacts and then working backwards to the outcomes required to achieve these impacts, the outputs required to achieve the outcomes, the types of activities needed to achieve these outputs and, finally, the required resources for the intervention. Your final intervention theory should not only be acceptable to stakeholders (see above), but also be realistic (i.e. timescales and financial resources should be commensurate with the expected impacts). It should also be testable: this means the envisaged outcomes must be specific enough to allow the evaluation to collect data on them[31].

2. Use your intervention theory / Theory of Change map to decide, together with the stakeholders, what your evaluation questions need to be so that useful evidence is generated from the research.

3. A key stage now is to examine with stakeholders the availability of existing data which can meet analytical requirements (taking into consideration the quality, relevance, coverage and accessibility of the evidence). If the existing evidence base does not fully meet the data requirements then consideration will need to be given to appropriate approaches to obtaining relevant data. This might include undertaking social research with the target population.

4. Identifying the appropriate methodology for informing gaps in the existing evidence base will be based on the evaluation questions, the requirements of the users of the evaluation and the timing of the evaluation and the intervention. Evaluators should consider the relative merit and appropriateness of qualitative methods (e.g. interviews, case studies, focus groups) and quantitative methods (e.g. surveys, analysis of monitoring data) to answer your evaluation questions particularly focusing on the methods required to collect data on specific aspects within the intervention logic.

5. Thinking through the evaluation process: the evaluator should seek to design the research activities so that they do not miss valuable and important information. When setting up the timeframe for the evaluation the following questions should be considered:

   - *What is a realistic timeframe for the evaluation to capture meaningful impact data?* If the evaluation happens too quickly after an intervention is introduced or finishes, then it is unlikely that it will generate data on overall impacts. Rather, the evaluation is more likely to generate data on outcomes (i.e. short to medium- term impacts).
   - *Are there any critical points at which particular data will need to be collected in order to capture important information?* For instance, stakeholder input needs to be gained as early as possible,

[31] See also: Blamey, A and Mackenzie, M (2007) "Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?" *Evaluation*, Vol. 13 No 4, p. 443

monitoring arrangements will need to be put in place before the intervention is implemented, and surveys should be undertaken at times when they would deliver the most meaningful input, etc.

- *At what intervals should your knowledge about stakeholders' thinking on the intervention be updated?* For instance, there might be crucial stages in the intervention at which their perspective would be particularly valuable. For interventions implemented over a long timeframe, it might be useful to speak to key stakeholders regularly (i.e. annually) in order to capture any changes in thinking. Additionally, opportunities should be taken to gauge stakeholders' views if unforeseen events occur that are likely to impact on the evaluation as this evidence can be used to refine the Theory of Change maps.
- *When do the evaluation findings need to be ready to be valuable to end users?* Having a clear focus on the forthcoming policymaking/ programme delivery timeframe will ensure the evaluation findings can make a timely and relevant contribution to the evidence base.

6. Plan how the data will be analysed to test the initial assumptions, the results verified and establish the feedback mechanisms with stakeholders. For example, mechanisms might be established for involving stakeholders in the interpretation of the data. You may also want to engage them in your updating of the original Theory of Change map with the new data and knowledge that the evaluation has generated to ensure their buy-in into the results.

7. An important final step will be establishing from the outset the opportunities for learning from the evaluation and considering how this will be fed back into the policy making and delivery cycle will ensure the evaluation findings will provide value.

6.2.2 Is a Realist Evaluation approach most appropriate?

Whilst Theory of Change tests implementation theory, Realist Evaluation seeks to identify those – often psychological – triggers that change human behaviour as a result of an intervention, taking into account the context within which the intervention sits. Realist Evaluation typically asks: **"what works, for whom, under what circumstances?"** It begins by developing a set of hypotheses (or theories) on those factors or processes that explain why an intervention has had a particular result (called a mechanism), and what effect the context of an intervention has on these mechanisms. A mechanism can be defined as capturing "people's reasoning and their choices. They describe how people react when faced with a policy measure"[32] .

Evaluation activities and methods are then geared around investigating which combination of mechanism and context factors is responsible for producing the observed outcome of an intervention. Like experimental designs, realist evaluation seeks scientific proof of cause and effect relationships. Unlike experimental designs, however, it pays attention to the external conditions (i.e. the context) in which an intervention is implemented and how these affect the outcomes and impacts achieved.

A Realist approach is particularly suitable in the following circumstances when:

- The interest is in understanding how an intervention in a particular context releases triggers that affect change. Thinking back to your intervention logic, and the example in Figure 3, Realist Evaluation, in essence, focuses on investigating the spaces between the different stages of the implementation of an intervention

[32] Befani, B et al (2007) "Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application", Evaluation, Vol. 13 No 2, p. 178

(these explain why and how change happens). Implementation issues (e.g. the activities carried out) are likely to be part of the range of possible context factors that form part of the testable theories but are not the main focus of the investigation.

- The **context** of the intervention is likely to be a key factor determining outcomes and impacts. For instance, the success of CCTV cameras in car parks for deterring crime might depend on where the camera is installed (e.g. whether there are 'blind spots' not covered), where the car park is located (in a remote or busy spot) or how much security staff is deployed in addition to the cameras. These factors will vary between car parks, and so will the outcomes and impacts of CCTV. Therefore, to be confident about attributing the observed change to an intervention, the evaluation needs to generate knowledge on "what works for whom, where and under what circumstances". By taking into consideration the importance of context, the evaluation then produces a tested theory on *why* and *under what conditions* an intervention works. Establishing these realist theories may require the exploration of a number of different sources such as academic literature, grey literature (including relevant evaluation reports) and individual experts (evaluators, academics, stakeholders). The greater diversity of sources is likely to improve the evaluator's understanding of the range of mechanisms and context factors that may interact in order to produce an impact which is then investigated.
- The intervention is implemented at a small geographical scale or is not overly complex. Realist Evaluation is about testing different theories about how an intervention triggers change in a particular situation (context). The more complex an intervention, the larger the number of possible theories is likely to be and hence the more expansive the evaluation (and related data collection) will need to be. It may therefore be more practical to apply the Realist approach for evaluations of interventions that are implemented at a smaller geographic scale and are not too complex (e.g. in terms of number of locations where it is implemented or range of expected outcomes).
- The intervention is being implemented over a short, medium or longer-term **timescale**. Any changes experienced during the implementation period would be factored into the initial theories (if regarded as potentially relevant for the outcome).
- The evaluation is looking to explain the expected and unexpected outcomes and impacts of an intervention by investigating what mechanism (or mechanisms) has produced them in a given context. The evaluation would be looking to identify the (range of) triggers that are set off by an intervention and are responsible for producing change in a particular situation. The results of the evaluation would therefore be universal / generalisable[33]. Findings would be applicable elsewhere as long as context and intervention are comparable. Understanding what produces change in a given context will help policy makers decide how to tailor interventions to different situations so as to achieve the intended results.

[33] See also: Blamey, A and Mackenzie, M (2007) "Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?" *Evaluation*, Vol. 13 No 4, p.p. 439-455

Case study 6: A Realist Evaluation to assess the impact of CCTV in car parks to reduce crime

Pawson and Tilley argue that there is nothing about CCTV in car parks which intrinsically inhibits car crime. Cameras must work by instigating a chain of reasoning and reaction. Thus, the mechanisms through which CCTV may enter the potential criminal's mind and the contexts needed if these powers are to be realised. The following context and mechanism factors might, among others, explain the outcome of reduced car crime following CCTV installation[34].

Outcome (O): Reduced car crime in a car park following CCTV installation.

Context factor 1 (C1): The car park is located in a CCTV 'blind spot' (i.e. all or parts of it are not covered by the installed CCTV)
Context factor 2 (C2): The car park is located in an area where possible alternative targets for car crime are available.

Mechanism 1 (M1): Potential offenders are deterred by CCTV because they do not wish to be captured, apprehended and convicted.
Mechanism 2 (M2): CCTV notices may remind drivers that their cars are vulnerable and they may therefore take greater care to lock them and taken other security measures.
Mechanism 3 (M3): CCTV may allow deploying security staff or police towards areas where suspicious behaviour is occurring. They then act as a visible barrier which may deter potential offenders and they may disable actual offenders.

These factors can be turned into the following context-mechanism-outcome configuration (CMOC) which the evaluation would then test:
CMOC 1: Though the car park is in a CCTV blind spot, notices mean drivers take greater security precautions and potential offenders are deterred because they do not want to be detected. Car crime goes down. [C1+M2+M1=O]
CMOC 2: The car park is located near possible alternative targets but displacement effects elsewhere are prevented because the installation of CCTV in the car park allows the deployment of security staff or police into areas where suspicious behaviour is occurring. They then act as a visible barrier which may deter potential offenders and disable actual offenders. Overall car crime is reduced with no displacement effects occurring. [C2+M3=O]

Source: Pawson, R and Tilley, N (1997) Realistic evaluation, pp 78-82 (modified and simplified)

If a Realist Evaluation design is chosen, the next step is to start thinking about the design process (see Box 2).

Box 2: Designing a Realist Evaluation

Getting started with designing an evaluation using the Realist approach

1. Make sure you think about the intervention in Realist terms by asking the following questions (you may find it useful to look at the CCTV in car parks example for further guidance) [35]:
    a. Mechanism: how may the intervention lead to an outcome in a given context?
    b. Context: what conditions are needed so that the intervention triggers the mechanism(s) and produces the envisaged outcomes?
    c. Outcomes: what change (or result) is produced by the mechanism(s) triggered by the intervention in a particular context?

    The more complex your intervention, the longer you can expect your list of possible mechanism and context factors to be. Draw on sources such as academic literature, grey literature, experts and others for this process.

2. Develop theories about how mechanism and context factors come together to produce outcomes. In Realist Evaluation these theories are called context-mechanism-outcome configurations (CMOC). They are the start

[34] Many more are imaginable and, in fact, listed by Pawson and Tilley (1997), but for explanatory purposes the list has been reduced and simplified.
[35] See also: Tilley, N (2000) Realistic evaluation: an overview, http://www.danskevalueringsselskab.dk/pdf/Nick%20Tilley.pdf

of the evaluation: it is these theories that your evaluation work will need to test. You can draw on a range of sources to develop your theories, including, for instance: transport research and relevant research from other fields, previous evaluations, appraisal and modelling work, those that have designed the intervention, experts and practitioners locally and nationally. Much of this information will have already fed into your intervention logic, and you can draw on this to develop your CMOC theories. The number of theories you will need to develop depends on the number of logically possible combinations of mechanism and context factors. The more complex your intervention the higher the number of theories is likely to be that you will need to test.

3. If the intervention is implemented in different locations (this is, for instance, often the case if you are looking to evaluate a programme), decide on the setting for your evaluation. Do you need to include all of the locations where your programme is implemented in your evaluation in order to test your theories, or is a selection of sites sufficient? As you have decided that context is an important factor to explain the outcome and / or impact of the intervention, you need to make sure that the locations you choose to include in your evaluation cover the range of conditions you have identified as important as part of your theory development in points 1.) and 2.) above. Only then will be you able to test these theories through your evaluation.

4. Starting from your set of theories, analyse your data needs. Think about what data are required to verify or falsify the different theories, what relevant data exist and you have access to and what additional data you need to collect. As your focus will be on understanding the importance of context, particular thought will need to be given to the level of detail your data will need to show, and whether this is available. Analyse where data is weakest or where there is contradictory information (the context, the mechanisms for change or the outcomes?). You can then use this insight to focus your evaluation design, both in terms of questions and in terms of methods to be used.

5. Decide what types of data will need to be generated considering your evaluation questions and your review of evidence undertaken to set out the hypotheses:

    - Decide which methods you will use to capture change, for instance monitoring data or other outcome measures, observations, interviews, surveys etc.
    - Decide what practical issues need to be considered for carrying out the evaluation (e.g. accessing individuals or data).

6. Decide whether you want to include stakeholders and other individuals in the interpretation of the data to generate conclusions on what works in what context. Such an activity is particularly advisable where the evaluation setting chosen is very small scale so that the input of local people in the interpretation of data adds valuable local knowledge.

7. Consider whether you want to achieve generalisable findings and what this means for the choice of methods incorporated in your Realist Evaluation design. If the intention for the evaluation is to produce lessons from multiple sites, this can be challenging using the Realist approach as context is considered to be the defining factor for impact. The ability to make generalisations is likely to depend on the use of additional specialist methods (for example qualitative comparative analysis (QCA)), so consider their inclusion in the study design[36]. This would allow examining patterns across sites (i.e. factors that are present and responsible for change) which would increase confidence that these are universal success factors.

8. Consider how the results of the evaluation will be used, and therefore how they will need to be presented in the final report. This may include considering the need for triangulation of data. The timing of the final report will, once again, be important to ensure that evaluation findings have an opportunity to feed into policy making.

[36] Befani, B et al (2007) "Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application", Evaluation, Vol. 13 No 2, p. 178

Case study 7: Evaluating an information campaign to encourage people onto public transport to reduce CO2 emissions (hypothetical example)[37]

In this hypothetical example, a Local Authority has launched an information campaign which aims to encourage people to use more sustainable modes of travel (public transport, walking and cycling etc) in order to reduce CO2 emissions from car use and make a contribution to averting climate change. A Realist design was chosen for the evaluation and started with developing a number of possible CMOC theories about how and why about the information campaign would influence car use. At the end of the campaign an increase in public transport use was recorded, but the degree of change seemed to differ between wards. The evaluation of the campaign allowed the investigation of why or why not modal shift had happened as a result of the campaign in order to learn for future interventions. The list below was developed before the campaign commenced by those leading on the evaluation in conjunction with the designers of the intervention and drawing on academic and other resources. It includes just a few of the possible factors as this is a highly complex area and a range of other factors would need to be considered in practice.

Outcome: A very small increase in public transport usage has occurred following the information campaign, but the degree of change was not large enough to rule out other factors (e.g. seasonal fluctuations).

Context factor 1: Accessibility of the public transport network differs between both households (e.g. the distance to/from their home to bus stops or train stations) and individuals (e.g. ability/willingness to walk to/from their home and the bus stop/train station).
Context factor 2: Particular individuals face varying barriers to using public transport such as their awareness/ knowledge of the public transport available to them; the actual cost of travelling by public transport rather than by car; the suitability of public transport for the journey they wish to make (e.g. whether bus / train routes go near to where they wish to travel; availability of public transport at the time of day or day of the week they wish to travel; whether they are travelling with children or luggage); and/or the extent to which car use is an ingrained habit for them.
Context factor 3: Individuals or groups can hold a variety of attitudes which may act as barriers to public transport use and/or cycling (e.g. the perception of non-car modes of transport being less convenient and/or more costly for all or some journeys; the fear of crime/anti-social behaviour on public transport; the fear of traffic accidents when cycling).
Context factor 4: The extent to which climate change can be used as a factor to motivate individuals to change their behaviour varies depending on individuals' values and value frames. Individuals who are interested in/ motivated by ethics or societal/environmental benefit are more likely to respond; individuals who are motivated by other kinds of needs (e.g. those relating to social status/aspirations; or those relating to more basic social/ economic survival) are less likely to respond to messages about climate change.

Mechanism 1: The information campaign increases people's awareness of car travel as a cause of climate change and motivates some individuals to change their travel behaviour because they are able and willing to walk/cycle/use public transport more and because they are motivated by messages relating to climate change. .
Mechanism 2: Some people who were motivated by the messages relating to climate change were unable to actually change their behaviour because various barriers (e.g. accessibility of the public transport network; higher cost of travelling by public transport) prevented them from being able to change their behaviour.
Mechanism 3: Some people were not motivated to change their behaviour in order to reduce CO2 emissions, either because of attitudinal barriers relating to transport or because they were not motivated by messages relating to climate change.

The following CMOC theories can be put together from this list that might then guide the evaluation:
CMOC1: Individuals with easy access to public transport which serves their journey needs; positive attitudes towards public transport; and who are motivated by messages relating to climate change; start using public transport more as a result of the information campaign.
CMOC2: Even though the information campaign increased some people's willingness to change their behaviour due to them being motivated by messages related to climate change, some individuals who were motivated by the messages relating to climate change were unable to change their behaviour due to external barriers (e.g. poor accessibility of the public transport network; higher cost of travelling by public transport).

[37] The information included in context and mechanism factors has been mostly taken from: 1) People, Science and Policy Ltd (2009) Exploring public attitudes to climate change and travel choices: deliberative research. Final report for Department for Transport; 2) IPPR (2009) Consumer Power: How the public thinks lower-carbon behaviour could be made mainstream; and 3) Anable et al (2006) An evidence base review of public attitudes to climate change and transport behaviour.

CMOC3: Some individuals who were motivated by the messages relating to climate change were still unwilling to change their actual behaviour due to negative attitudes towards public transport and/or cycling (e.g. fears about personal safety/crime/anti-social behaviour on buses; fears of traffic accidents when cycling).
CMOC4: Some individuals were not motivated by the messages relating to climate change because their needs are related to either social status/aspirations or more basic social/economic survival factors, rather than concerns about ethics/ethical living.

6.3. Choosing a combined approach

In some circumstances it can be beneficial to combine different elements of the three broad evaluation approaches we have outlined in Step 5. Indeed, the approaches are not mutually exclusive and evaluation designs can sometimes benefit from incorporating elements of other approaches to generate the best possible data. What combination of elements is most useful depends on the intervention and the evaluation questions. Below we briefly discuss two options for a combined approach, though this should not be taken as excluding other possible variations.

6.3.1 Combining theory-based approaches with experimental designs

A combined approach could bring together theory-based approaches with experimental designs. This is possible because experimental and quasi-experimental approaches are not only evaluation approaches but also research methods, so they can be used in conjunction with other evaluation designs to expand the range of data that is collected and therefore the robustness of the evidence generated. If this combined approach is chosen, the intervention (or parts of it) must satisfy the conditions for both experimental and theory-based approaches.

A combined approach may be useful where the conditions are appropriate for an experiment but there are questions to be answered about the underlying intervention logic. Understanding the anticipated sequence of steps between input and outcome (implementation/process of the intervention) or its context can assist the interpretation of findings from the experiment and provide a fuller picture about *why* the intervention was successful. This knowledge will assist policy makers with making future decisions about interventions.

Additionally, experimental methods can complement evaluation designs that are underpinned by theory-based approaches where they can offer a ‘scientific’ testing of hypotheses.

Case study 8: The evaluation of travel plans – a hypothetical example for a combined approach

An evaluation of travel plans might be interested in a range of questions and touch, for instance, on scheme design, what works where and when and the contributions the plans make to DfT objectives. Overall, the evaluation is therefore likely to have a knowledge purpose, and be interested in how the intervention as implemented generates outcomes and impacts. Under these conditions, a Theory of Change approach will be the most appropriate overall framework. The initial intervention logic would capture stakeholders' views on the conditions that need to be in place to achieve the expected outcomes and impacts. Evaluation activities would then explore delivery on the ground, including any variations (e.g. the extent to which individuals have personalised help to make decisions, or are just given general information, and the whether stakeholders are actively involved in the design) and what this has meant for success.

If there is, in addition, an accountability dimension to the evaluation, quasi-experimental methods could be added. This is possible because these schemes have a focus on the individual and the interventions are relatively straightforward. For example, one could compare schools or workplaces with and without travel plans using this evaluation approach. It is unlikely that randomisation of the recipients of the intervention could be achieved, as it would be difficult to target some individuals in a school, community or workplace, and not others. As such, a pure experimental approach is not viable.

The result of such an evaluation would, therefore, be both learning about how and why change happened and evidence about the degree of change produced against a baseline situation.

6.3.2 The extended intervention logic option

Where practical circumstances exclude the use Theory of Change using an extended intervention logic approach could be a 'second best' option to generate richer explanations about why and how change was produced.

The extended intervention logic option is a way of feeding in elements of a Theory of Change approach to supplement outcome studies. The key difference to a full Theory of Change evaluation is that no stakeholder consensus on the intervention theory is sought. Rather, their input into the evaluation is restricted to supplying information after introduction of the intervention and hence to add sophistication to the interpretation of data. The difference to a 'pure' outcome study is that this approach would provide some answers to the question of why change was produced (rather than just capturing that change happened). This approach therefore adds a knowledge component to what will mostly be an accountability evaluation.

In the extended intervention logic the evaluator would start by organising monitoring data, as well as data you have from programme or other relevant documents, into the format of the intervention logic (see Step 3). Stakeholder interviews would then be conducted in order to obtain their views on connections between outputs and outcomes and impacts and hence to obtain their views on why the measured change occurred and what this means for outcomes and impacts. Beyond this, new qualitative or quantitative data would only be collected if and where practical circumstances allow it and particular data gaps exist that need to be filled.

The combination of stakeholder interviews, monitoring data and drawing up the intervention logic means that the causal pathways as initially drawn up can be strengthened and, where necessary, refined or revised. The added learning generated from the evaluation would be useful to modify the intervention and support introduction elsewhere. However, confidence in conclusions will be lower as the research effort will be limited and therefore causal relationships would not be tested as well as during a full Theory of Change evaluation. Policy makers may therefore be less likely to act on the results from this evaluation.

Case study 9: Evaluating the Dutch Ecodrive initiative – an extended intervention logic

Ecodrive aimed to overcome the lack of consumer awareness on the benefits of energy efficient driving through activities including driver training and publicity campaigns. The purpose of the evaluation was to assess how Ecodrive activities led to reduced CO2 emissions (i.e. impact). Whilst the programme was well-monitored, a straightforward connection between the intervention and the outcome and impact observed was assumed. There was also a need for greater clarity about what constituted success. Against this background, the evaluation team began by identifying, from published documents, the logic between the intervention as planned and the consequent outcomes and impacts as well as the indicators needed to measure them. This programme logic was then refined through stakeholder interviews (representatives from the car industry, Ministry of Transport and Environment and Agency for Energy and the Environment). From this came a better understanding of appropriate indicators and use of monitoring data which then guided the analysis of the available data. The evaluation showed that the communication campaign within Ecodrive was the most effective mechanism, acknowledging that high costs for setting up the campaign were compensated by large reach. This type of result could only have been derived from the systematic mapping of the programme's policy assumptions. Overall, therefore, the approach proved to be a coherent way of focusing a high volume of monitoring data into logical policy assumptions with measurable indicators. This was reinforced by involving stakeholders during the policy mapping stages to refine the theory and add context to the evaluation.

Source: Van den Hoed, R., Harmelink, M. and Joosen, S. (2006) *The evaluation of the Dutch Ecodrive Programme*. Available from the AID-EE website: http://www.aid-ee.org/documents/000015Ecodriving-Netherlands.pdf

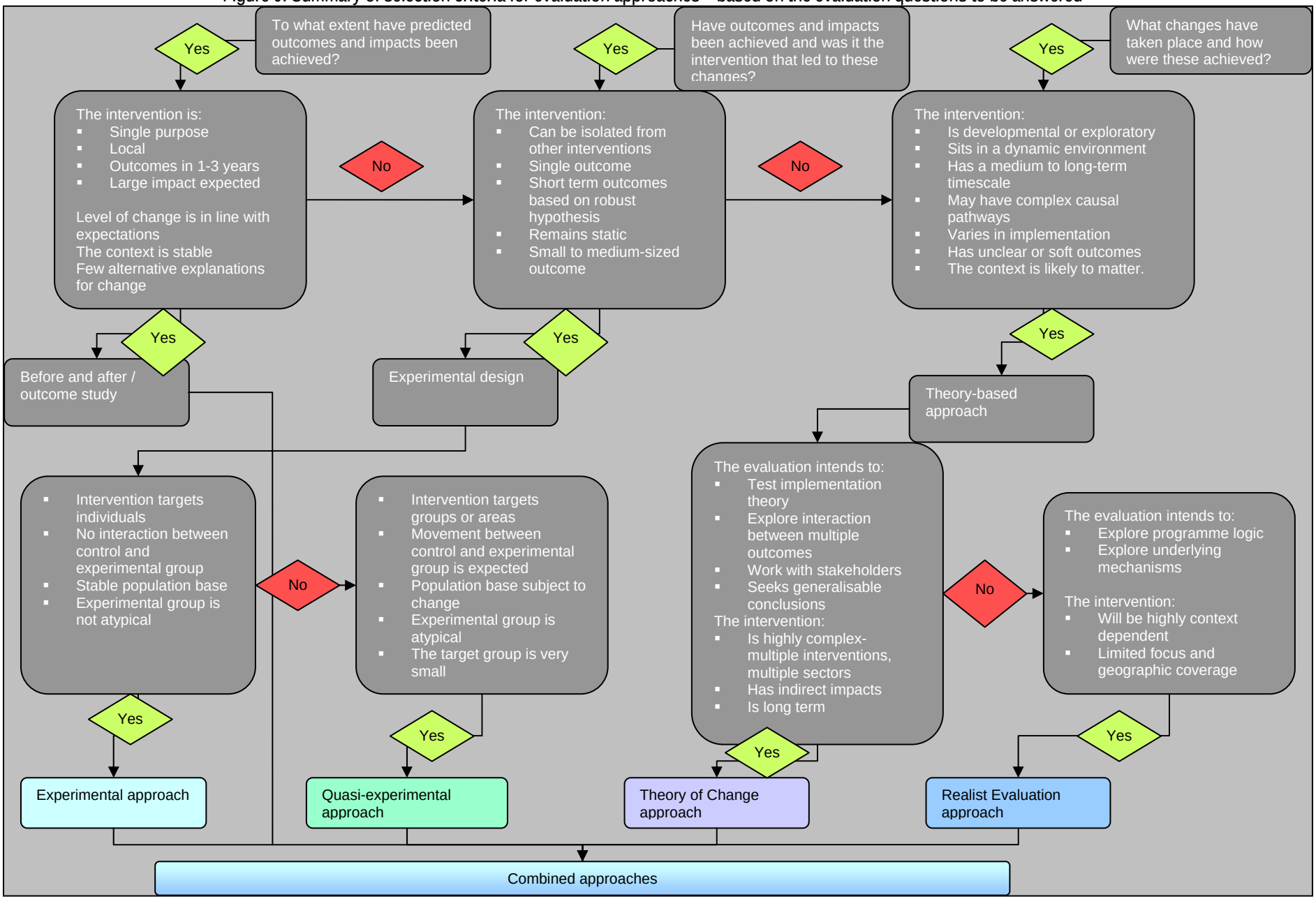
SUMMING UP: SIX ISSUES TO CONSIDER FOR BETTER ATTRIBUTION

These guidelines have set out in six steps how transport impact evaluations can be designed to produce better evidence on attribution. The key messages are:

- **Be clear about who will use the evaluation**, what evidence these users will need and what resources (both financial and in terms of personnel) you have available to undertake the evaluation work.
- **Understand the nature of the intervention** you are looking to evaluate: how 'complex' is it, what is its purpose and what broad evaluation questions will your evaluation need to answer?
- **Map the intervention logic** to help you understand what you and other stakeholders already know about the context, inputs, outputs, outcomes and impacts of the intervention you are looking to evaluate. This will help you understand where your greatest evidence needs are and therefore where you should focus your evaluation.
- **Define the evaluation purpose and frame the evaluation questions** drawing on your knowledge from the intervention logic.
- **Decide on the best overall evaluation approach**: an outcome study, an experimental approach or a theory-based approach.
- Finally, **refine this choice** by choosing between experimental and quasi-experimental designs, Theory of Change or Realist Evaluation or a combined approach.

At the end of this six-step process you will be able to start designing your evaluation, think about appropriate methods and begin a process which will give you sound evidence on the impact of your intervention which will be of use to your key stakeholder constituency. Figure 9 provides a quick check list of the key intervention features to take into account in deciding on each type of evaluation approach. If one or more of the characteristics in the main rectangular boxes are not present, then it is likely that the evaluation approach in the green diamond below that box will be inappropriate, and another approach should be considered.

Figure 9: Summary of selection criteria for evaluation approaches – based on the evaluation questions to be answered



GLOSSARY

Term	Definition
Activities	Specific elements of an intervention – may also be seen as an aspect of 'outputs' see below.
Appraisal	The process of defining objectives, examining options and weighing up the costs and benefits, risks and uncertainties of those options before an investment decision is made. See ex ante evaluation See DfT's WebTAG appraisal guidance www.dft.gov.uk/webtag/
Attribution	The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention. *Note:* Attribution refers to that which is to be credited for the observed changes or results achieved. It represents the extent to which observed effects can be attributed to a specific intervention or to the performance of one or more partner taking account of other interventions, (anticipated or unanticipated) confounding factors, or external shocks. www.worldbank.org/oed/ecd/docs/annex_e.pdf
Attrition	The number of participants opting to no longer take part in an experiment or who cannot be recontacted.
Audit	In the form of value -for-money (or performance) audit as opposed to financial audit, is the provision of independent analysis and assurance on the way in which public money has been spent. It is strongly concerned with questions of efficiency and good management and makes recommendations that lead directly to improvements in public services. There are parallels with evaluation but audit is generally less wide-ranging in scope and is not as focused on understanding the longer-term consequences of the intervention. www.nao.org.uk/about_us/what_we_do/financial_audit.aspx www.evaluation.org.uk/resources/glossary.aspx
Counterfactual	An estimate of what would have happened if the intervention (e.g. the project, programme policy or financial assistance) had not taken place. The status quo or other baseline option used in the original appraisal should normally inform the counterfactual. However, viewing events from a post hoc position, evaluators may judge that the counterfactual would actually have been quite different from what was envisaged at the time of the appraisal, due to, for example, alternative states of the world and/or alternative management decisions. In such circumstances it may be helpful to consider other counterfactuals in addition to the original baseline option. Green book: www.hm-treasury.gov.uk/data_greenbook_guidance.htm
DfT	Department for Transport
Ex ante evaluation	An evaluation conducted before the implementation of an intervention. See appraisal www.evaluation.org.uk/resources/glossary.aspx
Ex post evaluation	An evaluation conducted either on or after completion of an intervention. www.evaluation.org.uk/resources/glossary.aspx

Term	Definition
External validity	The confidence one can have about whether or not one's conclusions about the intervention can be generalised to fit circumstances, times, people, and so on, other than those of the intervention itself. A threat to external validity is an objection that the evaluation design does not allow causal inference about the intervention to be generalised to different times, places or subjects to those examined in the evaluation. www.evaluation.org.uk/resources/glossary.aspx
Experimental methods	A theoretical way of deriving the counterfactual situation, and hence the net impact of an intervention. It involves comparing two groups which are identical in all respects except one: exposure to the intervention. Differences between the groups which have been exposed (the programme group) and the group which has not (the control group) are then attributable to the intervention. Quasi-experimental designs are a class of causal evaluation designs which take a more practical approach than is the case with true experimental designs. Control groups can still be used, but these have to be assigned through some non-random process. Alternatively, one can examine beneficiaries before and after exposure to the intervention. www.evaluation.org.uk/resources/glossary.aspx
Impact	Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. http://www.worldbank.org/oed/ecd/docs/annex_e.pdf
Impact evaluation	The evaluation of the effectiveness of economic, social and environmental impacts of an intervention.
Internal validity	The confidence one can have in one's conclusions about what the intervention actually did accomplish. A threat to internal validity is an objection that the evaluation design allows the causal link between the intervention and the observed effects to remain uncertain. It may be thought of as a question of the following nature: could not something else besides the intervention account for the difference between the situation after the intervention and the counterfactual? www.evaluation.org.uk/resources/glossary.aspx
Intervention	Collective noun used to cover transport policies, programmes, schemes, projects and packages.
Intervention logic	The conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes. The examination of the programme's intervention logic will be of central importance in most evaluations. The evaluator needs to ask how the programme achieves its specific objectives, and how do the specific objectives contribute to the attainment of the general objectives? The terms "theory of action", "programme logic" and "programme theory" are sometimes used to mean more or less the same thing. www.evaluation.org.uk/resources/glossary.aspx
Non-equivalent group design (quasi-experimental method)	The non-equivalent comparison group (NECG) design involves the evaluator selecting a group of units similar to those receiving the new policy or programme that is being tested. Such a group is called a comparison group (similar to a control group in a social experiment) and acts as a counterfactual. www.nationalschool.gov.uk/policyhub/evaluating_policy/magenta_book/chapter7.asp

Term	Definition
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs. www.worldbank.org/oed/ecd/docs/annex_e.pdf
Outputs	The goods and services produced by an intervention (e.g. training courses for the long-term unemployed). See also intervention, intervention logic, operational objectives. These may also be referred to as activities. www.evaluation.org.uk/resources/glossary.aspx
Qualitative comparative analysis	The name given by Charles Ragin (The Comparative Method, 1987) to his proposed technique for solving the problems that are caused for comparative macrosociologists by the fact that they must often make causal inferences on the basis of only a small number of cases. The technique is based on the binary logic of Boolean algebra, and attempts to maximize the number of comparisons that can be made across the cases under investigation, in terms of the presence or absence of characteristics (variables) of analytical interest. www.encyclopedia.com/doc/1O88-qualitativecomparatvnlyss.html
Realist Evaluation	Realist Evaluation assumes that most interventions have varying impacts under different sets of circumstances. Therefore, the context in which a programme is implemented is an important determinant for the outcomes. Realist Evaluation addresses the psychological and motivational responses that lead to behaviour change and starts from the premise that causal outcomes follow from mechanisms acting in context. Put differently: Outcomes are explained by the action of particular mechanisms in particular contexts. Realistic Evaluation is therefore concerned with "understanding causal mechanisms and the conditions under which they are activated to produce specific outcomes." (Tilley 2000, 5).
Theory of Change	Theory of Change involves a systematic and cumulative study of the links between activities, outcomes and context of an initiative. It involves the specification of an explicit theory of how and why a programme or project might cause or have caused an effect and the use of this theory to guide the evaluation. The focus of the Theory of Change approach is therefore on causal pathways.
Time series design (quasi-experimental method)	An example of a quasi-experimental design. It involves obtaining several measurements over time both before and after exposure to a programme in order to create a time series of observations. www.evaluation.org.uk/resources/glossary.aspx

KEY TEXTS FOR FURTHER READING

Experimental and quasi-experimental approaches

Campbell, D. T. and Stanley, J. C. (1963) *Experimental and Quasi-experimental designs for research*, Chicago: Rand McNally

Glasziou, P., Chalmers, I., Rawlins, M. and McCullock, P. (2007) “When are randomized trials unnecessary? Picking signal from noise”, *MMJ*;334,349-351.

Cullen, J. and Hills, D. (1996), *The Role of RCTs in Assessing Services Effectiveness: a Critical Review*, EDRU Occasional Paper, London: The Tavistock Institute

Theory-based approaches

Pawson, R. and Tilley, N. (1997) *Realist Evaluation*, London: Sage

Connell J. P. and Kubisch A. C. (1998) *Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects, and Problems*, This paper is for sale from the Aspen institute which also has a number of on line ‘Theory of Change’ resources: www.aspeninstitute.org.

See also linked site: www.theoryofchange.org

Davidson, E. J. (2000), "Ascertaining Causality in Theory Based Evaluation" *New Directions in Evaluation*", Fall, vol. 87

Discussions of several approaches

Cook, T. D. (2000) “The false choice between theory-based evaluation and experimentation”, *New Directions in Evaluation*, Vol. 87 (Fall)

Blamey, A. and Mackenzie, M. (2007) “Theories of Change and Realistic Evaluation” *Evaluation*, Vol. 13 No. 4, pp. 439-455

GSR Magenta Book, www.nationalschool.gov.uk/policyhub/magenta_book/