



# Guidance Note

## Best Practice Monitoring, Evaluation and Review

December 2019

© Crown Copyright



This work is licensed under the Creative Commons Attribution 4.0 International licence. In essence, you are free to copy, distribute and adapt the work, as long as you attribute the work to the Crown and abide by the other licence terms.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>. Please note that no departmental or governmental emblem, logo or Coat of Arms may be used in any way which infringes any provision of the [Flags, Emblems, and Names Protection Act 1981](#). Attribution to the Crown should be in written form and not by reproduction of any such emblem, logo or Coat of Arms.

The Treasury URL at December 2019 for this document is  
<http://www.treasury.govt.nz/guide/best-practice-monitoring-evaluation-review>

# Contents

- Introduction .....1
- 1 The purpose of monitoring, evaluation and review .....2
- 2 Context .....3
- 3 What do we mean by monitoring, evaluation and review? .....5
- 4 Questions to ask .....5
- 5 Integrating MER into the policy process .....6
- Annex 1: Starter Template for MER planning table .....10

## Introduction

This Guidance Note describes the key elements of Monitoring, Evaluation and Review. These elements should underlie the development of all government policy initiatives that involve proposals to create, amend or repeal primary or secondary legislation (a 'government regulatory proposal').

The material in this document should be read in conjunction with the Regulatory Impact Analysis Handbook.

## Further information

The formal requirements for all government regulatory proposals are set out in the Cabinet Office circular: CO(17)3: Impact Analysis Requirements available at [dpmc.govt.nz/publications/co-17-3-impact-analysis-requirements](https://dpmc.govt.nz/publications/co-17-3-impact-analysis-requirements): There is further guidance in the revised Cabinet Manual, available at [dpmc.govt.nz/our-business-units/cabinet-office/supporting-work-cabinet/cabinet-manual](https://dpmc.govt.nz/our-business-units/cabinet-office/supporting-work-cabinet/cabinet-manual).

The Guide to Cabinet's Impact Analysis Requirements provides guidance on how to meet the formal requirements for regulatory proposals. The Guidance Note on Effective Consultation for Impact Analysis also provides useful guidance on the policy development of regulatory proposals. Both are available on Treasury's Regulation webpage ([treasury.govt.nz/regulation/impact-analysis](https://treasury.govt.nz/regulation/impact-analysis)).

## Acknowledgements

We are indebted to Clare Bear of the Ministry for Primary Industries' (MPI's) Research & Evaluation Team for her contribution to this guidance.

# 1 The purpose of monitoring, evaluation and review

Monitoring, Evaluation and Review (MER) are three of the most overlooked, but most important, steps in the policy development process.

A Regulatory Impact Assessment (RIA) must establish the agency's plans for monitoring, evaluating, and reviewing the performance of a proposal over time. The RIA must answer the following questions:

- How will the Agency determine when and whether the regulatory changes have performed well?
- How will the Agency assess whether the preferred option continues to have a greater net-benefit than alternatives?

A mere assertion that “the proposed policy will be monitored and evaluated in due course” is insufficient for an impact analysis. Past experience has demonstrated that in these cases MER is rarely implemented, adequately or at all.

Inadequate MER presents both a lost opportunity to assess and communicate the value of an effective regulatory approach (ie, the positive impacts it is making); and a potential risk, eg, an undetected and *ongoing* inefficient waste of public taxpayer money and resources, sub-optimal effectiveness of a regulatory approach, or damage to the reputations of individuals, sectors or the country.

Monitoring and evaluation involves costs, which should be factored in to the analysis of options, and implementation budgets.

The overall approach selected and scalability of resources required (time, upfront costs for data collection and analysis, etc) will depend on the magnitude, complexity, and importance/ public interest in the proposed policy. As a general guideline, based on international good practice, MER should be at least 1% – 5% of the total intervention time and cost, and ideally 5% – 10%.

Some agencies have internal research and evaluation staff with specialist expertise, who can provide support, eg, how to design the best evaluation approach or a survey questionnaire; and advise about existing information sources or useful indicators to measure. Such specialist support staff are often able to suggest ways to measure progress or effectiveness where policy analysts may assume it is not possible.

## 2 Context

This guidance is set within the context of a wider number of reviews and strategies:

### 2.1 Improving the quality of policy advice

The 2010 Scott Review on *“Improving the Quality and Value of Policy Advice”*<sup>1</sup> emphasised that it is important to:

1. Use more/rigorous evaluation and monitoring, from the beginning of a policy process;
2. Put more focus on the indicators of the effectiveness of a policy (not just whether it is delivered on time/budget).

### 2.2 Post legislative scrutiny

A Treasury intern research project<sup>2</sup> in 2018-19 undertook an empirically-based view of the extent to which *ex post* MER of government regulatory proposals is undertaken, and the factors which affect this. In this way, the project examined the “reality” of post-legislative scrutiny in New Zealand, and sought to reconcile it with the reported “intention” of departments. It was the intention that the project inform the Treasury’s training and guidance material to support good practice MER and contribute to better regulatory stewardship.

Interviews conducted for the report revealed that the “reality” of post-legislative scrutiny in New Zealand is still ad hoc, despite encouraging changes. Generally, agencies had not fully carried out proposals for MER. Some interviewees candidly acknowledged they had been unaware of proposals for MER contained in the relevant RIA, prior to the interviews. There was wide variation in practised MER, from extensive and comprehensive MER taking place, to none at all. For the majority of regulatory changes examined, only monitoring had taken place. In some, it was unclear whether such monitoring was actively contributing to regulatory effectiveness.

There was also variation in the “intention” to conduct post-legislative scrutiny. Generally, the MER section of the RIA was seen as a place for high-level, initial thinking, rather than for comprehensive plans. While some RIAs stepped through particular MER initiatives and plans that would be used, others simply noted key indicators to be tracked, or that some kind of review would be undertaken at some point.

Officials identified four main barriers which prevent post-legislative scrutiny from taking place on a more systematic basis. These barriers are the reverse of the factors which officials identified as having facilitated post-legislative scrutiny.

The most commonly cited barrier was the absence of political prioritisation for post-legislative scrutiny. Ministers determine both the policy direction and the priorities for their departments, and post-legislative scrutiny is rarely a priority for ministers unless there is the

---

<sup>1</sup> <https://treasury.govt.nz/publications/improving-quality-and-value-policy-advice-findings-committee-appointed-government-review-expenditure>

<sup>2</sup> *Post-legislative Scrutiny: Reconciling Intention and Reality*, Charles Cox, Intern Research Project, February 2019

perception of public accountability or it aligns with ad hoc political priorities. The consequence is a variable culture of inquiry, where arrangements for MER are left almost entirely to departmental discretion.

Secondly, officials pointed to a lack of guidance to support post-legislative scrutiny as a barrier. However, guidance is unlikely to address a lack of systematic prioritisation being given to MER, save to change perspectives about its importance. Thus, this barrier must be viewed as secondary.

Third, interviewees noted the inherent difficulties in collecting and analysing data for the purpose of post-legislative scrutiny. But because data difficulties can be mitigated by resourcing and prioritisation, this barrier is best seen as a subset of the first.

Finally, officials noted that it was sometimes difficult to establish the benchmark and intent of regulations because of the loss of institutional knowledge through staff turnover.

### 2.3 Regulatory stewardship

Regulatory stewardship is New Zealand's way to prompt government agencies to direct more attention to monitoring and assessing the costs and benefits of existing regulatory systems.

**The OECD Regulatory Policy Committee advocates the review of the stock of regulation to be an integral part of the regulatory**

- Existing regulatory systems are extensive, and their accumulative effects across communities, the economy, and time are pervasive. Even a small improvement in the quality of existing stock could bring large gains to a society.
- Regulatory endeavour is essentially experimental in nature. Even with smart and robust design and delivery informed by an excellent understanding of those being regulated, we need to pay continuous attention to:
  - the risks of unintended consequences and unnecessary costs, and
  - the way that regulatory systems interact with people and institutions, with other regulatory systems, and with technological and social change.
- Lessons from reviews and evaluations will inform and improve the design of new regulatory instruments.
- Reviewing the regulatory stock and its impacts will enhance public trust in government, particularly when reviews draw on stakeholder and wider public input.

Since 2015, the major regulatory departments have been required to report information on the "fitness-for-purpose" of the regulatory systems in which they have major roles. One objective of this reporting requirement is to encourage departments to develop systems and processes to assess and understand regulatory systems.

Fitness-for-purpose assessments are "whole-of-system"; looking across all aspects of the system, from policy to delivery, and the impacts on the ground.

This "whole-of-system" focus means assessments:

- are ideally contributed to by agencies from across the regulatory system, and
- include input and views from stakeholders and regulated parties.

Assessments are best done periodically according to a schedule agreed between the system agencies, probably with some identified “trigger points” that would prompt the agencies to consider an out-of-schedule assessment.

## 2.4 Relationships between fitness-for-purpose assessments, and monitoring, evaluation and review

- In contrast to fitness-for-purpose assessments, monitoring, evaluation and reviews will not necessarily take a whole-of-system focus, although regulatory stewardship practice should increasingly encourage system-wide approaches to these exercises.
- Monitoring, evaluation and review will either provide information to feed into assessments, or in the case of whole-of-system reviews, take the place of assessments.
- Assessments may also point to the need for monitoring, evaluation or review.

## 3 What do we mean by monitoring, evaluation and review?

The different terms and meanings, at a glance:

<i>What is happening?</i>	= monitoring (tracking progress as you go)
<i>What do people think about it?</i>	= research (collecting and analysing data)
<i>What's happened?</i>	= review (analysing how well things are progressing or 'what happened', maybe as part of statutory review requirement or with an independent expert's opinion)
<i>'So what' about what's happened?</i>	= evaluation (assessing the evidence against success criteria, to reach a judgement about 'how good' the implementation or impacts are/ were)

## 4 Questions to ask

When thinking about MER for a RIA, it is helpful to ask the following questions:

### How will the impact of the new arrangements be monitored?

- How will you know whether the impacts anticipated actually materialise?
- System-level monitoring and evaluation
- Are there already monitoring and evaluation provisions in place for the system as a whole (ie, the broader legislation within which this arrangement sits)? If so, what are they?
- Are data on system-level impacts already being collected?



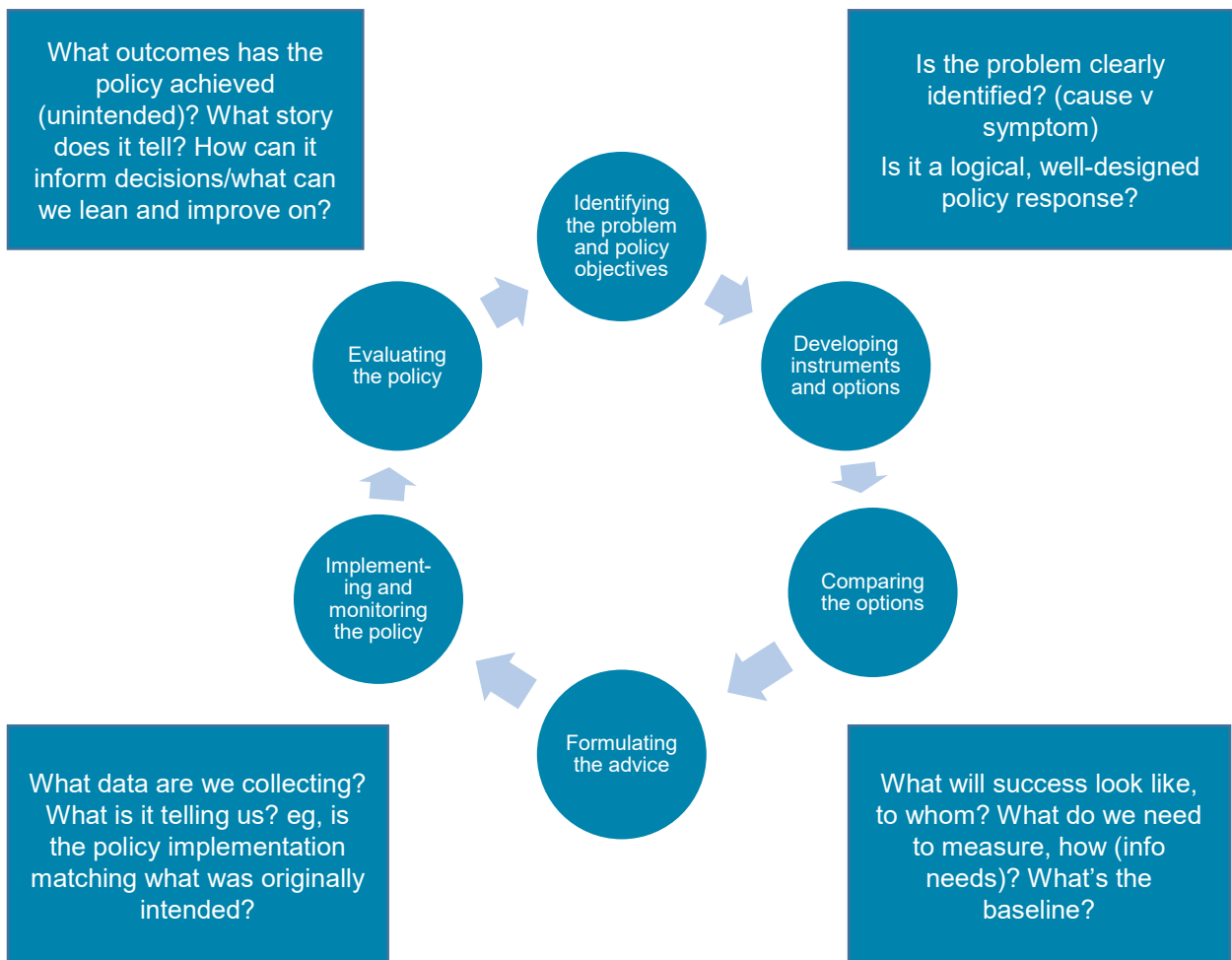
- Are data on implementation and operational issues, including enforcement, already being collected?
- Will you need to collect extra data that is not already being collected? Please specify.

**When and how will the new arrangements be reviewed?**

- How will the arrangements be reviewed? How often will this happen and by whom will it be done? If there are no plans for review, state so and explain why.
- What sort of results (that may become apparent from the monitoring or feedback) might prompt an earlier review of this legislation?
- What opportunities will stakeholders have to raise concerns?

## 5 Integrating MER into the policy process

It’s not enough to think about Monitoring Evaluation and Review as an after-thought – it must be integrated into the policy process from the beginning. Here are some questions that policy analysts should ask themselves throughout that process:



## 5.1 The three elements of monitoring, evaluation and review

There are three key elements to consider with monitoring, evaluation and review (MER):

1. A framework to measure the expected 'success' of the policy against;
2. A record of the baseline evidence methods and sources; and
3. A plan and budget to carry out monitoring and evaluation activities.

A regulatory impact analysis is not expected to deliver these three elements to completion; but it should contribute at least the foundation ideas, plans and priority measures, so that managers and Ministers can have confidence that there will be relevant evidence collected to assess the effectiveness of the regulatory policy.

## 5.2 MER Framework

A key question to determine is, what will success look like? This could mean 'success' in terms of how well the regulatory intervention is or remains fit-for-purpose within the wider system and dynamics that created the problem, or how well/ efficiently it is being implemented, or the effectiveness of its impact.<sup>3</sup>

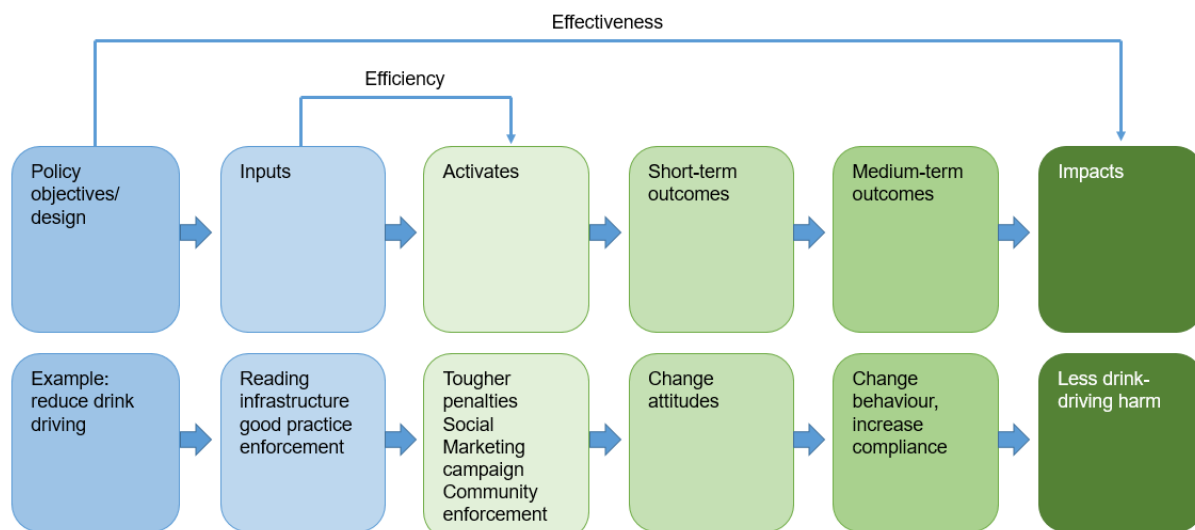
Creating clearly described policy objectives is a very helpful foundation for a framework around these approaches to defining 'success', ie, to know *what* to measure, monitor and evaluate as must-have vs nice-to-have evidence. Being explicit about the success criteria, values or principles underlying the policy objectives is also important.

Common types of framework include:

- A set of success criteria/values similar to the policy objectives; a 'rubric' of broadbrush criteria or expectations; and/or a set of specific performance indicators or targets.
- An outcomes or intervention logic model framework based on a theory of change analysis. The model diagram below (similar to that required by Treasury for budget bids) is ideal to identify the key short-, medium- and long-term elements and causal links expected between the intervention and policy's objectives, to create measures for (and test for underlying assumptions and unintended consequences).

---

<sup>3</sup> These three approaches are known as formative, process and summative/ impact evaluation.



### 5.3 Baseline evidence and counter-factual

Baseline evidence is an important resource to monitor, evaluate or review a policy, ie, to be able to compare the current or future impacts of a regulatory intervention back against the pre-intervention status quo, and counter-factual, ie, what difference has the regulatory intervention made since; or in comparison to what might trend have been expected without it?

Accordingly, it is helpful if officials developing the regulatory impact analysis *document* what the main baseline evidence is, and methodology notes where applicable. This will provide a basis for future monitoring and evaluation activities to repeat and update the evidence. For example:

Success criteria or expected outcome area	Baseline information/result	Information source	Information method (any limitations, budget or skills required etc)	How often to collect/ repeat information

Anticipating future evidential needs at the early policy development stage makes future MER much faster and easier, and much higher quality, because the evidence base is there to build upon, rather than future officials attempting to ‘reinvent the wheel’ some years later, potentially with a loss of institutional knowledge to guide them. In many cases it is not possible to retrospectively capture baseline or pre-intervention evidence, such as what the public attitudes were several years ago – the opportunity is lost.

## 5.4 Planning MER activities

Plans should be made for how and when the regulation will be monitored, evaluated and reviewed. Agencies should consider committing to a periodic review, either through a sunset-review clause in the regulation itself, or through committing to collect and monitor information for evaluating regulatory performance.

Using the monitoring and evaluation framework and existing baseline evidence above help to provide the basis for planning what is important to measure through implementation and evaluate or review periodically. MER does not happen by itself, it needs planning, resources, and skills.

Example of an MER Plan Table<sup>4</sup>:

Expected outcome or benefit	Measure (indicator, baseline, target etc)	Timeframe and frequency to measure	Actions, data and budget/ resources required	Data sources and management	Key roles/ who will be involved

Annex I sets out a starter template for a MER planning table.

## 5.5 Reviews

Reviews should consider the following issues:

- Is there still a problem (and is it the one originally identified)?
- Are the objectives being met?
- Are the impacts as expected? Are there any unforeseen problems? Are there any indirect effects that were not anticipated?
- Is intervention still required? Is the current intervention still the most appropriate, or would another measure be more suitable?

On-going or periodic consultation with stakeholders may be appropriate, in which case the arrangements for this should be agreed. It may be appropriate to establish a feedback mechanism (eg, a way for stakeholders to ask questions or lodge complaints). Regular, public reporting on the effectiveness of the regulation may also be considered.

---

<sup>4</sup> Based on MPI Outcomes or Benefits Plan template

# Annex 1: Starter Template for MER planning table

1. **State the main Actions / Decisions that the Review / Evaluation will likely lead to**
2. **State the main audiences for the Review / Evaluation findings and their main likely questions**
3. **Main milestones of the Review/Evaluation**

For example:

- Report on the implementation provided by x date to y audience
- Interim evaluation report to inform further development/refinement of the legislation by xx date to yy audience
- Full evaluation report (after all foreseeable impacts of the legislation are likely to have become apparent) by xxx date to yyy audience

4. **Main foreseeable challenges, obstacles, risks, limitations and assumptions around Monitoring, Evaluation, and Review**

For example, certain data being difficult to obtain, difficulty of predicting when certain impacts of the legislative change will be fully realised, difficulty of obtaining feedback from certain groups of people most impacted by the legislative change.

5. **Main data sources and tools for Monitoring, Evaluation, and Review**

Briefly state this. For example, surveys of x using Stats NZ survey guidance, data from y source or potential data gaps and the need for additional data collection (may require a budget).

**Standard Caveat:** This plan may change, and is based on current knowledge and resources.