

# Understanding impact evaluation

What Works Growth (WWG) aims to make local growth policy more cost-effective, through better use of evidence and evaluation. ‘Evidence-based policy’ calls for policy decisions based on systematically accumulated objective knowledge on what works and what works best. WWG supports impact evaluation as a component of evidence-based policy making because it can tell us if an intervention produces its intended results.

This page explains why evaluation, and particularly impact evaluation, is important for policy design. It aims to help policymakers understand what impact evaluation is, why it is important and when it is appropriate.

## Evaluation and types of evaluation

### What is evaluation and why is it important?

Ex-ante analysis tries to predict the effects of an intervention *before* it is implemented. Evaluation is ex-post analysis that tries to understand the effects of a policy *after* it is implemented. Evaluation can tell us what has, and has not, worked in the past.

### Types of evaluation

Different types of evaluation answer different questions. While WWG focuses on impact evaluation, which asks whether a policy affects its intended outcomes (and is described below) other types of evaluation can help with different questions:

**Process evaluation** looks at the way an intervention was implemented. Was it implemented as intended? What challenges were encountered? How do the administrators and participants feel it went? Did the context influence implementation? - Methods vary but usually involve interviews, surveys, and collection of monitoring data on outputs (i.e. things the policy did). Good monitoring data can give some idea of what might be changing due to the policy.

**Theory-based evaluation** goes into the “black box” of an intervention to consider the logic of how and why it affects specific outputs and outcomes (and perhaps doesn’t affect others). This usually starts by constructing a theory of change, and then trying to understand, using data and interviews, whether activities and outputs are changing in the way the theory of change would suggest. Theory-based evaluation can help think about mechanisms through which an intervention might work but cannot establish whether they occur in practice.

**Value for money (VfM) evaluation** focuses on estimating the return on investment from the resources spent on an intervention, to understand its cost-effectiveness. VfM evaluation can use different methods – the most well-known being cost-benefits-analysis – but they all seek to quantify the benefits of the intervention relative to costs. VfM evaluation does not determine whether an intervention has had an impact, but rather whether any impact was worth the time and money.

**Economic impact assessment (EIA)** examines effects on intervention specific outcomes and standard economic variables (e.g., Gross Domestic Product [GDP], Gross Value Added [GVA], expenditure, employment, and wages). EIA usually starts with a logic model that identifies the outputs, outcomes, and impacts that the intervention is expected to produce. The estimation of the benefits (direct, indirect, and induced), and net additional effect of a project, are based on assumptions on additionality, deadweight, and multiplier effects, especially relating to employment and expenditure. EIA cannot determine causal impact and attribute the changes to an intervention.

### Where can I learn more?

- Magenta book – [https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf)
- Green book – <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020>
- DLUHC appraisal guide - <https://www.gov.uk/government/publications/dluhc-appraisal-guide/dluhc-appraisal-guide>

## Impact evaluation

### What is impact evaluation and why is it important?

Impact evaluation examines whether a policy had an impact on specific outcomes. This is known as ‘causal impact’ or ‘**causality**’ – as it aims to establish that the intervention is the cause of the outcome.

Impact evaluation does this by trying to answer the question “what would have happened if the intervention didn’t occur?” or, in some cases, “what would have happened if we tried a different intervention instead?” Impact evaluation helps to answer these questions by using counterfactuals.

Impact evaluation provides evidence that can help inform decision-making about resource allocation, program design, and policy choice.

### What are counterfactuals and why are they important?

The main difference between impact evaluation and other types of evaluation is that impact evaluation establish causality by using **comparison**. Counterfactuals provide this comparison and answer the questions “did the intervention work?” or “would the effects have happened without the intervention?”

Ideally, evaluators would have access to a parallel world where they could compare the same

individuals, businesses, or places under two different scenarios – one where the intervention was implemented and one where it was not. They could then see what would have happened in the absence of the intervention and whether the intervention made a difference.

As this is not possible in real life, counterfactuals approximate that parallel world by constructing or finding a comparison group which is as similar as possible to the group who receive the intervention (i.e. treatment group). A similar comparison group isolates the effects of the intervention - the only difference between treatment group and comparison group is that one group got the intervention, and the other did not. Any difference in outcome can be directly attributed to the intervention.

The main challenge of impact evaluation is ensuring that those in the comparison group are similar to those in the treatment group. Comparison groups can be created in different ways and how they are created gives the name to the different impact evaluation methodologies: Randomised control trials (RCTs), regression discontinuity design (RDD), difference in difference (DiD), panel data, and propensity score matching (PSM).

Some evaluations use a before and after comparison for participants, but do not use a comparison group and therefore cannot isolate other factors that affected participants. Before and after comparison is useful for seeing what has changed for participants but cannot tell us if we should attribute these changes to the intervention or to something else.

## Impact evaluation methodologies

### Which type of impact evaluation methodology should be used?

There are many impact evaluation methodologies which construct the treatment and comparison groups in different ways. They all require different assumptions about what make the treatment and control group comparable in the absence of treatment and may have different data requirements.

The evaluation question, how the beneficiaries are identified and the options available for creating a comparison group affect the evaluation methodology that can be used:

- A randomised control trial chooses the treatment and comparison groups at random from within the eligible population. The random allocation means the two groups should be similar allowing the effect of the policy to be estimated by looking at differences between the groups. A simple rule: if it is possible to randomise who gets treatment, think about using a randomised controlled trial.
- Sometimes there is a form of ‘natural randomness.’ For example, there is a cut-off for treatment, so comparing a subset of the treated just within cut-off and a subset of the comparison group just outside should give us groups that are similar. Evaluation methods using cut-offs are called regression discontinuity design. Phased roll-out provides another example providing timing of roll-out is random.
- If no randomness can be used, another option is selecting the comparison group based on similarities on observable characteristics and comparing the treated and control groups before and after treatment. This relies on the groups being similar based on what can be observe about them. These evaluation methods include difference-in-differences, panel data methods (fixed effects or first differences) and propensity score matching.

**Remember:** When thinking about impact evaluation, methodologies that require weaker assumptions and less data are preferred.

The table below summarises the main impact evaluation methodologies, their data requirements, the assumptions they rely on and how the comparison group is created.

Table 1: Impact evaluation methodologies

Methodology	Description	Comparison group	Assumptions	Data requirements
<b>Randomised controlled trials- RCT</b>	RCT is defined by random assignment to treatment or comparison group.	Units* randomly assigned to the comparison group.	<ul style="list-style-type: none"> <li>Randomisation is implemented correctly so units* in the treatment and comparison groups have identical observed and unobserved characteristics (i.e. they are statistically identical)</li> </ul>	<ul style="list-style-type: none"> <li>Before ('baseline') and after ('follow-up') data on outcomes for treatment and comparison groups.</li> <li>Data on characteristics are useful for checking randomisation and improving precision of estimates.</li> </ul>
<b>Regression discontinuity design- RDD</b>	RDD uses the 'quasi-randomness' produced by cutoffs for treatment eligibility (i.e. units* are treated as long as they are above or below a certain observable threshold)	Units* that are just below the threshold and ineligible to receive the intervention (untreated)	<ul style="list-style-type: none"> <li>The cut-off cannot be manipulated so units* either side (immediately below and above) of the cut-off have identical observed and unobserved characteristics (i.e. they are statistically identical). Only the treatment changes at the cut-off.</li> <li>Population close to the threshold is representative of the whole population (otherwise the results only apply to those close to threshold)</li> </ul>	<ul style="list-style-type: none"> <li>Before ('baseline') and after ('follow-up') data on outcomes for treatment and comparison groups.</li> <li>Data on characteristics are useful for checking randomisation and improving precision of estimates. Ranking index and eligibility cutoff.</li> </ul>

<b>Difference in difference- DiD</b>	DiD compares changes in outcome for a treatment group to changes in outcome for a comparison group before and after treatment.	Units* that did not participate in the intervention	<ul style="list-style-type: none"> <li>• The treatment group follows the same trend as the comparison group (i.e. in the absence of the intervention, outcomes for both groups would have moved in parallel over time).</li> <li>• There is a specific and known time period for treatment (so that the groups can be compared before and after treatment).</li> </ul>	<ul style="list-style-type: none"> <li>• Before ('baseline') and after ('follow-up') data on outcomes for treatment and comparison groups.</li> <li>• Data on characteristics are useful for checking randomisation and improving precision of estimates.</li> </ul> <p><b>Panel data</b> uses data that follows the same units* over time allowing to control for things that remain constant for each unit across time.</p> <ul style="list-style-type: none"> <li>• Fixed effects (FE) include dummy variables for unit* characteristics and anything that happened in a given time period.</li> <li>• First differences (FD) subtracts the outcome of a previous period in the final regression, cancelling out anything that remains constant across time.</li> </ul>
<b>Propensity score matching- PSM</b>	PSM matches every treated unit* with the most similar unit* in the comparison group, according to a propensity score (the closest match based on observable characteristics).	For each treated unit*, the corresponding matched untreated unit*.	<ul style="list-style-type: none"> <li>• Propensity scores for treated and untreated units* sufficiently overlap and treated units* are matched with similar untreated units*.</li> <li>• The matching criteria is relevant to selection into treatment.</li> <li>• There is no characteristics that affects participation in the intervention beyond the observed characteristics used for the matching.</li> </ul>	<ul style="list-style-type: none"> <li>• Before ('baseline') and after ('follow-up') data on outcomes for treatment and comparison groups.</li> <li>• Data on characteristics are useful for checking randomisation and improving precision of estimates.</li> </ul>

\*Units can be individuals, businesses or places depending on the nature of the intervention.

## The Maryland Scientific Methods Scale (SMS)

The Maryland Scientific Methods Scale (SMS) ranks policy impact evaluations from 1 (least robust) to 5 (most robust) according to the method used and the quality of implementation. Robustness, as judged by the Maryland SMS, is the extent to which the method deals with the selection biases inherent to policy evaluations and reduces assumptions.

We use SMS to score the evidence for our evidence reviews and toolkits on local economic growth policies, and we exclude studies that do not use robust methods of evaluation. More information on the variety of commonly employed methods we examine and how we place them on the Maryland SMS can be found [here](#).

## Where can I learn more?

What Works Growth resources:

- How to evaluate – <https://whatworksgrowth.org/resource-library/how-to-evaluate/>
- An 8-step guide to better evaluation – <https://whatworksgrowth.org/resource-library/8-step-evaluation-guide/>
- Guide to scoring the evidence – [https://whatworksgrowth.org/wp-content/uploads/16-06-28\\_Scoring\\_Guide.pdf](https://whatworksgrowth.org/wp-content/uploads/16-06-28_Scoring_Guide.pdf)

Other resources:

- Magenta book – [https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT\\_Magenta\\_Book.pdf](https://assets.publishing.service.gov.uk/media/5e96cab9d3bf7f412b2264b1/HMT_Magenta_Book.pdf)
- World Bank, Impact evaluation in practice – <https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice>

## Thinking about impact evaluation

### When is impact evaluation appropriate?

It is not feasible to do impact evaluation for all interventions, and policymakers need to think carefully about which interventions are worth evaluating – what is the purpose (i.e. how will results be used) and will it produce useful findings (i.e. robust results). Budget and time restrictions, data constraints, and intervention scale may all affect the viability of an impact evaluation and the value (robustness) of its results.

A clear purpose means defining the question to be answered. In many cases, the question may be more suited to a process evaluation. The table below presents a summary of types of evaluation and evaluation questions with some typical examples.

Table 2: Types of evaluation and evaluation questions

Type of evaluation	Learning objective	Type of question	Typical evaluation questions
<b>Impact evaluation</b>	What difference did the intervention make?	<i>Casual question</i> – Question whether a relationship is actually causing change.	<ul style="list-style-type: none"> <li>• Did the intervention achieve the expected outcomes?</li> <li>• Did the intervention cause the difference?</li> <li>• To what extent have different groups been impacted in different ways, how and why?</li> </ul>
<b>Value-for-money evaluation</b>	Was this a good use of resources?	<i>Cost-effectiveness questions</i> – Question whether the effects on outcomes were worth the resources invested.	<ul style="list-style-type: none"> <li>• How cost-effective was the intervention?</li> <li>• Is the intervention the best use of resources?</li> </ul>
<b>Process evaluation</b>	What can be learned from how the intervention was delivered?	<i>Mechanism question</i> – Questions about the process of how something works.	<ul style="list-style-type: none"> <li>• Was the intervention delivered as intended?</li> <li>• What worked well or less well, and why?</li> <li>• What could be improved?</li> <li>• How has the context influenced delivery?</li> </ul>

Impact evaluations are particularly useful in areas with less evidence. This includes innovative interventions, pilot programmes to be scaled up, and interventions for which there is a lack of robust evidence in a specific context.

Some policies areas are easier to evaluate than others because the design or implementation process helps with the selection of a comparison group (e.g. interventions with random participant selection or implemented in cohorts and with waiting lists).

Measuring changes and finding an appropriate comparison group is easier for larger interventions. The size of the expected impact can also affect the findings - if the expected impact is very small, it may be too small to be observed.

There are also cases where an impact evaluation is not possible. For instance impact evaluation of umbrella programmes or overarching economic strategies (e.g. an industrial strategy) pose methodological issues and causal effect cannot be identified. First, they are very broad and have multiple interventions making complex to disentangle effect on outcomes (i.e. attribution). Second, delivery can vary significantly among participants and interventions, and comparison is not possible. Third, objectives are higher order and often effects cannot be observed or measured. Forth, in the

case of overarching strategies there is no comparison group. In these cases, it is advisable to select the most relevant and impactful projects and consider doing separate impact evaluations.

When thinking about conducting an impact evaluation, consider:

- **What is the purpose of the evaluation?** It should have a clear purpose, including how results will be used.
- **Will the findings inform important decisions?** For example, try to evaluate interventions with important budget implications or affecting many people.
- **Is there any evidence that show if the intervention works and the scale of the impacts it can produce?** Is there evidence available from similar intervention under similar circumstances? Impact evaluation is more valuable when there is not a solid evidence base on the interventions and its outcomes, or it does not include evidence for a similar context.
- **Is the cost of the evaluation proportional to the benefits of the findings and the scale of the intervention?** The cost should be proportionate to the size and expected impact of the intervention and the usefulness of the findings.

## What to consider if thinking of doing an impact evaluation?

If you are thinking about an impact evaluation, keep in mind the following:

- **Logic model** – Is there a consistent logic model? Are the outcomes well defined and measured? Are they realistic? A good logic model helps you evaluate your intervention.
- **Timescale** – Are expected outcomes likely to have materialised within the timescale of the evaluation?
- **Size and quality of the treatment and comparison group** – Can a comparison group be found? Both groups need to be sufficiently large to ensure that the findings are statistically significant, allowing for some participants to drop-out.
- **Data requirements** – Does the data required exist? If not, how will it be collected? Will the data be available within the timescale of the evaluation? Outcome data is required for treatment and comparison groups, both before and after treatment.

**Remember:** Programme objectives are key. They represent what an intervention aims to achieve. The more concrete the objectives in terms of target population, magnitude, and timing of the expected changes, the easier it will be to track progress and conduct an evaluation.

In addition, a set of defined indicators and data collection techniques at all the levels of the intervention (i.e. a monitoring system) are required to track implementation and results.



## Making evaluation easier

Evaluation should be part of programme design, helping policymakers think through what success would look like, the intervention's objectives and outcomes, and the data required to measure them.

There are some instances where administrative or monitoring data can be used to carry out an impact evaluation that was not planned in advance, but this often makes it harder to construct a suitable comparison group or have the required data. Thinking about impact evaluation at an early stage can help design the intervention in a way that impact evaluation methodologies with weaker assumptions and fewer data requirements can be used (e.g. RCT). It can also help to ensure the data collection required is planned for at the different stages, and across both treatment and comparison groups.

Impact evaluation is not always suitable or advisable. Consider if an impact evaluation is appropriate (see subsection above), if the intervention is worth evaluating and what questions you want to answer. If the answer is a definitive 'Yes,' consider carefully what would be a good comparison group, what data would be needed and how could it be collected, and what adjustments to the design would be required to make that possible and how they could affect the delivery and results.

## Where can I learn more?

What Works Growth resources:

- How to evaluate – <https://whatworksgrowth.org/resource-library/how-to-evaluate/>
- An 8-step guide to better evaluation – <https://whatworksgrowth.org/resource-library/8-step-evaluation-guide/>
- Guide to scoring the evidence – [https://whatworksgrowth.org/wp-content/uploads/16-06-28\\_Scoring\\_Guide.pdf](https://whatworksgrowth.org/wp-content/uploads/16-06-28_Scoring_Guide.pdf)
- Logic model guide – <https://whatworksgrowth.org/resource-library/using-logic-models/>

This work is published by the What Works Centre for Local Economic Growth, which is funded by a grant from the Economic and Social Research Council, the Department for Business and Trade, the Ministry of Housing, Communities and Local Government, and the Department for Transport. The support of the Funders is acknowledged. The views expressed are those of What Works Growth and do not represent the views of the Funders.

Every effort has been made to ensure the accuracy of the report, but no legal responsibility is accepted for any errors omissions or misleading statements.

The report includes reference to research and publications of third parties; What Works Growth is not responsible for, and cannot guarantee the accuracy of, those third party materials or any related material.

October 2024

What Works Centre for Local  
Economic Growth

info@whatworksgrowth.org  
@whatworksgrowth

[www.whatworksgrowth.org](http://www.whatworksgrowth.org)

© What Works Centre for Local  
Economic Growth 2024



HM Government



Economic  
and Social  
Research Council