RICARDO

Ricardo Energy & Environment

Guidance document for ex-post evaluation of climate policies in Effort Sharing sectors

Report for DG Climate Action
Contract No 340201/2018/791993/SER/CLIMA.C.2

Öko-Institut e.V. Institut für angewandte Ökologie Institute for Applied Ecology

WAGENINGEN UNIVERSITY & RESEARCH

vito

ED 11784 | Guidance ex-post evaluation | Date 23/09/2020

Customer:

DG Climate Action

Customer reference:

Contract No 340201/2018/791993/SER/CLIMA.C.2

Confidentiality, copyright & reproduction:

.

Contact:

James Tweed
Ricardo Energy & Environment
Gemini Building, Harwell, Didcot, OX11 0QR, United Kingdom
VITO - EnergyVille
Belgium

t: +32 (0) 1235 75 3007
e: james.tweed@ricardo.com

Ricardo is certificated to ISO9001, ISO14001 and OHSAS18001

Author:

Renders Nele, Dauwe Tom, Pasquel Garcia Tatiana, VITO – Förster Hannah, Gores Sabine, Schumacher Katja, Noka Viktoria, Öko – de Boer Florianne, Forster Daniel, Morgan-Price Samantha, Ricardo

Approved By:

Daniel Forster

Date:

23 September 2020

Ricardo Energy & Environment reference:

Ref: ED11784- Guidance ex-post evaluation

Table of contents

1 Introduction

Ex-post evaluation of climate policies and measures is necessary to understand what the actual effects of climate policies have been, along with their overall effectiveness (i.e. emission savings) and efficiency (cost of delivering the emission savings). This evaluation shows how the outcomes compare with initial expectations, as well as identifying where improvements can be made in the design of the evaluated policy and/or future policies. In this context, the ex-post evaluation of national policies, and effective use of the results from these evaluations, can play a major role in the cost-effective delivery of Member State targets under the Effort Sharing legislation.

Context and rationale of the guidance

This guidance provides support to European Member States to further enhance systematic ex-post evaluation of Effort Sharing related policies. The Effort Sharing legislation establishes binding annual greenhouse gas emission (GHG) targets for Member States for the periods 2013 - 2020 and 2021 - 2030. These targets concern emissions from most sectors not included in the EU Emissions Trading System (EU ETS), such as transport, buildings, agriculture and waste. The national targets will collectively deliver a reduction of around 10% in total EU emissions from the sectors covered by 2020 and of 30% by 2030, compared with 2005 levels. Together with a 21% cut in emissions covered by the EU ETS by 2020 and 43% by 2030, this will allow the EU to achieve its climate targets for 2020 and 2030.

This guidance covers reporting of evaluation results as well as the learning of lessons from ex-post evaluations for the design of future policies from Member States. The guidance is one of the outcomes of the DG Climate Action project '*Capacity Building to Facilitate Implementation of the Effort Sharing Legislation, with Focus on Ex-post Evaluation and Policy Lessons Learned*". In this project, a survey was launched to understand Member States' needs in relation to – among others – ex-post evaluation. The results of this survey indicate a strong interest from Member States in receiving support on the following topics:

- Collection of evidence for ex-post evaluations;
- Quantitative methodologies that can be used to evaluate policies (impact on GHG emissions and costs), including cost effectiveness and cost benefit analysis of emission reductions;
- Key evaluation issues or methodological challenges, such as:
    - Interaction of policies;
    - How to include quantitative emission impacts results from ex-post evaluations in GHG emissions projections; and
    - Quantification of GHG impacts of policies within the scope of the Effort Sharing legislation where the impacts occur outside of the scope (i.e. within ETS sectors).

In a first step to support Member States with the identified needs, a bibliography has been compiled of previous studies that have evaluated ex-post impacts of Member States' policies in Effort Sharing sectors. The studies are collated in a repository, capturing key details, and allowing the identification of methodologies that can serve as good examples for future evaluations. In addition to the bibliography, this guidance has been developed to support Member States in improving their approaches for ex-post evaluation of policies and measures. As there are already a number of guidance documents, the aim is not to replicate what is provided in other sources. Rather, the aim is to provide additional guidance on the **identified needs** of Member States, as well as to illustrate common practice among Member States based on **selected case studies** from the bibliography. Therefore, the compiled guidance has the following key characteristics.

Key characteristics of the guidance for ex-post evaluations:

The guidance …

1. … focuses on priority evaluation needs expressed by Member States;
2. … provides practical and hands-on guidelines on key topics to support policy evaluation;
3. … is concise with links to additional material and documents;
4. … is supported by examples and case studies;
5. … focuses on ex-post evaluation only, but multiple evaluation criteria;
6. … is aligned with the European Commission's Better Regulation Toolbox;
7. … aims to improve completeness of reporting under the Energy Union Governance; and
8. … should support transparent reporting of evaluation results.

Structure of the guidance

The guidance is organized in line with identified needs of Member States (Table 1).

Table 1. Structure of guidance in line with Member States' needs to support ex-post evaluation.

Member States' needs	Specific methodologies or challenges	Chapter of guidance
General frameworks for policy evaluation	1. intervention logic model, 2. counterfactual analysis	Chapter 2
Methodologies for collecting evidence	3. surveys, 4. systematic literature review, 5. focus groups, 6. interviews, 7. monitoring performance data	Chapter 3
Analytical, quantitative evaluation methods	8. indicator analysis, 9. cost effectiveness analysis and cost benefit analysis, 10. regression analysis, 11. decomposition analysis, 12. multi-criteria analysis	Chapter 4
Approaches to deal with key evaluation issues or methodological challenges	13. policy interactions, 14. rebound effect, 15. free riders, 16. uncertainty, 17. emission factors, 18. co-benefits, 19. consistency between ex-post information and projections, 20. splitting the impacts between Effort Sharing and ETS	Chapter 5
Case studies or existing evaluation studies from Member States		Chapter 6

Three elements of an **evaluator's toolbox** are covered by this guidance: (1) frameworks for policy evaluation, (2) evidence collection methods, and, (3) analytical, quantitative methods. The next three chapters therefore describe these elements in a more technical, but concise and step-wise, manner. Each section or methodology of the evaluator's toolbox follows a fixed structure, as explained in the 'How to read the guidance' BOX below. Approaches are based on multiple existing guidance documents[1], such as the European Commission's Better Regulation Toolbox, the European Environment Agency (EEA) evaluation framework, the UK government's Magenta Book, and the World Resource Institute Policy and Action Standard – GHG protocol, as well as specific technical literature or documents. This compilation of multiple sources reflects possible Member States' priorities. Moreover, this guidance document provides clear pointers to other sources, where further information can be obtained on certain steps of the process. To improve the transparency of the guidance and to allow a comparison of methodologies, a fixed set of characteristics is rated, briefly described and presented as a summary table at the beginning of Chapter 3 and Chapter 4. These characteristics include level of complexity, resources needed and ease of communication of the results.

Following the evaluator's toolbox, explanation or guidance is given in Chapter 5 on **how typical evaluation issues or challenges can be tackled** (e.g. how to deal with policy interactions or uncertainty).

The technical guidelines for most of the methodologies or challenges are illustrated in **case studies from the bibliography**. Case studies of ex-post evaluation from Member States are described in more detail in Chapter 6. Attention is given to how the evaluation was performed (evaluation approach) and how the relevant evidence was collected (collection of evidence). It should be noted that while the guidance includes several thorough and elaborated case studies, there was no showcase example of a study that took into account all methodological aspects suggested in the guidance. Instead, a common finding is that many case studies apply simplifications and shortcuts to the proposed methodologies. One main reason for this is presumably difficulty in obtaining the necessary data, as some guidance documents argue that finding ex-post evidence can often be more problematic than forecasting outcomes ex-ante.

[1] Full references of existing guidance documents are given in the 'Want to know more?' sections of Chapters 2 – 5, , where further information on the described methodologies can be found.

How to read the guidance?

Each approach is presented in the guidance according to a fixed structure and contains the following:

Description

☛ Explaining briefly the evaluation methodology in the context of ex-post evaluation of Effort Sharing-related policies.

When to use it?

☛ Describing the conditions or evaluation context under which the methodology can be appropriately applied (e.g. availability of evidence, stakeholders involved, evaluation questions).

How to use it?

☛ Describing the methodological steps to be taken in a concise and illustrative way.

Advantages/disadvantages

☛ Explaining & scoring the strengths and weaknesses of each methodology against a fixed set of characteristics. An explanation of the characteristics is given in the table below.

Characteristics	Score	Explanation
Data requirements	Low	How much data or evidence is required, to apply the method? Are many specific statistics, monitored data, … required? …
Complexity	Medium	How complex is the method? Are specific software requirements needed (other than Excel)?
Usefulness	High	How useful is the method to assess the evaluation criteria? E.g. decomposition analysis has low usefulness because can't directly be linked to one policy or measure.
Resources		How much time and other resources are needed to apply the method?
Evaluation criteria		What and how many of the evaluation criteria can be assessed by applying the method?
Communication / visualisation of results		Are the results easily communicated / visualised and thus raise understanding?

Tools

☛ What tools are publicly available online to support the evaluation methodology?

Data sources

☛ Which publicly available data sources or statistics, other than common sources (e.g. Eurostat), could support the evaluation?

Related topics

☛ Which other sections of the guidance relate to this topic? The other sections are indicated by means of a HYPERLINK.

Case studies of relevant examples of the approach

☛ Which case studies or existing evaluation practices from Member States illustrate this evaluation methodology very well. The cases are indicated by means of a HYPERLINK.

Want to know more?

☛ References to other, existing guidance documents or studies giving more explanation or interesting insights on the methodology.

2 What is an evaluation?

2.1 Introduction

What is an evaluation?

--- *" Evaluations compare critically what has actually happened to what was expected to happen "* ---

There is a wide variety of definitions of the concept of evaluation. In context of the European Commission's Better Regulation initiative – which guides the design of EU policies so they can achieve their objectives at minimum cost in an open and transparent manner - evaluation is defined as an integral part of the whole policy cycle (EC, 2017):

"Evaluation is a tool to learn about the functioning of interventions and to assess their actual performance compared to initial expectations. Evaluations should critically determine whether interventions are fit for purpose and deliver intended objectives at minimum cost. This way, it also draws conclusions on whether the intervention continues to be justified or should be modified to improve its effectiveness, relevance and coherence and/or to eliminate excessive burdens or inconsistencies or simply be repealed. Evaluation goes beyond the assessment of what has happened and considers the wider perspective on why this change has happened and look for causality between the changes and the intervention. Beyond listing outputs and describing changes, evaluations should investigate any links between the observed changes and the intervention. Evaluations are evidence-based and multiple quantitative and qualitative methods have been developed and used over the years to collect, analyze, synthesize and report evidence objectively."

Scope of guidance: intervention, action, measure and/or policy?

This guidance is aimed to support ex-post evaluation of interventions for achieving greenhouse gas emission reductions in sectors under the Effort Sharing legislation. Although not always explicitly stated, in most evaluations, these interventions are policies, policy measures or policy actions. Programmes, schemes and policy packages are also in the scope of this guidance.

Why do an evaluation?

Evaluation not only helps in understanding and improving interventions, but also engages policy makers, stakeholders and the general public. Evaluations are an important part of the policy cycle as they should also provide lessons learnt on why interventions worked well or what are key barriers to an intervention becoming more effective and efficient.

Therefore, evaluations can generate valuable information and contribute to a wide range of initiatives and objectives. The Magenta Book (HM Treasury, 2011) indicates that good evaluation can …

- *"… provide **a sound scientific basis** for policy making, by providing reliable understanding of which interventions work and are effective. An understanding of how and why policies work can also be used to inform the development of new policies, and to improve the effectiveness and reduce the burden of existing ones;*
- *… underpin **practical resourcing and policy making exercises** such as spending reviews and the formulation of new strategies. They can contribute to the setting of policy and programme objectives, and can be used to demonstrate how those objectives are being met; and*

- *they can therefore provide **accountability**, by demonstrating how funding has been spent, what benefits were achieved, and assessing the return on resources. This can help to satisfy external scrutiny requirements and comply with sunset clauses and other formal requirements that make a link between evaluation and the continuation of the policy.*

A good evaluation is therefore a normal and natural part of policy making and effective government and is a powerful tool available to the policy maker."

Similarly, the European Commission's Better Regulation guidelines underline the integral role of evaluation in the policy cycle. The guidelines cover the whole policy cycle from policy design and preparation, adoption, implementation, application (including monitoring and enforcement) towards evaluation and revision (Figure 1). For each phase of the policy cycle, there are a number of better regulation principles, objectives, tools and procedures to make sure that the EU has the best policy possible. These are explained in the Better Regulation Toolbox (EC, 2017) and relate to planning, impact assessment, stakeholder consultation, implementation and evaluation.

Figure 1. EU policy cycle (EC, 2017).

The earlier that an evaluation can be planned in the policy development process, the more appropriate and effective the evaluation will be, as early planning ensures appropriate evaluation methods and identifies data to be monitored during the intervention. The later in the policy process the evaluation is considered the fewer options there are for undertaking it.

What are evaluation criteria?

Some of the most used evaluation criteria are:

- Effectiveness: "How successful the intervention has been in achieving or progressing towards its objectives."
- Efficiency: "The relationship between the resources used by an intervention and the changes generated by the intervention (which may be positive or negative)."
- Coherence: "(…) looking at how well or not different actions work together."
- Relevance: "Relationship between needs & problems in society and the objectives of the intervention and hence touches on aspects of design."

- Other: there are also several other evaluation criteria which might be considered, depending on the type of intervention and the timing of the evaluation. The most common additional criteria are equity, utility, complementarity, coordination, sustainability, and acceptability.

In the BOX below, criteria are further explained and illustrated by typical evaluation questions, as described in the Better Regulation Guidelines (EC, 2017).

Evaluation criteria and typical example questions (Better Regulation, EC, 2017)

Effectiveness: How effective has the intervention been?
The evaluation should analyse the progress made towards achieving the objectives of the intervention, looking for evidence of why, whether or how these changes are linked to the intervention. So, the answer to this question should go further than showing if the intervention is on track. It should seek to identify the factors driving or hindering progress and how they are linked (or not) to the intervention.
Examples of different ways to formulate effectiveness questions are:

- What have been the (quantitative and qualitative) effects of the intervention?
- To what extent do the observed effects link to the intervention?
- To what extent can these changes/effects be credited to the intervention?
- To what extent can factors influencing the observed achievements be linked to the intervention?

Efficiency: How efficient has the intervention been?
The evaluation should always look closely at both the costs and benefits of the intervention as they accrue to different stakeholders, identifying what factors are driving these costs/benefits and how these factors relate to the EU intervention. Efficiency analysis is a key input to policymaking, helping both policy-makers and stakeholders to draw conclusions on whether the costs of the intervention are proportionate to the benefits.
Typical examples of efficiency questions:

- To what extent has the intervention been cost effective?
- To what extent are the costs of the intervention justified, given the changes/effects it has achieved?
- To what extent are the costs associated with the intervention proportionate to the benefits it has generated? What factors are influencing any particular discrepancies? How do these factors link to the intervention?
- Are there opportunities to simplify the legislation or reduce unnecessary regulatory costs without undermining the intended objectives of the intervention?
- If there are significant differences in costs (or benefits) between stakeholders, what is causing them? How do these differences link to the intervention?
- How timely and efficient is the intervention's process for reporting and monitoring?

Coherence: How coherent is the intervention internally and with other actions?
The evaluation should look at how well the intervention works internally as well as with other interventions. The answer to this question should provide evidence of where and how interventions are working well together (e.g. to achieve common objectives or as complementary actions) or point to areas where there are tensions (e.g. objectives which are potentially contradictory, or approaches which are causing inefficiencies).

Typical examples of coherence questions:

- To what extent is this intervention coherent with other interventions which have similar objectives?
- To what extent is the intervention coherent internally?
- To what extent is the intervention coherent with wider, other policy?
- To what extent is the intervention coherent with EU or international obligations?

Relevance: How relevant is the intervention?
The evaluation must look at the objectives of the intervention being evaluated and see how well they (still) match the (current) needs and problems. Relevance analysis is very important because if an intervention does not help to address current needs or problems then it is no longer appropriate regardless of how effective, efficient or coherent it is. Therefore, this is key information that will assist policymakers in deciding whether to continue, change or stop an intervention.

Typical examples of relevance questions:

- To what extent is the intervention still relevant (to the citizens)?
- To what extent have the (original) objectives proven to have been appropriate for the intervention in question?
- How well do the (original) objectives of the intervention (still) correspond to the needs within the Member State?
- How well adapted is the intervention to subsequent technological or scientific advances?

There are also several other evaluation criteria which might be considered:

- Equity: How fairly are the different effects distributed across the different stakeholders / regions / genders / Social groups?
- Utility: To what extent do the changes/effects of an intervention satisfy (or not) stakeholders' needs? How much does the degree of satisfaction differ according to the different stakeholder groups?
- Complementarity: To what extent do policies and interventions support and usefully supplement other policies (in particular those pursued by EU and local/regional authorities)?
- Coordination: To what extent are interventions organised to maximise their joint effects, e.g. by mobilising resources combined with harmonising measures?
- Sustainability: How likely are the effects to last after the intervention ends? It is often hoped that the changes caused by an intervention are permanent. It can be important to test this expectation for interventions which have a finite duration.
- Acceptability: To what extent can we observe changes in the perception of the intervention (positive or negative) by the targeted stakeholders and/or by the general public?

How can a policy evaluation be designed?

To allow for high quality evaluations, multiple steps should be taken during the policy process. These are illustrated in Table 2. Early planning of evaluation in the policy process leads to design and delivery of more appropriate and effective evaluations. Already at the stage of defining policy objectives, possible interlinkages should be identified between the intervention and the outputs/outcomes/impact. Moreover, by considering all policy options or elements during the design stage of the policy cycle, the potential for useful evaluations (e.g. including elaborated monitoring provisions and evaluation milestones) will be enhanced.

Table 2. Steps involved in policy evaluation.

Steps involved in evaluation	Questions to consider
Defining the policy objectives and intended outcomes	• What is the programme logic or theory about how inputs lead to outputs, outcomes and impacts, in the particular policy context?
Considering implications of policy design for evaluation feasibility	• Can proportionate steps be taken to increase the potential for good evaluation? • What adjustments to policy implementation might improve evaluation feasibility and still be consistent with overall policy objectives?
Defining the audience for the evaluation	• Who will be the main users of the findings and how will they be engaged?
Identifying the evaluation objectives and research questions	• What do policy makers need to know about what difference the programme made, and/or how it was delivered? • How broad is the scope of the evaluation?
Selecting the evaluation approach	• Is an impact, process or combined evaluation required? • Is an economic evaluation required? • How extensive is the evaluation likely to be? • What level of robustness is required?
Identifying the data requirements	• At what point in time should the impact be measured? • What data are required? • What is already being collected / available? • What additional data needs to be collected? • Who will be responsible for data collection and what processes need to be set up?
Identifying the necessary resources and governance arrangements	• How large scale / high profile is the policy, and what is a proportionate level of resource for the evaluation? • What budget is to be used for the evaluation and is this compatible with the evaluation requirements? Has sufficient allowance been built in? • Who will be the project owner, provide analytical support, and be on the steering group? • What will the quality assurance processes be?
Conducting the evaluation	• Will the evaluation be externally commissioned or conducted in-house? • Who will be responsible for specification development, tendering, project management and quality assurance? • When does any primary data collection need to take place? • Is a piloting or cognitive testing of research instruments required? • When will the evaluation start and end?
Using and disseminating the evaluation findings	• What will the findings be used for, and what decisions will they feed into? • How will the findings be shared and disseminated? • What will be included in the evaluation report: stages of data collection, assumptions to estimate GHG reductions, methodology description, etc.?
Reflection on the evaluation process	• Which evaluation steps went well? Which evaluation steps can be further improved and how?

Source: Magenta book (HM Treasury, 2011)

An evaluator of any intervention knows that there are many practical problems that can make it difficult exactly to follow all the steps involved in evaluation, listed above. A list of some **typical problems** that can be encountered is (AID-EE, 2006):

- Lack of monitoring data: Lack of monitoring data is maybe the most common problem in an evaluation. For this reason, data gathering ex-post can easily become very costly and time consuming. The best way to handle scarce data sources is to already identify monitoring needs in the design of the policy instrument and to include monitoring from the outset. A description of the consequences of low data availability for the evaluation results is strongly recommended;
- Lack of time and resources. The lack of time and resources often means that compromises have to be made regarding the approach and collection of evidence. Therefore, decisions have to be made regarding what the focus of the evaluation should be;
- Difficult to determine cause-effect relations (input/outcomes): The causality between the observed effects and influencing factors and relating these effects to an individual policy or measure are often difficult to establish. This is because there might be parallel policies or interventions that lead to the same effect (interacting or overlapping policies). Sometimes, there are various exogenous developments influencing the effect as well. In such cases, it is important to double-check the identified relations with other evaluators and stakeholders.

Given these difficulties, the validity and reliability of the evaluation need to be carefully monitored from the beginning of the policy process. Evaluators should determine the desired level of evaluation robustness and completeness depending on a range of factors, for example, the type of intervention being evaluated (e.g. financial, informative), scope of the evaluation (e.g. national, pilot project), evaluation timing (e.g. interim, ex-post evaluation), available resources & time, and the availability and measurability of effects (e.g. existing monitoring data and key indicators). The choice of evaluation method will therefore vary for every evaluation.

Related topics

Intervention logic model
Counterfactual analysis
Assessing policy interactions
Uncertainty

Want to know more?

- (AID-EE, 2006) Guidelines for the monitoring, evaluation and design of energy efficiency policies - How policy theory can guide monitoring & evaluation efforts and support the design of SMART policies, AID-EE Intelligent Europe, 2006, https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/aid-ee_guidelines_en.pdf
- (DAC Network, 2018) DAC Criteria for Evaluating Development Assistance, OECD DAC Network, https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm
- (deGEval, 2016) Standards für Evaluation. (in German), deGEval - Gesellschaft für Evaluation, 2016, : https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf
- (EEA, 2016) Environment and climate policy evaluation, EEA, 2016, https://www.eea.europa.eu/publications/environment-and-climate-policy-evaluation
- (EC, 2017) Better Regulation: guidelines and Toolbox, European Commission, 2017, https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en
- (Ecologic, 2005) Cost-effectiveness of environmental policies - An inventory of applied ex-post evaluation studies with a focus on methodologies, guidelines and good practice, Ecologic, 2005, https://www.ecologic.eu/sites/files/project/2013/1731_Cost-effectiveness_conclusions.pdf
- (HM Treasury, 2011) The Magenta Book, Guidance for evaluation, HM Treasury, 2011, https://www.gov.uk/government/publications/the-magenta-book
- (IOB, 2009) Evaluation policy and guidelines for evaluations, IOB - Policy and Operations Evaluation Department – Dutch Ministry of Foreign Affairs, 2009, https://www.oecd.org/dac/evaluation/iob-evaluation-policy-and-guidelines-for-evaluations.pdf
- (Tavistock, 2010) Guidance for transport impact evaluations - Choosing an evaluation approach to achieve better attribution, Tavistock Institute, 2010, https://www.tavinstitute.org/wp-content/uploads/2013/01/Tavistock_Report_Guidance_for_Transport_Evaluations_2010.pdf

- (WHO, 2013) Evaluation practice guidebook, WHO, 2013, https://apps.who.int/iris/bitstream/handle/10665/96311/9789241548687_eng.pdf;jsessionid=76346C2119DBA79300BD99301C35B774?sequence=1
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

2.2 Frameworks for policy evaluation

The starting point for an evaluation is to consider how the intervention was expected to work. This requires identification of the different steps and actors involved in the intervention to allow the identification of the expected cause and effect relationships (EC, 2017).

Establishing a framework for the evaluation provides a consistent and systematic means to design the evaluation, and to collate and analyse the existing and newly collected evidence, as well as to generate and interprete the results. It will improve the understanding of what existing evidence tells us and will allow identification of those gaps in the evidence base on which the evaluation should focus. The evaluation framework is most likely to be based on an intervention logic model and/or counterfactual analysis to estimate the (typically quantitative) impact.

2.2.1 Intervention logic model

Description

--- " Intervention logic is 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 (GOV.UK, 2018)" ---

Taking most often the form of a narrative accompanied by a diagram, the 'intervention logic' is an important analytical tool to structure, conduct and communicate retroactive assessments of interventions. The framework will also be helpful in identifying what the evaluation should measure and in deciding on how to evaluate (evaluation approach or methodology). EEA (2016) describes the intervention logic framework to identify the cause and effects relationships between different elements of the intervention. The following elements of the intervention are described (Figure 2):

- *"**Inputs** are the resources dedicated to the design and implementation of a measure (staff, administrative structures, financial investment, training, awareness raising, etc.);*
- ***Outputs** are the tangible effects of a measure (e.g. the number of renewable installations, the number of organisations certified, number of completed training courses);*
- ***Impacts** are the ultimate effects of these changes in behaviour on the environment and human health. Impacts may occur, after a certain period, among direct addressees or indirect addressees (e.g. change in personal income, wellbeing);*
- ***Results** are, in turn, more short-term effects: they are immediate changes that arise for direct addressees at the end of their participation in an intervention (e.g. reduced costs of training opportunities provided, turnover);*
- ***External factors** (e.g. the weather) and other policies can intervene on – i.e. support or weaken - the effect of policies."*

The yellow balloons represented in Figure 2 are the criteria typically used in the evaluation of policies. So, an *"intervention logic is about how an intervention or a policy is intended to operate in order to achieve objectives along the objectives–inputs–outputs–impacts chain. The direct participation of key stakeholders of the policy process can help to clarify this. Evaluators can address specific aspects of* *this chain or isolate specific elements of the intervention logic that may put the achievement of objectives at risk."*

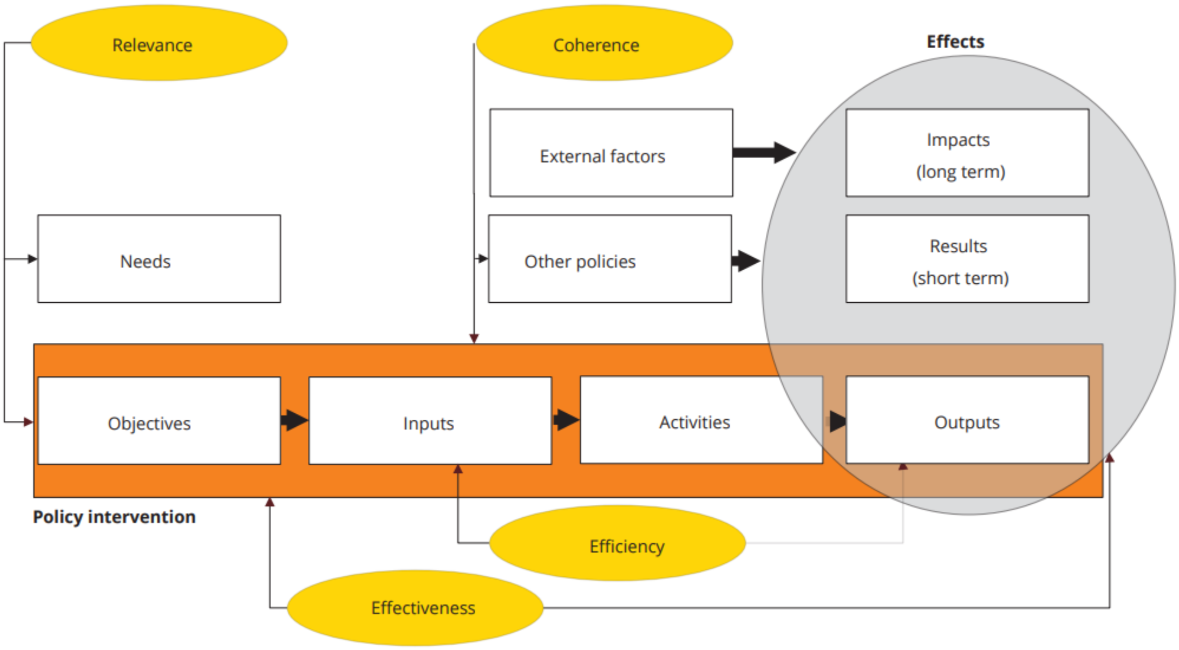


Figure 2. Policy evaluation framework: intervention logic (EEA, 2016).

The intervention logic does not include detail about absolutely everything that happens in, and as a result of, an intervention, but *"summarises the aspects that are critically important in explaining how the intervention produces the changes that it is aiming to achieve."* Therefore, it is important when creating a logic model to decide what these critical aspects are that contribute to its outcomes. Logic models tread a fine line between being simple, easy to understand and use, and reflecting the complexity of the real world. Therefore, wherever this line is drawn, logic models will *"remain relatively mechanistic, linear representations of processes of change which do not reflect the full complexity of the real world but provide a simplicity that has advantages for planning and conducting evaluations (GOV.UK, 2018)".*

How to use it?

To develop an intervention logic, it will be helpful to draw on, for instance, a range of evaluation evidence, primary research, intervention documents and strategy reports. If the intervention logic being used was not drawn up as part of the policy-making process, it is also worthwhile consulting with those stakeholders who were involved in conceptualising the intervention so their insights can inform the understanding of the intervention logic. Moreover, considering a diversity of questions, as illustrated for UK transport policies in Table 3, might support the – sometimes challenging – mapping of cause and effect relationships between the different elements of the intervention (GOV.UK, 2018). A possible resulting intervention logic diagram is illustrated for the evaluation example of 'EU Framework for Metering and Billing of Energy Consumption' as performed by the EU Commission (EC, 2016).

Table 3. Questions to consider in creating intervention logic (example UK transport policy).

	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.

Source: (GOV.UK, 2018)

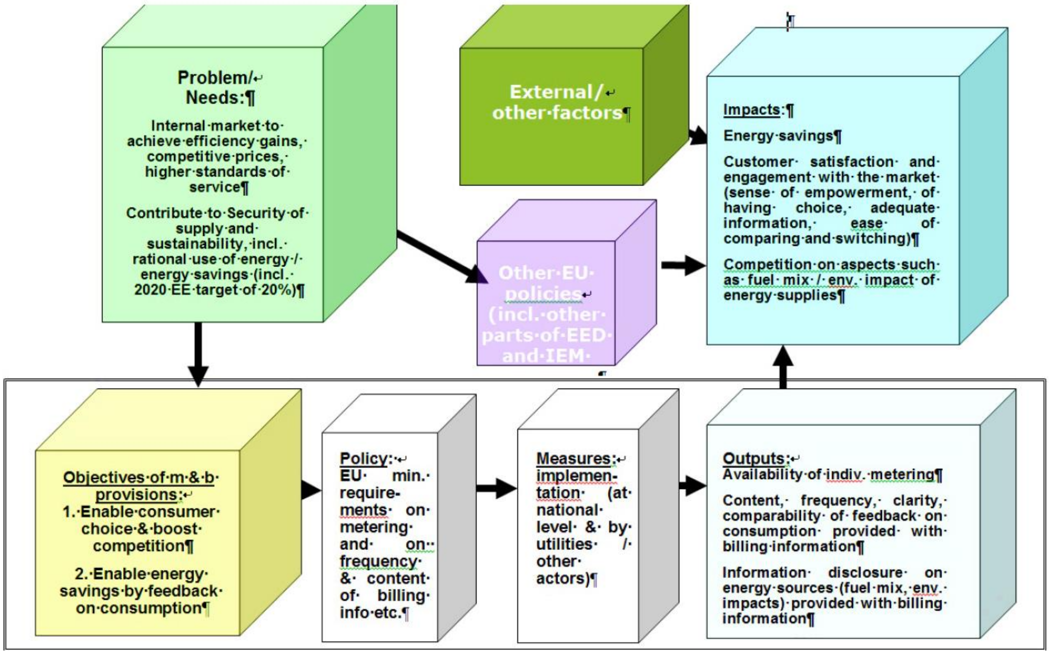


Figure 3. Intervention logic diagram of Commission's evaluation on the EU Framework for Metering and Billing of Energy Consumption (EC, 2016).

Related topics

How can a policy evaluation be designed
Counterfactual analysis
Toolbox for ex-post evaluation: methodologies for collecting evidence

Want to know more?

- (CEE, 2012) Theory-Based Approaches to Evaluation: Concepts and Practices, CEE Centre of Excellence for Evaluation – Government Canada, 2012, https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/theory-based-approaches-evaluation-concepts-practices.html#toc4
- (EC, 2016) SWD(2016)399 final - Commission staff working document – Evaluation of the EU Framework for Metering and Billing of Energy Consumption {COM(2016) 761 final} {COM(2016) 401 final}, European Commission, 2016, https://ec.europa.eu/info/sites/info/files/swd-2016-399-final_en_0.pdf
- (EC, 2017) Better Regulation: guidelines and Toolbox, European Commission, 2017, https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en
- (EEA, 2016) Environment and climate policy evaluation, EEA, 2016, https://www.eea.europa.eu/publications/environment-and-climate-policy-evaluation
- (GOV.UK, 2018) Guidance: Introduction to logic models, Public Health England – GOV.UK, 2018, https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/introduction-to-logic-models#references
- (HM Treasury, 2011) The Magenta Book, Guidance for evaluation, HM Treasury, 2011, https://www.gov.uk/government/publications/the-magenta-book
- (Tavistock, 2010) Guidance for transport impact evaluations - Choosing an evaluation approach to achieve better attribution, Tavistock Institute, 2010, https://www.tavinstitute.org/wp-content/uploads/2013/01/Tavistock_Report_Guidance_for_Transport_Evaluations_2010.pdf

2.2.2 Counterfactual analysis

Description

The counterfactual analysis measures what would have happened in the absence of a certain intervention, and the impact is estimated by comparing possible alternative (i.e. counterfactual) outcomes to those observed under the intervention. Observed actual development is a reference level, to which alternative possible developments are compared. Thus, counterfactual evaluation provides scenarios depicting possible alternative pasts (typically quantitatively) (ENVIEVAL, 2013).

How to use it?

There are well-established methods for estimating the counterfactual or baseline scenario(s), each with their own pros and cons. Some examples are given in the BOX below. Other illustrations for counterfactual analysis are in the selected case studies described in Chapter 6.

Example approaches to estimate the baseline scenario(s) (Europe Economics, 2016)

Example 1. Bottom-up approach: This approach uses granular estimates of energy uses or savings, (e.g. energy used by an electric vehicle or the change in energy use due to building insulation), and multiplies these granular estimates up to the total affected population. An advantage of this approach is that the individual components of the counterfactual are individually specified and estimated. The disadvantage is that some individual components may be difficult to estimate practically.

Example 2. Extrapolation of observed trends from the pre-intervention: An advantage of this approach is that it does not require finding a comparator unit that has not been influenced by intervention. However, a simple extrapolation of the past trends can fail to account for changes over time, e.g. due to another policy or technology.

Example 3. Comparison with similar (regional or national) markets where the intervention is not introduced: An advantage of this approach is that it takes into account other developments from the post-intervention, in cases where these developments are common to both markets. The difficultly is finding a suitable comparator, i.e. one which is likely to experience similar developments to the market in question.

In general, the next steps can be followed to **determine the effects of an intervention aiming to reduce GHG emissions** from different source/sink categories included in the evaluation scope (WRI, 2014):

(1) Estimate **baseline** emissions from each source/sink category:

The first step in applying the counterfactual analysis is to define the baseline scenario. For each source or sink category included in the GHG assessment, a baseline scenario is defined that represents the conditions most likely to occur in the absence of the intervention. The most likely counterfactual scenario depends on the drivers that would affect emissions in the absence of the intervention. Therefore, identifying key drivers and determining reasonable assumptions about their "most likely" values has an important impact on the evaluation results. Drivers that affect the emissions can be divided into two types:

- Other interventions than the policy being assessed that are expected to affect the emissions sources and sinks included in the assessment;
- Non-policy drivers or other conditions such as socioeconomic factors and market forces that are expected to affect the emissions sources and sinks.

In certain cases, multiple baseline options may seem equally likely. In such cases, a range of results based on multiple alternative baseline scenarios will be obtained. A **sensitivity analysis** is preferred to see how the results vary depending on the selection of scenarios. It is also important to identify potential **free rider effects** when identifying the baseline scenario. For example, the baseline scenario for an insulation subsidy should consider that a fraction of consumers receiving the subsidy may have installed the same insulation, even without the subsidy.

(2) Estimate **policy scenario** emissions for each source/sink category.

(3) For each source/sink category, subtract baseline emissions from policy scenario emissions to estimate the GHG effect of the policy or action for each source/ sink category.

(4) Aggregate GHG effects across all source/sink categories to estimate total GHG effect of the policy or action.

This way, the quantitative causal effects for a specific intervention can be determined against a counterfactual scenario in the absence of that intervention.

Related topics

Toolbox of ex-post evaluation: analytical methods
Rebound effect
Free rider effect
Uncertainty
Consistency between ex-post information and projections

Want to know more?

- (EC, 2017) Better Regulation: guidelines and Toolbox, European Commission, 2017, https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en
- (EEA, 2016) Environment and climate policy evaluation, EEA, 2016, https://www.eea.europa.eu/publications/environment-and-climate-policy-evaluation
- (ENVIEVAL, 2013) Strengths and weaknesses of Counterfactual Evaluation, ENVIEVAL & MTT Agri-food Research Finland, 2013, https://www.envieval.eu/fileadmin/envieval/dissemination/Events/First_international_stakholder_workshop/Strengths_and_weaknesses_of_counterfactual_evaluation.pdf
- (Europe Economics, 2016) Evaluation of Fiscal Measures in the National Policies and Methodologies to Implement Article 7 of the Energy Efficiency Directive, Europe Economics, 2016, https://ec.europa.eu/energy/sites/ener/files/documents/final_report_on_fiscal_measures_used_under_article_7_eed_0.pdf
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

3 Toolbox for ex-post evaluation: methodologies for collecting evidence

3.1 Summary

Whatever evaluation approach is taken, the accuracy of the results will depend on the quality of the evidence and related analysis. Important elements of high quality evidence are accuracy, verifiability, lack of bias, and data availability over the necessary time period of the ex-post evaluations, as emphasized by the European Commission's Better Regulation Toolbox (EC, 2017). This also states: *"Evaluations should be based on the best available evidence, which should be drawn from a diverse and appropriate range of methods and sources and combined (triangulation). This could be a combination of quantitative and qualitative methods. Not all sources of evidence are equally robust and consideration must be given as to when and how the evidence was collected and whether there is any bias or uncertainty in it. Where possible, sensitivity and/or scenario analysis should be conducted to help test the robustness of the analysis. Any limitations to the evidence used and the methodology applied, particularly in terms of their ability to support the conclusions, must be clearly explained."* In this Chapter **a toolbox of methodologies to collect evidence for ex-post evaluations** is described and illustrated by Member States' current practices.

The table below summarizes the strengths and weaknesses per methodology for a fixed set of characteristics (these are explained in Chapter 1), allowing a comparison of the methodologies. The ratings (Low-Medium-High) are explained in the corresponding methodology sections. In the summary table reference is also made to the case studies in Chapter 6 that illustrate common practice among Member States.

Methodologies	Short description	Data requirements	Complexity	Usefulness	Resources	Evaluation criteria	Communication/ Visualisation	Cases (See Chapter 6)
Surveys	Surveys are a common methodology for collecting large sets of data in a structured way from a sample. It is an effective way to gather new data, so typically used when there is a lack of existing primary data.	Low	Low	Medium	Low - Medium	High	High	case #1
Systematic literature review	A systematic review refers to a focused literature review that seeks to answer research question(s) using pre-defined eligibility criteria (inclusion/exclusion criteria) for documents and outlined and reproducible methods.	Low	Medium	Low - Medium	Medium	Low - High	Low - Medium	most cases
Focus groups	The focus group uses structured discussion that involves the progressive sharing and refinement of participants' views and ideas. It is well suited to cases where the views on evaluation topics are very divergent, but where discussion in groups may lead to a deeper viewpoint.	Low	Medium	Medium	Medium	High	Low	
Interviews	Interviews are used to collect qualitative information in a (semi-) structured way and to collect the opinions of persons affected by a particular intervention. They have the most added value in an exploratory context, often as complement to a survey.	Low	Medium	Medium	Medium - High	High	Low	case #1, case #4, case #5
Monitoring performance data	Monitoring data are regularly collected, quantitative data about a policy and can include data relating to each component of the intervention logic model and to each evaluation criteria.	Low	Low-Medium	High	Medium - High	High	High	case #4, case #6

3.2 Surveys

Description

Surveys are a common methodology for collecting large sets of data from a sample, which can help to answer one or more of the evaluation questions. They consist of a series of questions targeted at a group (or groups) of relevant stakeholders. Surveys are typically used to collect quantitative data, but it is also possible to collect qualitative data.

In Chapter 6, the survey case study #1 draws on two policy evaluations to demonstrate how surveys can be used to support the ex-post evaluation of climate policies. Both examples are from the agriculture sector and consider the actions taken by farmers, in response to a policy driver, to reduce GHG emissions at farm level. The two examples are explored in more detail in Chapter 6, but are also drawn on in the following guidance.

(1) The first case study comes from a review of the Greenhouse Gas Action Plan (GHGAP) in England. The GHGAP is a voluntary industry-led programme, and the principal mechanism for delivering reductions in emissions from agriculture in England (DEFRA, 2017).

(2) The second evaluation focuses on the EU Common Agricultural Policy (CAP), in the period 2013-2018 when the current form of CAP has been in force. It sought to understand the impact certain measures of the CAP have had on reducing GHG emissions, agriculture's vulnerability to climate change and its ability to provide adaptation and mitigation services to society (EC, 2018).

When to use it?

There are a number of circumstances in which a survey is a useful evaluation tool. First and foremost, surveys are an effective way to gather new data, so they are typically used when there is a **lack of existing primary data**. Surveys are particularly useful when data is held by many different stakeholders. In both case study examples there were a large number of implementing entities from which data was needed to estimate the emissions impacts.

As the survey questions are generally fixed for the duration of the survey, it is important that the data that is being collected can be clearly defined at the start. Designing the questions so that they are framed in the right way can be difficult without a deep understanding of the topic. It can therefore be useful to carry out some scoping research prior to implementing a survey e.g. some exploratory interviews (Public Health England, 2018). In addition to the scoping research, focus groups and pilot studies can be used to ensure the survey achieves its objectives and is understood by all participants.

Surveys can also be used to **validate or enhance other data sources**. For example, the surveys that were applied in the EU's CAP evaluation were used to test the conclusions found via other methods (e.g. literature review). In this way, surveys can triangulate results with literature reviews and interviews to provide greater confidence and clarity.

Representative coverage of the population and statistical significance is essential in case the survey is used as a single tool for ex-post evaluation of a specific policy. In other cases, a survey does not need to be statistically representative to be useful. Looking to the EU CAP example, the evaluation acknowledged that the surveys were never intended to provide statistically representative data, but they were still able to provide valuable insights. In the EU CAP example, the survey results were triangulated and were therefore not solely used to evaluate the policy directly.

Preparing, promoting and analyzing a survey can be time consuming. The size and complexity of the survey should therefore be balanced with the expected usefulness of results. It can therefore be advisable to conduct preliminary investigations based on stakeholder's knowledge before investing resources in a survey.

How to use it?

(1) Design

The survey design should reflect the specific objectives of the policy, the scope, scale and design of the policy instrument, and the characteristics of the target sector. It should therefore be informed by the intervention logic.

The survey will usually be designed to gather data which can help to answer one or more of the **evaluation questions**. For instance, when evaluating the greenhouse gas impacts of a policy, understanding how different sources of emissions have been affected by the policy, as well as by other drivers, is important. For the two agriculture example cases, this meant asking respondents about the different measures/actions that had been undertaken on the farms, while acknowledging that the actions may not necessarily have been taken with a primary aim of reducing emissions. Other drivers, such as cost reductions and the potential to increase production, were also recognised in the survey. These surveys were carefully designed to enable respondents to answer in a way that meant the data collected was relevant to the policy evaluation (see more on this in (2) Question types below).

Questions on the topic area are also typically accompanied by some more **general questions about the stakeholder**. These can collect demographic data (e.g. age, gender, employment, place of residence) or sector data (Public Health England, 2018). Examples of sector relevant data from the agriculture case studies include farm type (e.g. poultry, cereals, diary), size of holding (i.e. as defined by Standard Labour Requirements terms; the theoretical number of workers required each year to run a holding). This information is useful to understand if the policy impacts are more relevant for certain stakeholders groups, and can also be important when scaling results to a larger population as it explains how representative the survey sample is for the full population.

To determine the **sample size** needed for a survey the following parameters will need to be set:

- Population: Total number of entities affected by policy;
- Accuracy level: This concerns how accurate the results of the survey need to be, and includes both the margin of error for the results of the survey as well as a confidence level, which refers to how well the sample that you chose represents the full population. Often the margin of error is set between 1-10% and the confidence level between 90-99% (i.e. 90-99% of the time your results would be the same if you did another survey);
- Response rate: The percentage of people that will be reached out to that are expected to complete the survey.

Based on the above parameters the sample size required can be calculated using standard statistical methods[2]. The examples described in case study #1 show that in the two studies examined, different sampling approaches were used. The Farm Practices Survey (FPS) to study the GHGAP survey is designed to have a high confidence level (i.e. capture information that is representative of the full population) and is run in a similar way each year. The CAP surveys, on the other hand, were not designed to have a high confidence level. Instead, the evaluation notes that surveys were used to test that conclusions found in a literature review reflect the experiences of stakeholders. The surveys were also followed by interviews to better understand areas that need further clarification.

In addition, when designing the survey it is important to consider the effect of a sample selection bias. This refers to the situation whereby participants responding to a survey are more likely to have a stronger opinion about the survey topic than the average targeted stakeholder. This could result in

[2] (Statistics How To, 2020): https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/find-sample-size/

biased results. The strength of this effect may need to be evaluated via a small focus group first, before the survey is distributed among all participants.

(2) Question types

Both quantitative and qualitative data can be collected in a survey. Which data is collected depends on the design of the question and the needs of the survey. There are three main types of questions: closed (or pre-coded or multiple choice) questions, scalar questions and open questions. Examples of each of these are given in Table 4 below. These examples are taken from DEFRA's Farm Practices Survey (FPS).

- Both closed and scalar questions present pre-defined answers to the survey respondent. Answers to this question will therefore be quantifiable. That means these questions are useful when statistical analysis needs to be performed. Furthermore, the information gained from closed questions enables respondents to be easily grouped based on the options they have selected. The major drawback to closed questions is there must be a clear understanding of the topic of the questions and how it relates to the overall aims of the evaluation. Without this, the closed question's may be incorrect or incomplete, leading respondents to fill in the open text 'other' field. This increases the work for anyone analysing the survey later. There is a further issue with this as the question may not properly reflect the surveys purpose.
- Open ended questions are more exploratory in nature and provide qualitative data. They are therefore typically not used for statistical significance, but instead to explore issues that quantitative data alone cannot explain. Open ended questions require more effort to analyse the results as each response must be read manually. To limit the time required to analyse results, surveys with many open ended questions can be limited to smaller populations or be used in preliminary research on a small sample.

The FPS survey to study the GHGAP mainly used closed and scalar questions, to collect quantitative data, for two reasons. Firstly, the use of these question types increased the ease of completing the survey. Secondly, it allowed a more structured response, which made it more efficient to analyse the responses. Two surveys were run as part of the GHGAP: one targeted at farmers and the other at advisers. Several topics were common across the surveys and therefore a similar question structure and list of possible answers were used to enable comparison between the two sets of data.

When formulating questions, it is important to be aware the effect the framing of the question can have on the results. In other words, how a question is phrased or the choice of scale of closed ended questions, can affect the respondents' replies. For example, a stakeholder that is negative towards a policy might have incentives to underreport positive effects or overreport negative effects, if questions are aimed at these aspects. It may therefore be necessary to test the questions and their framing first with a smaller group to see if such a bias exists.

Table 4. Types of survey question (DEFRA, 2015).

Question type	Description	Example
Closed/pre-coded/ multiple choice question	Predefined multiple-choice options, appearing as a list. Care should be taken in the drafting of pre-coded responses. The codes must cover what are likely to be the most frequent survey responses. When using these questions, consider including an 'other' option. This will allow respondents to add their own answers if the list included does not include all options	What actions are you taking to reduce greenhouse gas emissions from your farm? Tick all that apply □ Improving energy efficiency (e.g. reducing electricity use, using reduced tillage) □ Recycling of waste materials from the farm (e.g. tyres, plastics) □ Improving nitrogen feed efficiency, livestock diets (e.g. using a ration formulation programme) □ Improving efficiency in manure and slurry management and application (e.g. controlled application rate, improved timing) □ Improving nitrogen fertiliser application accuracy (e.g. using a fertiliser recommendation system, regularly checking and calibrating fertiliser spreaders) □ Increasing use of legumes in arable rotation □ Increasing use of clover in grassland
Scalar questions	Multiple choice where the answer to the question is labelled either at the end points or at every point on the scale. Often includes a midpoint and the option to select 'don't know'.	How important do you feel it is to consider greenhouse gases (GHGs) when taking decisions about your land, crops and livestock? Please tick one box only □ Very important □ Fairly important □ Not very important □ Not at all important □ My farm does not produce GHGs
Open questions	Free-form survey questions that allows a respondent to answer in open text format. This means that responses are not limited to a set of options.	The FPS does not include any completely open questions. There are open fields in the form of the 'Other, please specify' to capture answers not included in the predefined closed/multiple choice questions.

(3) Timing: when to run survey?
Using a survey before or at the implementation of the policy can be used to construct the counterfactual scenario. A follow up survey can then be carried out once the policy has been in place for a suitable period (as part of the interim evaluation) or at the end of the policy (final evaluation), to determine the impacts of the policy ex-post.

The 2016 DEFRA review of the GHGAP was able to use data from the FPS conducted in 2012 as a counterfactual or baseline as the latter survey is conducted every year the policy is in place. Progress between 2012 and 2016 could therefore be observed (DEFRA, 2016). Whereas, the EC CAP review noted limitations in the ability to make comparisons to what was happening on the farm, before the measures were introduced. To enable such comparisons in future, it was recommended that an environmental component should be added to the Farm Structure Survey (i.e. an EU agriculture survey carried out every 3-4 years by Member States (Alliance Environment, 2018).

(4) Administration
There are numerous ways of administering surveys. The table below contains some examples, and the advantages and disadvantages of these methods.

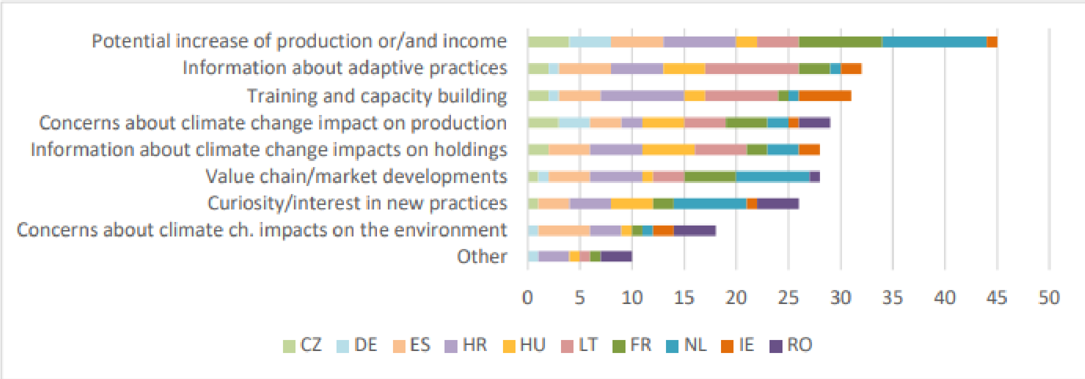
Table 5. Survey administration methods, advantages and disadvantages ((Public Health England, 2018).

Administration Method	Advantages	Disadvantages
In person	• Able to lead participant through appropriate sections, explaining what is required, if necessary • High response rate • Minimal errors or missing responses • Good for overcoming disability and language problems	• Expensive • Time consuming • Geographically limited • Problematic for sensitive topics • Possible interviewer bias (closed questions and multiple interviewers can minimise this bias)
Online	• Quick • Cheap to run • Wide geographical reach • Functionality can 'force' respondents to answer, reducing missing data • Can be completed anytime • Can ensure anonymity	• Response rate may be low • Requirement for internet access may exclude certain groups (e.g. those with sensory impairments or older people) • Requires technical knowledge to set up
Post	• Anonymous • Cheap	• Risk of very low response rates • Reminders or incentives may be required • Must be clearly designed and self-explanatory

(5) Presentation and analysis of results
Quantifiable data has another benefit, in that it is easy to present results in a graph or other visual form. While there will be many forms available, it is important that the presentation and interpretation of the data should link back to the objectives of the policy being evaluated. Examples of presenting survey results taken from the agriculture evaluations are presented below. These demonstrate that the results can be shown in various ways, depending on what it is important to communicate.

The data collected from surveys often allows for comparisons between groups and subgroups. Analysis of these sub-groups can reveal differences and reasons for these differences (Public Health England, 2018). The chart below from the EC evaluation shows the number of respondents who indicated each option, and splits the total respondents into the Member State they are from. From this it is easy to see that 'potential increase of production or/and income' was the key driver overall.

Figure 4. Presenting data example - EC evaluation of CAP survey results: 'Importance given by farm advisors and representatives to the main drivers of on-farm adaption (84 respondents)' (EC, 2018).



Source: survey of 84 farm and forest advisers and representatives in 10 Member States

The data can also be presented in other ways, which link the results of the survey to the broader aims of the policy. During the review of GHGAP, tables of quantitative findings were presented for each activity area of the policy (/survey). Where possible a comparison was made with 2012 data. Symbols were used to provide an indication of progress between the two surveys.

Additional activity area	**6. Progress on energy efficiency and renewables**	
Corresponding GHGAP activity	• A range of information published by the GHGAP partners including updates on energy efficiency, renewable energy technologies and Government incentives.	
Related survey data	• % farmers taking action to reduce GHG emissions who were improving energy efficiency as part of this (comparison with 2013).	79% ✓
	• % of farmers processing waste[16] by anaerobic digestion (comparison with 2012)	5% ✓
	• % of farm businesses currently undertaking energy generation practices[17] (Farm Business Survey 2011/12).	16% …
	• % of farm businesses intending to undertake additional energy generation practices[1] within the next 2 years. (Farm Business Survey 2011/12).	21% …
Data sources	Defra Farm Practices Survey, 2016, Defra Farm Business Survey, 2011/12.	

Indicator description	Indicator symbol
Little or no change	≈
Insufficient or no comparable data	…
Improvement	✓
Deterioration	×

[16] Slurries, crops, other feedstocks from on the holding, other feedstocks from outside the holding.

[17] Solar panels, Biomass fuelled heating boilers, exporting of slurry / grass / manure /miscanthus / biomass, Wind turbines, other renewable energy technology, anaerobic digestion, other.

Figure 5. Presenting data example - 'Progress on Energy Efficiency and Renewables from the FPS Survey' (DEFRA, 2016).

The governments can also create online dashboard tools to present the data online in a dynamic form. An example is the UK Office of national statistics 'Measures of Well-being' Dashboard[3]. Developing such tools can be time and resource intensive. It is therefore only worthwhile when a large amount of engagement is expected, and/or the survey will be repeated several times.

The analysis of quantitative survey results is often relatively simple. It can therefore be conducted in excel. For example, it is common to report the percentage of respondents who selected each option to a closed question. Excel can also be used to present the results of a survey in graph from. For more complicated statistical analysis, there are specific packages and tools which can also be used (e.g. Statistical Package for the Social Sciences (SPSS) or the statistical software 'R') which have complex

[3] https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuresofnationalwellbeingdashboard/2018-04-25

statistical calculations built in which can aid analysis. It should be noted that when a survey receives a very large number of qualitative question responses (e.g. several thousand responses), other types of analysis can be considered. For example, machine learning techniques can identify sentiments (positive or negative) or themes in respondents answers. This process can have limited applications, and may not pick up on nuances in results.

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low	Surveys are a data collection tool, they therefore have low data requirements. Sometimes complimentary data sources or statistics can be used to better understand survey results (e.g. Eurostat). Similarly, information collected before setting up a survey can improve its results, e.g. understanding of the target audience can shape the survey questions so more useful answers are collected.
Complexity	Low (– High)	Surveys can be simple to set up and analyse. However, designing the survey to ensure it provides the evidence requires more expertise. Free online tools and excel analysis will cover the needs of most evaluation surveys. If machine learning techniques are applied complexity of analysis increases.
Usefulness	Medium	Surveys are very useful for gathering evidence from a large number of stakeholders, although it is more difficult to check the accuracy of the data. However, they can be very helpful when using in combination with other data sources (such as interviews).
Resources	Low - medium	How much time and other resources are needed largely depends on the size of the survey and the type of questions included. A survey of mainly closed questions will be simple to analyse, and the resources needed is not dependent on the number of respondents. Open answers can increase the time required as responses will need to be read. It is important to balance the initial time required to set up the survey and the analysis, with the usefulness of the final results.
Evaluation criteria	High	Surveys are used for most evaluation criteria.
Communication / visualisation of results	High	Results (of quantitative questions) can be easily visualised in an appealing and easily understandable manner.

Tools

There are a number of free online survey tools available. The European Commission has its own free tool ‘EUSurvey’[4].

[4] https://ec.europa.eu/eusurvey/home/welcome

Related topics

- Intervention logic model
- Counterfactual analysis
- Systematic literature review
- Interviews
- Uncertainty

Case studies of relevant examples of the approach

- **Case study #1:** Evaluation study of the impact of the CAP on climate change and greenhouse gas emissions, European Commission (Directorate-General for Agriculture and Rural Development) and Alliance Environment (Mottershead, D.; Maréchal, A.; Allen, B.; Keenleyside,C; Lóránt, A.; Bowyer, C.; Brèche, O.: Martin, I.; Daydé, C.; Bresson, C.; Panarin, M.; Martineau, H; Wiltshire, J.; Menadue, H.; Vedrenne, M; Coulon, A), 2018, https://ec.europa.eu/agriculture/sites/agriculture/files/evaluation/market-and-income-reports/2019/cap-and-climate-evaluation-report_en.pdf
- **Case study #1:**The Greenhouse Gas Action Plan for Agriculture Review, Department for Environment, Food & Rural Affairs DEFRA, 2016, , https://www.gov.uk/government/publications/greenhouse-gas-action-plan-ghgap-2016-review

Want to know more?

- (Alliance Environment, 2018) Evaluation of the CAP Greening Measures, Alliance Environment and European Commission Agriculture, 2018, https://ec.europa.eu/agriculture/sites/agriculture/files/leaflet_en.pdf
- (DEFRA, 2015). FPS February 2015 - greenhouse gas mitigation - statistics notice, UK Government National Statistics and DEFRA, 2015, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/431938/fps-ghg2015-statsnotice-03june15.pdf
- (DEFRA, 2016) The Greenhouse Gas Action Plan for Agriculture Review, Department for Environment, Food & Rural Affairs DEFRA, 2016, , https://www.gov.uk/government/publications/greenhouse-gas-action-plan-ghgap-2016-review
- (DEFRA, 2017) The Greenhouse Gas Action Plan for Agriculture - Review 2016, Department for Environment Food & Rural Affairs DEFRA, 2017 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/599129/ghgap-2016-review.pdf
- (EC, 2018) Evaluation study of the impact of the CAP on climate change and greenhouse gas emissions, 2018, European Commission (Directorate-General for Agriculture and Rural Development) and Alliance Environment (Mottershead, D.; Maréchal, A.; Allen, B.; Keenleyside,C; Lóránt, A.; Bowyer, C.; Brèche, O.: Martin, I.; Daydé, C.; Bresson, C.; Panarin, M.; Martineau, H; Wiltshire, J.; Menadue, H.; Vedrenne, M; Coulon, A), https://ec.europa.eu/agriculture/sites/agriculture/files/evaluation/market-and-income-reports/2019/cap-and-climate-evaluation-report_en.pdf
- (GOV.UK, 2018) Guidance to Evaluation methods, Public Health England – GOV.UK, 2018, https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/evaluation-methods#references
- (HM Treasury, 2011) The Magenta Book, Guidance for evaluation, HM Treasury, 2011, https://www.gov.uk/government/publications/the-magenta-book
- (Public Health England, 2018) Guidance - Evaluation - Choosing Methods for Evaluation, UK GOV Public Health England, https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/evaluation-methods
- (Statistics How To, 2020) Sample Size in Statistics (How to Find it): Excel, Cochran's Formula, General Tips, Statistics How To, consulted in April 2020, https://www.statisticshowto.com/probability-and-statistics/find-sample-size/
- (World Bank, 2004) Monitoring & Evaluation: Some Tools, Methods & Approaches, World Bank, 2004, http://documents.worldbank.org/curated/en/829171468180901329/pdf/246140UPDATED01s1methods1approaches.pdf

3.3 Systematic literature review

Description

A systematic review refers to a focused literature review that seeks to answer research question(s) using pre-defined eligibility criteria (inclusion/exclusion criteria) for documents and outlined and reproducible methods. It incorporates an explicit layer of methodological systematization, to add transparency, objectivity and reproducibility to the review process. The process generally includes a number of formal methodological steps that can be followed to identify and analyse literature (Berrang-Ford et al, 2015). These steps serve to minimize bias and work toward consensus among stakeholders on the status of the evidence base from literature (Collaboration for Environmental Evidence, 2018).

Increasing interest in systematic approaches to synthesize qualitative literature has led to 'middle-ground' approaches which are often called meta-synthesis. Many reviews using a meta-synthesis approach include a description of methods for document selection, though apply a more iterative process less strictly defined than the formal eligibility criteria endorsed for many quantitative systematic reviews (Berrang-Ford et al, 2015). This way, they can provide a good balance between time & resources on the one hand, and the need for a robust approach on the other hand.

When to use it?

The systematized review process is designed to ensure that the selection of documents and information sources included is based on a set of clearly defensible criteria rather than ad hoc selection or being subject to undisclosed reviewer's bias (Berrang-Ford et al, 2015). This way, the reviewers try to gather a maximum number of the reported relevant bibliographic evidence in articles and studies while minimizing publication bias (i.e. higher probability that studies are published that report positive results) and selection bias (i.e. higher probability that studies are selected that are readily accessible, or are only published in major databases) (Collaboration for Environmental Evidence, 2018; HM Treasury, 2011).

How to use it?

As indicated in Chapter 2 – What is an evaluation? – it is important to first define an answerable, evaluation or research question and the scope of the study, as the synthesis of evidence is driven by the question(s) it is trying to answer (Collaboration for Environmental Evidence, 2018). Once the scope is defined, the following steps can be taken to allow a proper literature review:

(1) Involving stakeholders

Different people or organisations may perceive the evaluation question from different perspectives. Therefore, it is helpful to involve a broad range of stakeholders at certain stages of an evidence synthesis so that different users' viewpoints are considered, aiming to remove bias through a narrow focus. Some of the types of stakeholders that should be considered when planning a synthesis are: academics, government decision-makers (national, regional, local), intergovernmental decision makers, private sector (businesses, service providers), non-governmental or civil society organisations and general public (Collaboration for Environmental Evidence, 2018).

(2) Developing and testing a search strategy

Systematic and comprehensive searching for relevant studies is essential to minimise bias; therefore, the searching requires more planning and preparation than other steps. A good search strategy can make a substantial difference to the time and costs required for the synthesis (Collaboration for Environmental Evidence, 2018). A step-by-step overview of the search process for systematic review is illustrated in Figure 6.

(3) Selecting documents, including defining inclusion and exclusion criteria

The use of pre-specified and explicit eligibility criteria ensures that the inclusion or exclusion of articles or studies from a systematic review is done in a transparent manner, and as objectively as possible. The eligibility criteria for a systematic review should reflect the evaluation question and therefore follow from the 'key elements' that describe the question structure, so studies can be identified in searches that are relevant for answering the review question (Collaboration for Environmental Evidence, 2018).

(4) Extracting data or information from studies or articles

Data extraction refers to the process of systematically extracting relevant information from selected studies and articles. It includes recording of (Collaboration for Environmental Evidence, 2018):

- relevant characteristics (meta-data) such as when and where the study was conducted and by whom;
- aspects of the study design and conduct;
- results of the study (e.g. in terms of effect size means and variances).

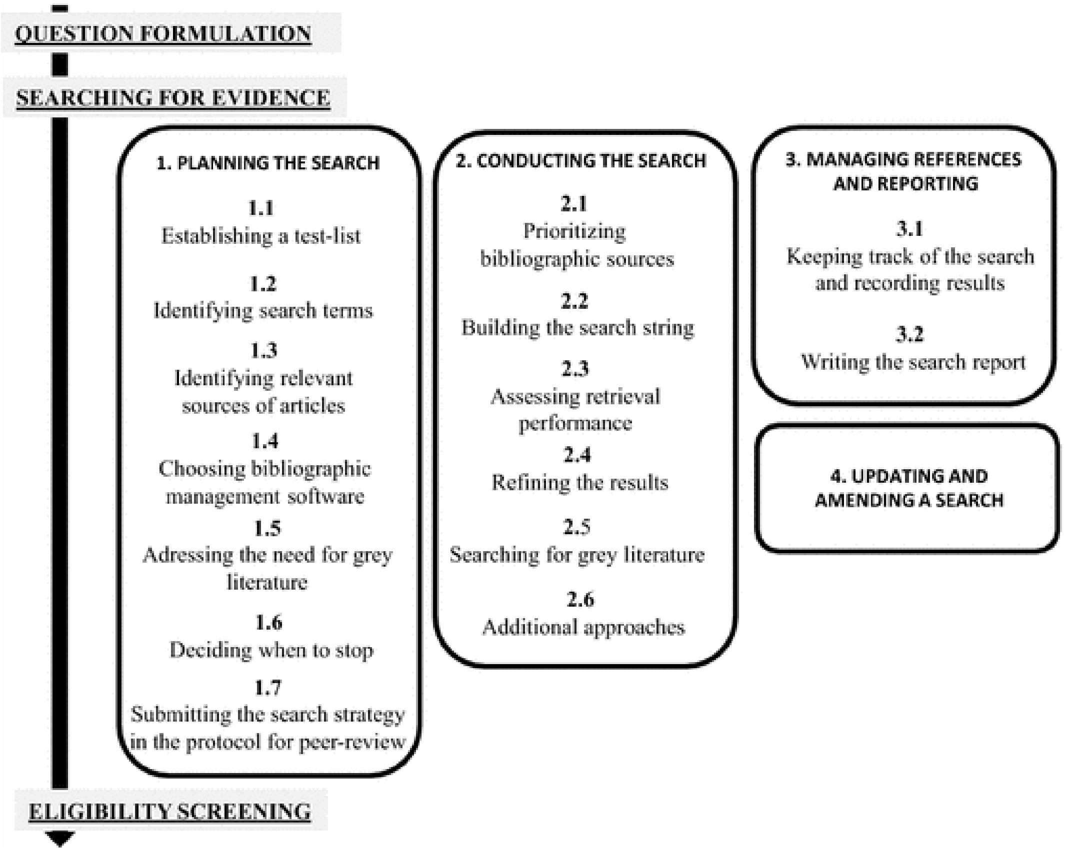


Figure 6. A guide to the planning, conduct, management and reporting of the searching step of systematic review (Collaboration for Environmental Evidence, 2018).

(5) Critical appraisal of study quality

During the critical appraisal, explicit and transparent criteria are used to judge the quality and strength of the identified studies, and hence the weight attached to their findings. Studies which do not meet sufficient quality standards can be rejected (HM Treasury, 2011). So, in short, critical appraisal is the process of assessing whether the evidence is valid for answering the evaluation question. Validity can cover "internal validity", which is the extent to which evidence is free from bias, and "external validity", which is the extent to which the evidence is relevant to the question being asked (Collaboration for Environmental Evidence, 2018).

(6) Reporting of conclusions based on data synthesis (quantitative and/or qualitative) and interpretation of findings

This step comprises the synthesis and interpretation of data in order to present reliable evidence in relation to the evaluation question. The strength of the evidence base and implications of the results for decision-making require careful consideration and interpretation. The analysis of data from the selected studies will depend on the policy question(s) being asked, the type of methodology used in the collected studies and how the findings will be used. Elements that may be reported in the conclusion section include (Collaboration for Environmental Evidence, 2018):

- quality/reliability of the included studies;
- relevance/external validity of the included studies;
- size and statistical significance of observed effects or impacts (in case of quantitative review);
- consistency of the effects across studies and the extent to which this can be explained by other variables (effect modifiers);
- clarity on the relationship between intensity of the intervention and outcome;
- lack of other plausible explanations of the observed effects.

In short, the review should be as clear as possible about what can and cannot be concluded from the existing evidence. It should identify any weaknesses or limitations in the existing evidence on the topic in question (HM Treasury, 2011).

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low	Data or evidence relies completely on relevant, available literature on the topic in question. Sources may involve electronic sources, print sources and “grey” literature.
Complexity	Medium	The complexity of the methodology is related to the multiple, formal steps that should be followed to minimise bias in the selected evidence. Therefore, the review is preferably undertaken by a multidisciplinary team because one person is unlikely to possess all skills required to conduct all stages of the review and because several review stages require independent checking. No specific software requirements are needed, although software for systematic review is available.
Usefulness	Low - Medium	Usefulness is strongly dependent on the evidence base from literature (i.e. availability of valid studies evaluating specific policies or measures in the Effort Sharing domain). As the scope from literature and evaluation concerned will very likely not coincide, conclusions from review won’t cover complete evaluation scope.
Resources	Medium	Review is undertaken by a multidisciplinary team. Stakeholders are preferably involved (stake in question formulation and findings of the synthesis). The search strategy or planning should include decision criteria defining when to stop the search given resource constraints. Care should be taken to lower the impact of resource constraints on the results.
Evaluation criteria	Low - High	Depending on the evidence base from literature, some or many evaluation criteria can be assessed: effectiveness, efficiency, coherence, relevance and other criteria.
Communication / visualisation of results	Low - Medium	Given the dependence of the analysis on the policy question(s), the type of methodology used in the collected studies and how the findings are used in the literature review, there might be a risk of misinterpreting the conclusions, as appropriate framing of findings should be clearly described (e.g. reliability, validity). Results or effects can be visualized, although appropriate framing is required as well.

Related topics

Toolbox for ex-post evaluation: methodologies for collection of evidence

Case studies of relevant examples of the approach

No specific case study using systematic literature review is included, as systematic review is often not applied in ex-post evaluations related to Effort Sharing policies & measures. However, in **most case studies described in Chapter 6**, a simplified literature review is included to either collect evidence or to get a better understanding of the background and context of the intervention.

Want to know more?

- (Berrang-Ford et al, 2015) Systematic review approaches for climate change adaptation research, Berrang-Ford L., Pearce T. & Ford J.D., Regional Environmental Change, 2015 15: 755, https://doi.org/10.1007/s10113-014-0708-7
- (Collaboration for Environmental Evidence, 2018) Guidelines and Standards for Evidence Synthesis in Environmental Management - VERSION 5.0, Collaboration for Environmental Evidence, 2018, https://www.environmentalevidence.org/information-for-authors
- (HM Treasury, 2011) The Magenta Book, Guidance for evaluation, HM Treasury, 2011, https://www.gov.uk/government/publications/the-magenta-book

3.4 Focus groups

Description

The focus group is a well-established method, taking the form of structured discussion that involves the progressive sharing and refinement of participants' views and ideas. The technique is particularly valuable for analyzing themes or fields which give rise to **divergent opinions or which involve complex issues** that need to be explored in depth (EVALSED, 2013). Rather than collect information individually, this method involves relatively homogenous groups of six to eight people. Each group meets once, during a physical or virtual meeting of around an hour and a half to two hours. To maximize the output, the discussion and the group interaction is facilitated by the evaluator or researcher who supplies the topics or questions for discussion.

When to use it?

Focus groups are well adapted in cases where the views on evaluation topics and issues are very divergent, but where discussion in groups may lead to a **deeper and more considered viewpoint**. Best practice is to work with different contrasting groups but internally homogeneous, so discussions bring out distinctive perspectives, experiences and views of each group in the evaluation.

Example of a focus group

EVALUATION OF THE WASTE REWARD AND RECOGNITION SCHEME: EMERGING FINDINGS (LYNDHURST, 2013)

The UK's Waste Reward and Recognition Scheme supports different schemes looking at innovative ways of tackling issues around food waste, recycling, re-use and waste prevention and reduction. It tests out how positive waste behaviour is affected through different kinds of reward and recognition schemes and what factors help or hinder such schemes in achieving this behavioural change. Eight schemes were assessed and tested and focus groups were used in several of these assessments. The Gloucestershire Waste Partnership worked closely with community groups to see if offering rewards and recognition would achieve higher recycling tonnages. The scheme was delivered by community groups, with the community rewards being funded by any increase in recycling credits ('Recycle for your Community Incentive Scheme' (CRIS)). For the evaluation of the recycling incentive scheme four focus groups were also used: one with residents aware of the scheme, one with residents unaware of the scheme, one with community organisations and one with the project team. A survey showed that a minority (19%) were aware of the CRIS. This was confirmed in the focus group discussions. Both the project team and community focus groups considered the scheme to be positive and useful experience even though they had not increased recycling tonnages. The focus group discussions further revealed that rewards were meant to be funded by recycling credits, but these rewards for the community groups were small in comparison to the communication budget provided and other funding streams available.

How to use it?

A critical element in focus group discussions is the role of the facilitator. She or he needs good communication skills, knowledge on the issues under discussion, and the capacity to question and challenge the group to achieve a more in-depth result. Considering these specific requirements, focus groups are mostly organized by policy evaluators.

The main steps are:

(1) Selection of participants:

It is advisable to form groups with a certain degree of homogeneity because of the objective to highlight where agreement exists within the group. Different groups representing different stakeholders or interest groups could be used. This means that the composition of the groups and the number of focus groups, depends on the particular requirements of the evaluation, which could range from one to several groups. Most guidelines on focus groups, advise to keep the groups relatively small, with the optimal number of participants of around six to eight per group. Special attention should be given to ensure high attendance (e.g. personal invitations, follow-up, etc.) and avoiding volunteer-bias (bias caused by the fact that people volunteering to focus groups or interviews might be different from the general population).

(2) Defining the topics:

Define and limit the topics to be discussed. A list of four to five open questions is usually sufficient. These can be tailored to the specific group. The aim is to use the questions or topics to engage the participants and to promote discussion.

(3) Course of the discussion:

The discussion may be launched fairly openly by introducing the subject of the session and asking a simple question of general interest. This will enable each participant to give an initial opinion or remark on the subject. As the discussion moves on, the aim is to clarify, delve deeper and to cover all angles. The facilitator's aim is to allow as much relevant discussion as possible to be generated from within the group, while at the same time ensuring that the topics and questions of interest to the evaluation are covered within the allotted time. This involves deciding when to move the discussion on to another topic, keeping the discussion relevant and focused, and choosing when to allow more free-ranging discussion with minimal intervention.

(4) Analysis of the results:

This final phase consists of interpreting and comparing the information given by the participants, and looking for shared and divergent opinions within each group. The collected information has to be reorganized, so the results can be analysed in relation to the objectives of the evaluation. The interpretation of data must take into account and distinguish two major aspects of the discussion: what the participants consider as interesting and what they judge as being important. The analysis will depend on the number of focus groups questioned, and on the nature of the interviews (for example, did the focus group discussion take a structured approach, or not?). The results from the different groups are compared to identify any convergence there may be. The report may quote the most noteworthy statements made by the participants, together with a summary of the discussion. The findings could be shared with the participants of the study to validate the results thereby increasing the credibility of the report or study.

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low	Focus groups are a data collection tool, they therefore have low data requirements. A good understanding of the intervention is needed to select the different groups and their participants for discussion. Sometimes complementary data sources or statistics can be used as a resource during the focus groups to facilitate the discussion.
Complexity	Medium	Focus groups require preparation to select the participants, arrange the group discussion and draft guiding questions. The outcome will also be more robust, when the facilitator has sufficient background knowledge on the subject.
Usefulness	Medium	Focus groups provide in-depth information on the opinions of selected participants. The group discussion can lead to a more balanced outcome, although there is a risk of a biased outcome, when a group discussion is not managed well.
Resources	Medium	It requires time for careful preparation and analysis of the discussions.
Evaluation criteria	High	Focus groups can be used for most evaluation criteria.
Communication / visualisation of results	Low	The qualitative results, including anecdotal evidence or opinions, from the focus group discussion can be more difficult to represent transparently and clearly in the evaluation. Results can be difficult to visualize.

Related topics

Surveys
Interviews

Case studies of relevant examples of the approach

No specific case study describing focus groups in detail is included, however the short example introduced above illustrates how MS can apply focus groups.

Want to know more?

- (EVALSED, 2013) Evalsed Sourcebook: Method and Techniques, European Commission, 2013, https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf
- (Lyndhurst, 2013) EV0530 Evaluation of the Waste Reward and Recognition Scheme: Emerging findings, BrookLyndhurst - a report to the Department for Environment, Food and Rural Affairs, Defra, 2013, http://randd.defra.gov.uk/Default.aspx?Module=More&Location=None&ProjectID=17989
- (BetterEvaluation, 2020) BetterEvaluation – Sharing information to improve evaluation, Retrieved April 2020, https://www.betterevaluation.org/en/evaluation-options/FocusGroups

3.5 Interviews

Description

Interviews are used to collect qualitative information and the opinions of persons affected by a particular intervention, its context, implementation and results. Several forms of interviews can be distinguished, each of which fulfils a different purpose: the structured interview (the most rigid form), the semi-structured interview and the informal interview. The **structured interview** resembles a survey, meaning mostly closed-ended questions. The **semi-structured interview** falls between the standardized, often closed-ended, surveys or interviews and the unstructured informal interview or focus groups (Adams, 2010). **Informal interviews** include only topic areas and themes and do not rely on standardized questions. They take the form of a more natural conversation between two people and allow the interviewer to pursue follow-up questions or new lines of discussion as they see fit. A question for example could be for the interviewee to identify the information they feel is most important for the discussion. If more people are involved, it resembles a focus group.

Interviews use a combination of closed and open-ended questions with the option of follow-up why or how questions. They are therefore useful for obtaining information on all aspects of an intervention's inputs, activities, outputs, results and impacts. It is a way of learning about and examining the views of different stakeholders on an intervention. Interviews are particularly valuable in exploring the ways in which an intervention has been implemented by the target group, identifying factors that have contributed positively or negatively to the result and the impact of the policy and for identifying good practices.

When to use it?

Interviews are time consuming to organize and to analyze the results, while the **number of respondents can be relatively small** compared to the number of potential respondents to an online survey, for example. But interviews do often have a clear added value, which depends on the circumstances and conditions of the evaluation. The interview has the most added value in an **exploratory context**, often as complement to a survey. It can also be a relevant technique when the number of stakeholders involved in the intervention is too small to be the subject of a statistically representative survey. The validity of the evidence provided in the interview is often determined by the context rather than the frequency or probability of occurrence (as in surveys).

How to use it?

Similar to focus groups, a critical element is the role of the facilitator in a semi-structured interview. She or he needs good communication, listening and note taking skills, as well as knowledge on the issues under discussion. An interview consists of four steps, which are (EVALSED, 2013):

(1) Selection of respondents

The number of respondents needed for interviews is smaller than for surveys. The reason is that the information obtained is validated by the context (e.g. level of expertise) and not by the probability of occurrence. The number of interviews depends on the topic, the variety of views, and on the resources available. Selecting respondents does not have to be done randomly, but could be based on purposive sampling. In this case, respondents are selected by the expert judgment of the evaluators starting from a variety of criteria such as specialist knowledge or capacity and willingness to participate in the research. It remains of importance however that the selected respondents reflect different views to avoid that the selection introduces a bias in the result.

(2) Planning the interview

This includes drafting guiding questions and an interview guide by specifying the topics that the interviewer wants to address to ensure completeness (i.e. all essential questions addressed). The added value of interviews is that new follow-up questions can be formulated during the interview to have more in-depth input and discussion. This requires sufficient expertise and knowledge of the interviewer, who also must have skills in communicating, listening and note taking.

(3) Course of the interview

The views of the respondents have to be recorded as accurately as possible. Where needed, interviews could also be kept anonymously.

(4) Analysis of the results

This final phase consists of analysing the interviews, interpreting and comparing the information given by the respondents, and finding common and divergent viewpoints so as to draw up a review of the evaluation (EVALSED, 2013).

Example of interview

EVALUATION OF MEASURES TO REDUCE FREIGHT TRANSPORT GHG EMISSIONS (TOURATIER-MULLER ET AL, 2019)

France has taken several measures to reduce greenhouse gas emissions of freight transport. This study investigated the environmental behaviour of small and medium sized enterprises (SMEs) among shippers and carriers in response to various transportation schemes introduced in France. Since 2008, carriers have been able to adhere voluntarily to a specific charter of commitments, the *Charte objectif CO2*. They can also obtain a compliance certification if, after three years, their environmental performance passes an audit led by an independent agency. By investigating the impact of both options, two main research questions are addressed:

- Are the governmental schemes effective, insofar that they have been implemented by SMEs?
- What are the main difficulties encountered and improvements that can be proposed to facilitate the implementation of these environmental protection schemes?

One of the research stages included the development of an interview guide. The interview guide was prepared to ensure that the same basic lines of inquiry are pursued with each person interviewed. The final interview guide included 26 questions covering six main themes: (1) Sustainable Development Strategy, with a target to reduce CO_2 emissions; (2) Knowledge of Decree 2011-1336 and its deployment; (3) CO_2 information processing; (4) Utilisation of the environmental data transmitted; (5) Shipper-Transporter Relationship; and (6) Implementation of an environmental collaboration, including CO_2 emissions reduction.

As case diversity was a crucial criterion, a total of 14 small- and medium-sized shippers and carriers were interviewed in different regions of France, as well as a transportation consultant.

During the coding process, which consisted of reading, analysing, and underlining key sentences, categories and sub-categories progressively emerged. This enabled the grouping of answers from different respondents. Furthermore, specific analyses for any pair of companies that worked together within a carrier–shipper relationship was done.

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low	Interviews are a data collection tool, they therefore have low data requirements. Sometimes complementary data sources or statistics can be used as a resource during the interview.
Complexity	Medium	Interviews require preparation to select the respondents, arrange the interviews and draft guiding questions. The outcome of interviews will be more insightful when the interviewer has sufficient background knowledge on the subject.
Usefulness	Medium	Interviews can be useful (especially in combination with a survey) as they allow follow-up questions, so more detailed information can be collected. Particular care has to be paid in the selection of respondents with different views.
Resources	Medium – high	An interview will take more time than a survey to collect evidence from a similar sized group. Not only because each individual has to be interviewed independently, but also because analysis will take more time. The use of the resources need to be weighed against the potential added value and evidence interviews give.
Evaluation criteria	High	Interviews can be used for most evaluation criteria.
Communication / visualisation of results	Low	In case of structured interviews, results can be visualised to some extent. The largest added value from interviews comes mostly from details that respondents provide, which are difficult to visualize. The qualitative results, including anecdotal evidence or opinions, from the semi-structured interviews can be more difficult to represent transparently and clearly in the evaluation.

Related topics

Surveys
Focus groups

Case studies of relevant examples of the approach

No specific case study focused on interviews solely is included, although **case study #1, #4 and #5** in Chapter 6 illustrate how interviews can be used as an integral part of ex-post evaluation.

Want to know more?

- (Adams, 2010) Conducting semi-structured interviews. In: Handbook of Practical Program Evaluation, Adams W., 2010, http://www.blancopeck.net/HandbookProgramEvaluation.pdf
- (EVALSED, 2013) Evalsed Sourcebook: Method and Techniques, European Commission, 2013, https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf
- (Touratier-Muller et al, 2019) Impact of French governmental policies to reduce freight transportation CO2 emissions on small- and medium-sized companies, Touratier-Muller N., Machat K., Jaussaud J., Journal of Cleaner Production Volume 215 Pages 721-729, 2019, https://www.sciencedirect.com/science/article/pii/S0959652619300605

3.6 Monitoring performance data and new data collection

Description

Whatever evaluation approach is used, quantitative data will form an integral and essential element to evaluate a policy. What data is required will depend on the types of evaluation proposed and the research questions to be answered. There are four main types of data which, if planned for, might be able to play a key role in supporting evaluations (HM Treasury, 2012):

(1) existing administrative data that has not been collected specifically for the evaluation;

(2) long term, large scale, often longitudinal, structural survey data that is often managed by central governments or the national statistics office;

(3) monitoring data or performance management data that are already being collected to support the administration of the policy; and

(4) new data collection needed to support the evaluations information needs.

As administrative and long-term data will, by their nature, be collected anyway this section focuses on (3) monitoring data (which in some cases will be a sub-set of general administrative data relevant to the operation of the policy or programme), and (4) new data collection.

The advantages of monitoring of an intervention are multiple. It establishes a historic data series that can be a valuable resource in any policy evaluation. It can form the basis of quantitative and statistical analysis, such as regression analysis, to establish the correlation and causality between changes in quantitative information and the policy. This section differs from surveys, focus groups and interviews in that the emphasis in this chapter is on quantitative data only, while surveys, focus groups and interviews collect mainly qualitative information. Surveys, interviews or focus groups are also often set-up specifically for an evaluation and not repeated at regular time intervals.

(1) Monitoring performance data

Monitoring data are **regularly collected information about a policy** and can include data relating to **each component of the intervention logic model** and to each evaluation criteria. As indicated in Table 6, it can consist of information on inputs that is needed to implement the policy, activities deployed, the outputs and the short- to long-term results and impacts (HM Treasury, 2012).

Table 6. Overview of types of data that could be monitored (HM Treasury, 2012), (WRI, 2014).

Data	Description	Use
Inputs	Resources that go into implementing a policy or action, such as financing	This can inform the evaluation of the efficiency of the intervention.
Activities	Administrative activities involved in implementing the policy or action, such as permitting, licensing, procurement, or compliance and enforcement	This can inform the evaluation of effectiveness and help determine whether the policy is being implemented correctly.
Outputs	Changes in behavior, technology, processes, or practices that result from the policy or action.	This can inform the evaluation of effectiveness by assessing whether the intervention has delivered the target outputs to the anticipated quality.
Results and Impacts	Changes in relevant conditions, such as energy consumption, or GHG emissions.	This can inform the evaluation of effectiveness by assessing whether the intervention has delivered the target results/impacts to the anticipated quality. This might not be monitored directly, but rather derived from monitored data in combination with additional information or assumptions.

While monitoring data are frequently administrative and are often not generated primarily for evaluation, it can be a very useful resource for evaluation purposes. The availability of this type of data, and whether there is any opportunity to adapt it in a way that best support the evaluation, can really strengthen the quality of an evaluation and should ideally be considered **at the design stage of the intervention**. In order to enable such a collection of performance data, requirements concerning collection and reporting might have to be imbedded in the relevant legislation. This would ensure the collection, aggregation and reporting of official data. Care should also be taken to establish good quality of the monitoring data being collected as poor or partial data will affect the scope and scale of monitoring data's contribution to an evaluation (HM Treasury, 2012).

(2) New data collection

Where monitoring data is not feasible or appropriate, bespoke research can be used to collect quantitative evaluation data. The data collection is not intended to recreate time series, but to be used in specific quantitative (statistical) methods, such as difference-in-difference analysis. In order for the results to be sufficiently robust, the data will have to be collected in **a standardized manner, with sufficient sample size**, and, where possible, from a treatment and control group to allow for comparison against the counterfactual. The type of quantitative research and data collection is very context-dependent and driven by the evaluation criteria and the evaluation question(s). New data collection can be done using surveys, asking for quantitative information (see section Surveys), or metering and measuring. In metering and measuring, activity data are recorded via technology or via billing or audits (MultEE, 2016).

Table 7. Examples of data that could be monitored.

Data	Examples
Input	Monitoring the budget or other resources allocated to the implementation of the policy. For example: • Funds allocated per year to operationalize the policy • Funds allocated per type of technology • Number of staff with necessary expertise to implement the policy • Enforcement costs • Number of new funds created to support the development of renewable energy
Activity	Monitoring actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilised to produce specific outputs. For example: • Number of projects funded by subsidy scheme • Number of visits to check regulatory compliance
Output	Monitoring the outputs directly realised by the policy. These are the first steps towards realising the operational objectives of the intervention and are measured in physical or monetary units. For example: • Kilometers of new railroad tracks • Number of transmission lines built to import power and facilitate grid integration • Number of developers receiving project development assistance • Number of renewable energy certificates issued • Supported agricultural land area
Result and Impact	Monitoring the direct and immediate effects of the intervention (results) and the benefits beyond the immediate effects on its direct beneficiaries both at the level of the intervention, but also more generally in the policy area (impact). For example: • Jobs created in supported projects • Total investments for energy efficiency • Improvements in the average new vehicle fuel consumption • Production of renewable energy • Percentage of agricultural land under management contracts to improve soil management • Share of agricultural area under greening practices • Greenhouse gas emission trends in a (sub)sector

When to use it?

(1) Monitoring performance data

In theory, monitoring needs can be made endlessly great, so that compromises have to be made between the needs for monitoring and what is practically possible to do. Some important challenges have to be taken into account while defining the monitoring approach (AID-EE, 2006):

- Is monitoring justifiable for the particular policy or measure? Depending on the **importance of the policy and measure** under consideration, more or less resource could be dedicated to monitoring. The choice of monitoring should be based on a conscious decision.
- How much **time and resources** should be spent on monitoring? How important is monitoring regarded in the policy process? Often monitoring activities have a low priority in the design and implementation of policy instruments. However, especially for policies and measures with an expected high impact or that have a high cost, sufficient resources should be foreseen for proper monitoring.
- Which aspects are most important to monitor to give reliable and consistent measurement against the intervention's objectives? This depends on **what type of information** is needed to monitor the

progress of the policy and what is needed for its evaluation at a later stage. It will be difficult or impossible to recreate missing information afterwards, although some of the gaps might be filled by new data collection to give quantitative information over a shorter time period.

- There are effects of policy instruments that are very hard or **impossible to monitor**. In particular, long term effects, that occur years after the policy instrument is terminated, cannot be monitored effectively. In designing monitoring schemes, a pragmatic selection of data that can be easily collected and is relevant has to be made.
- How to identify and adequately deal with **uncertainties**? It is also important to identify and take into account to what extent monitored data are uncertain. One element to reduce uncertainty is to include quality control measures

(2) New data collection

Where monitoring data is not feasible, appropriate or available, new data collection could be an alternative or even complementary approach to monitoring. Some quantitative data can be difficult or time/resource consuming to collect in a monitoring scheme, but could still be very relevant in the context of policy evaluation. This applies in particular for information and data that are more linked to the results and the impacts of the policy and measure. Bespoke quantitative research, for example a difference-in-difference analysis, can in this context be used to fill these information gaps and to assess the impact. New data collection will in this case be needed to collect the necessary quantitative information.

How to use?

(1) Monitoring performance data

As the monitoring data have to be of high quality and consistent in time, monitoring therefore requires a **high level of planning**. Ideally data collection is already started before the implementation of the intervention to establish the baseline. Unlike most policy evaluations, it is not a one-off exercise. Resources and time are however in most cases scarce when it comes to monitoring activities. For some policy instruments it might be worth to ask the question if it is reasonable to dedicate resources to monitoring activities. This could be the case if the instrument is very small, if it is primarily meant as a supporting instrument or if the outcomes of the instrument are very hard to measure (AID-EE, 2006). This means that compromises have to be made between the needs for monitoring and what is practically possible to do, as not all relevant information can be monitored. **A selection has to be made of the most important indicators** for evaluation purposes, which could be based on:

- Already available data and statistics: map existing data and statistics that are available and can be used to evaluate an intervention. For example, the combination of monitored data with governmental statistics allows the calculation of relative indicators that could be used in indicator analysis. Therefore, **make maximum use of existing data to save time and increase coherence**.
- The aspects of the policy being most important to monitor: based on the intervention logic, is it most relevant to monitor inputs, activities, outputs, results and impacts? **Collect only what is relevant.**
- The resources that can be spent on monitoring: weigh resources needed for monitoring activities against the design and implementation of the policy instrument. **Automate as much as possible to shorten data collection and processing time.**
- The availability and quality of data: some data are more difficult to monitor and are more uncertain than others. Uncertainty of the monitored data will also mean uncertainty of the end result of the evaluation. **Use common reporting standards.**

See Figure 7 (HM Treasury, 2012) for a schematic overview of key considerations in setting up a monitoring system.

Figure 7. Key consideration in setting-up a monitoring system (HM Treasury, 2011).

What data needs to be gathered to give reliable and consistent measurement at each stage of the Theory of Change (and if necessary evaluate impact)?

1. Identify key indicators that can be clearly defined and used to monitor progress against the Theory of Change. Targets and target dates can be set but are not essential.
2. Identify the data required to measure inputs, outputs, outcomes and impacts.
3. Identify data needed to create a comparison group (if needed for an impact evaluation).
4. Identify whether this data is already available.

↓

What new data should be collected to fill gaps in measurement?

The following data should be considered:

1. Stakeholder perceptions/attitudes towards the intervention/behavioural change
2. Financial data relating to the intervention's expenditure
3. Process data to assess whether the intervention has been implemented as intended
4. Data to track the outcomes and impact of an intervention

↓

Who will have responsibility for gathering the data?

1. Identify the most appropriate individuals to gather the data, e.g. programme/project delivery team, an existing performance monitoring team, evaluators, etc.
2. What resources are required? Do those responsible have the time and skills?

↓

When will the data be gathered?

1. How often should administrative/monitoring data be gathered? (e.g. monthly, quarterly, annually). Can this be aligned with the auditing/reviewing process of the funding body?
2. What is the timetable for the collection of new data and is this aligned with the reporting schedule for the evaluation?

↓

How will the data be gathered and stored?

1. What format should the system use? Can it be aligned with existing monitoring systems?
2. Data protection protocols are needed to meet security and data sharing requirements.
3. Likewise, ethical considerations need to be taken into account (e.g. informed consent).
4. Where will the data be stored?

↓

How will the data be verified to ensure accuracy and consistency with requirements?

1. Who are the most appropriate individuals to verify the data, e.g. analyst, programme/project lead at the funding body, independent evaluators, etc.?
2. What resources are required to undertake the task?

↓

Design and implement the monitoring system

(2) New data collection

There is a large interaction between new data collection, the type of quantitative research and methodologies used and the evaluation criteria and question(s). Two examples of data collection are explained in the BOX below: (1) collection of metered data on gas consumption from building owners in the Netherlands, and (2) a quasi-experimental approach to measure energy savings of public sector buildings in the UK.

When designing and implementing data collection there are some key challenges that need to be taken into account.

- Data has **to be collected in a standardized way** within the sample (e.g. metered data over the same period of time) and across years. High standardization of data collection reduces the uncertainty of the outcome so that the results reflect reality more accurately.
- In some cases the entire population cannot be sampled and data collection can only be done on a subset. In such cases the **sample size** (e.g. the number of households, companies, car owners, etc.) needs to be sufficiently large to ensure that the results are representative for the entire group. Small sample sizes increase uncertainty. However, even with larger sample sizes, confounding factors could have an effect on the result. To make quantitative results more robust, **repetitive sampling over time and use of control groups,** both in monitoring and new data collection, can help to correct for these confounding factors.
- The selection also has to be done unbiased and completely random. Sampling or **selection bias** can be introduced unintentionally. For example, voluntary reporting increases the risk that respondents are more interested and motivated than non-respondents.

Examples of new data collection

(1) COLLECTION OF METERED DATA ON GAS CONSUMPTION FROM BUILDING OWNERS IN THE NETHERLANDS

ECN (2010) evaluated the impact of a tightened building code for new houses in 2006, changing the energy performance coefficient (EPC) from 1.0 to 0.8. The evaluation focused on 4 key questions:

- What are the experiences of companies involved in construction? (Interviews)
- What was the effect on energy consumption? (New data collection)
- Was there a measurable effect on the health of occupants? (Survey)
- Did it increase the construction cost of a new house? (Cost calculation)

For each of these questions, different methods were applied as indicated in brackets above. Here we only focus on the second question, as it included new data collection. For the data collection, all new houses completed between June 2006 and January 2008 were considered (excluding houses connected to heat networks). An invitation letter to contribute to the study and data collection was sent to approximately 15,000 houses, of which 1,076 owners responded positively and were selected. Contribution to the data collection consisted of reporting metered data on gas consumption at regular time intervals via an online platform. This was complemented with a questionnaire to collect essential information on the characteristics of the buildings (e.g. number of rooms, additional heating systems such as solar heating or wood burners), and its occupants and their behavior (e.g. number of people in the family, occupancy of the house, average daily indoor temperature).This additional data collection is essential as it allowed correcting for these variables that have an important and significant effect on gas consumption when comparing gas consumption in houses complying to the EPC 1.0 with the houses complying to the EPC 0.8.

The final conclusion of the study was that the annual standardized gas consumption of houses with an EPC of 0.8 was lower than houses with an EPC of 1. These differences are however not statistically significant. In addition, the differences between average gas consumption were lower than were expected from theoretical engineering calculations.

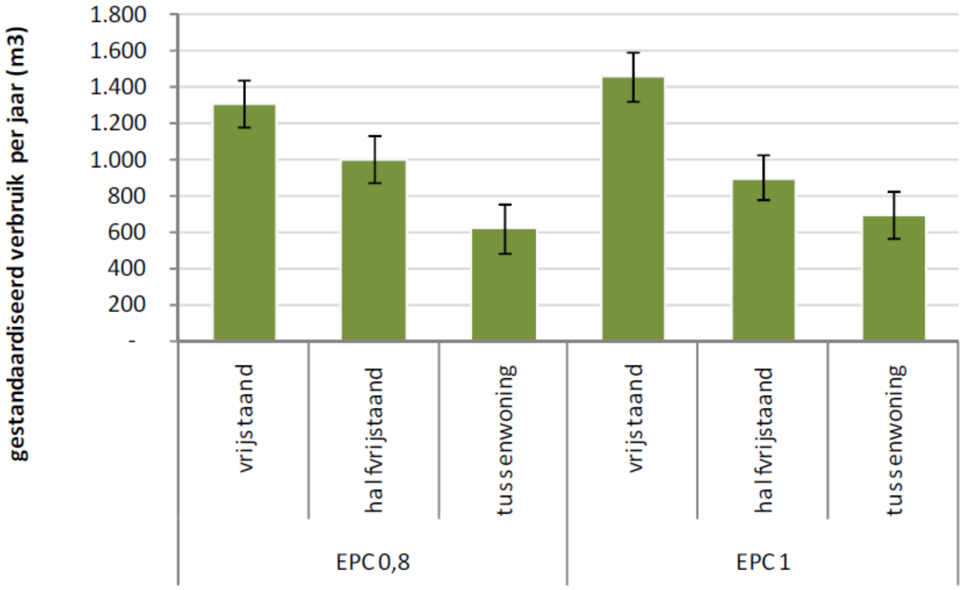


Figure 8. Average annual standardized gas consumption in detached, semi-detached and terraced houses with EPC of 0.8 and 1 (ECN, 2010).

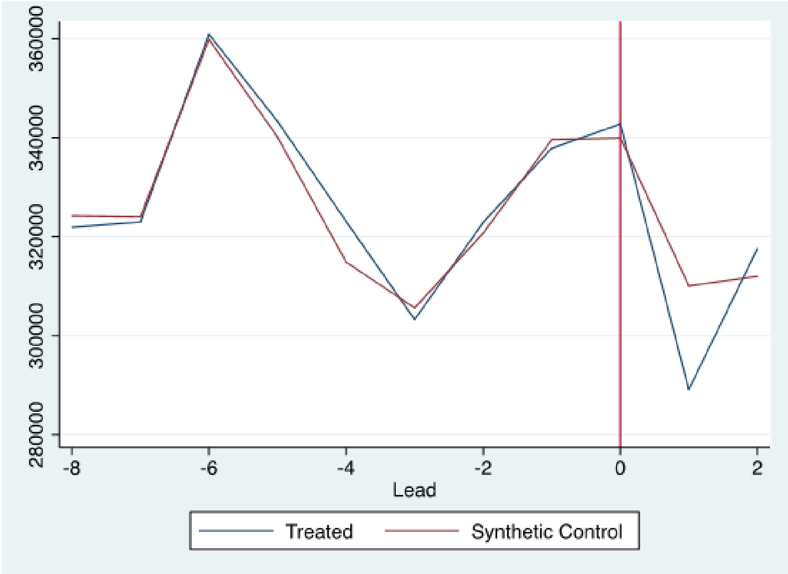
(2): QUASI-EXPERIMENTAL APPROACH TO MEASURE ENERGY SAVINGS OF PUBLIC SECTOR BUILDINGS IN THE UK:

A comparison of a treatment and control group was done for the evaluation of the Public Sector Energy Efficiency Loan Scheme of the UK (BEIS, 2018). This scheme provides interest free loans to public sector bodies to support the installation of energy efficiency measures, thereby reducing energy consumption, greenhouse gas emissions and energy bills. The loans are available to local authorities, National Health System / Foundation Trusts, schools (including academies), further and higher education institutions and provides two funding models: the Salix Energy Efficiency Loans Scheme and the Recycling Fund.

Monitoring activities to measure actual savings on site appeared to be variable, with few organisations undertaking concerted attempts and limited to large organisations (with experienced energy management teams). Reasons for not monitoring included cost (relative to benefit), hassle and challenges associated with working with old systems across complex portfolios of buildings. Several respondents also noted that there were too many other potential factors influencing energy use to enable reliable isolation of the impact of any given project. Monitored data thus did not provide additional evidence on the effectiveness of the intervention.

Therefore a quasi-experimental approach was used for three types of projects implemented in primary schools: lighting projects, insulation projects, and all other projects affecting natural gas consumption. Schools were selected for this pilot as the meter data required for the analysis were readily available for the buildings where projects were implemented. The analysis explores the distribution of the changes in energy consumption between a school implementing a project through the scheme and its synthesized control unit. This control unit represents, as close as possible, a school with similar characteristics, but without the implemented energy efficiency improvements.

Figure 9. Natural gas use in the average treated unit and related synthetic control (BEIS, 2018).



Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low	Setting up monitoring or new data collection does not require much data. It requires a good understanding of the intervention's logic model, so relevant indicators can be selected for monitoring.
Complexity	Low - Medium	Complexity depends on the type of method that is used to collect data, which can range from relatively simple indicators or statistics that might be easy to collect to more complex measurements, for example in-situ measurements of energy savings.
Usefulness	High	Depending on the type of monitoring indicators, it can be hard to link the results directly to a particular policy or measure. But their usefulness can be improved in combination with other data sources (e.g. surveys and interviews).
Resources	Medium - High	Depending on the complexity of the selected indicator, monitoring might require more resources.
Evaluation criteria	High	Monitoring/data collection can be linked to several evaluation criteria, although mostly to the effectiveness and efficiency of an intervention.
Communication / visualisation of results	High	The collected data can be easily visualised in an easily understandable manner.

Related topics

Intervention logic model
Counterfactual analysis
Toolbox for ex-post evaluation: methodologies for collection of evidence
Uncertainty

Case studies of relevant examples of the approach

No specific case study using monitoring performance data or new data collection is included, although in **case study #4 to case study #6** described in Chapter 6, it is illustrated how monitored performance data can be used as an integral part of ex-post evaluation.

Want to know more?

- (AID-EE, 2006) Guidelines for the monitoring, evaluation and design of energy efficiency policies - How policy theory can guide monitoring & evaluation efforts and support the design of SMART policies, AID-EE Intelligent Europe, 2006, https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/aid-ee_guidelines_en.pdf
- (BEIS, 2018) Evaluation of the public sector energy efficiency loan scheme, BEIS, 2018, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730976/Public-sector-energy-efficiency_loan_scheme_evaluation_Interim_Report_Final.pdf
- (ECN, 2010) Evaluatie EPC-aanscherping woningen (in Dutch), ECN? 2010, https://publicaties.ecn.nl/PdfFetch.aspx?nr=ECN-E--10-043
- (HM Treasury, 2011) The Magenta Book, Guidance for evaluation, HM Treasury, 2011, https://www.gov.uk/government/publications/the-magenta-book
- (Sing N., Vieweg M., 2015) Monitoring implementation and effects of GHG mitigation policies: steps to develop performance indicators. WRI working paper, Sing N., Vieweg M., 2015, https://www.wri.org/publications/performanceindicators
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf
- (EC, 2015) Technical handbook on the monitoring and evaluation framework of the Common Agricultural Policy 2014–2020 https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=21095&no=3
- (EC, 2017) Better Regulations Toolbox #41. https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-41_en_0.pdf
- (MultEE, 2016) Data Collection Process for Bottom-up Monitoring. https://multee.eu/system/files/D2.3_Data_collection_process_for_bottom-up_monitoring_online_0.pdf

4 Toolbox of ex-post evaluation: analytical methodologies

4.1 Summary

Analytical methodologies make it possible to isolate the effects of an intervention, such as greenhouse gas mitigation effects, in a complex environment. These methodologies are, however, very data-sensitive and require a large quantity of reliable information to allow analysis of the cause – effect relationships of policies and measures. In this chapter a toolbox of quantitative analytical methodologies for ex-post evaluations is described and illustrated by Member States' current practices: **How can the impact, costs and or benefits of an intervention be estimated?**

The following table summarizes the strengths and weaknesses per methodology for a fixed set of characteristics (these are explained in Chapter 1), allowing a comparison of the methodologies. The ratings (Low-Medium-High) are explained in the corresponding methodology sections. In the summary table reference is also made to the case studies in Chapter 6 that illustrate common practice among Member States.

Methodologies	Short description	Data requirements	Complexity	Usefulness	Resources	Evaluation criteria	Communication/ Visualisation	Cases (See Chapter 6)
Indicator	Indicator analysis uses single or multiple indicators to track progress towards meeting an objective or obtaining a certain effect. It requires quantitative data that usually comes from a monitoring framework and represents a time series and/or different groups.	Medium – High	Low - Medium	Medium-High	Medium	High	High	case #5
Cost effectiveness	CEA will appraise a policy in terms of effectiveness in achieving single desired outcome for given level of cost, relative to its counterfactual. It is most useful when the objectives are clearly identified, connected to a quantified target and a clear baseline.	Medium	Low - Medium	Medium	Low - Medium	Medium	Medium	case #4 case #5
Cost benefit	CBA assesses whether a policy or measure is worth implementing (i.e. whether benefits outweigh costs) from a societal perspective. The major difficulty with CBA is to monetize all costs and benefits.	High	Low - Medium	High	High	Medium	Medium	case #6
Regression	This statistical method aims to investigate the relationship between two or more variables. It is a useful method to estimate impacts of a policy if (a) the policy effect can be characterised by a specific variable and; (b) good data is available on the trends for this variable as well as for the other variables.	High	High	Medium	Medium	Medium	Medium	
Decomposition	Decomposition analysis can be used when one wants to quantify how various key drivers influence GHG emissions. These drivers do not directly depict policies and measures. The effects of policies will be indirectly visible through the changes in the drivers.	Medium – High	Medium	Low – Medium	Low - Medium	Low	High	case #2 case #3
Multi-criteria	In case where policy options may have different environmental and social impacts that are measured with different units. An MCA can provide a method for comparing different indicators and ranking the options while providing a transparent rationale for evaluation.	Low – Medium	Low	Low	Low	Medium	High	case #8

4.2 Indicator analysis

Description

Indicator analysis can be used as evidence to evaluate multiple evaluation criteria, although it will be most often used to evaluate effectiveness and assess the impact of a certain intervention. The analysis uses single or multiple indicators to track progress towards meeting an objective, mobilisation of a resource, obtaining a certain effect, as a gauge of quality, etc.

When to use it?

Indicator analysis can be used to evaluate the effectiveness, efficiency, relevance and coherence of a policy. It requires quantitative data that usually comes from a monitoring framework, either pre-existing or specifically set-up for evaluation purposes. For meaningful analysis of the indicators, data from a time series and/or different groups (e.g. for difference-in-difference analysis[5]) is needed.

How to use it?

(1) Selection of indicators

The selection of relevant indicators is very dependent on the instrument type, the sector and other characteristics of the policy or measure under evaluation. It is therefore not possible to give very concrete predefined indicators. Instead this report focuses on guidelines of what to consider when selecting indicators to evaluate policies or measures. There are several frameworks that could be used to characterize and select indicators, such as the DPSIR framework (driver, pressure, source, impact and response). For these guidelines, the differentiation will be made based on the **intervention logic**. The intervention logic can be used to map the expected inputs, outputs, results and impacts of the intervention. It can thus also inform evaluators of which indicators to analyse.

Different type of indicators can be identified:

- **Input indicators**: measure the input that is necessary to enable the intervention to be implemented. Input indicators can be used to assess the efficiency of a policy or measure. Examples are funding, available staff and infrastructure;
- **Output indicators**: represent the direct product of the intervention. Output can often be more easily monitored than result and impact indicators as these tend to be also under the responsibility of the implementing entity. Examples are number of subsidies, kilometers of rail built, number of energy audits performed, etc.;
- **Result and Impact indicators**: represent indicators of the change that the intervention wants to achieve over the short, medium and long term. The indicators could include both the direct, indirect or unintended effects of an intervention. An example is outlined below in the BOX.

Example of output indicators

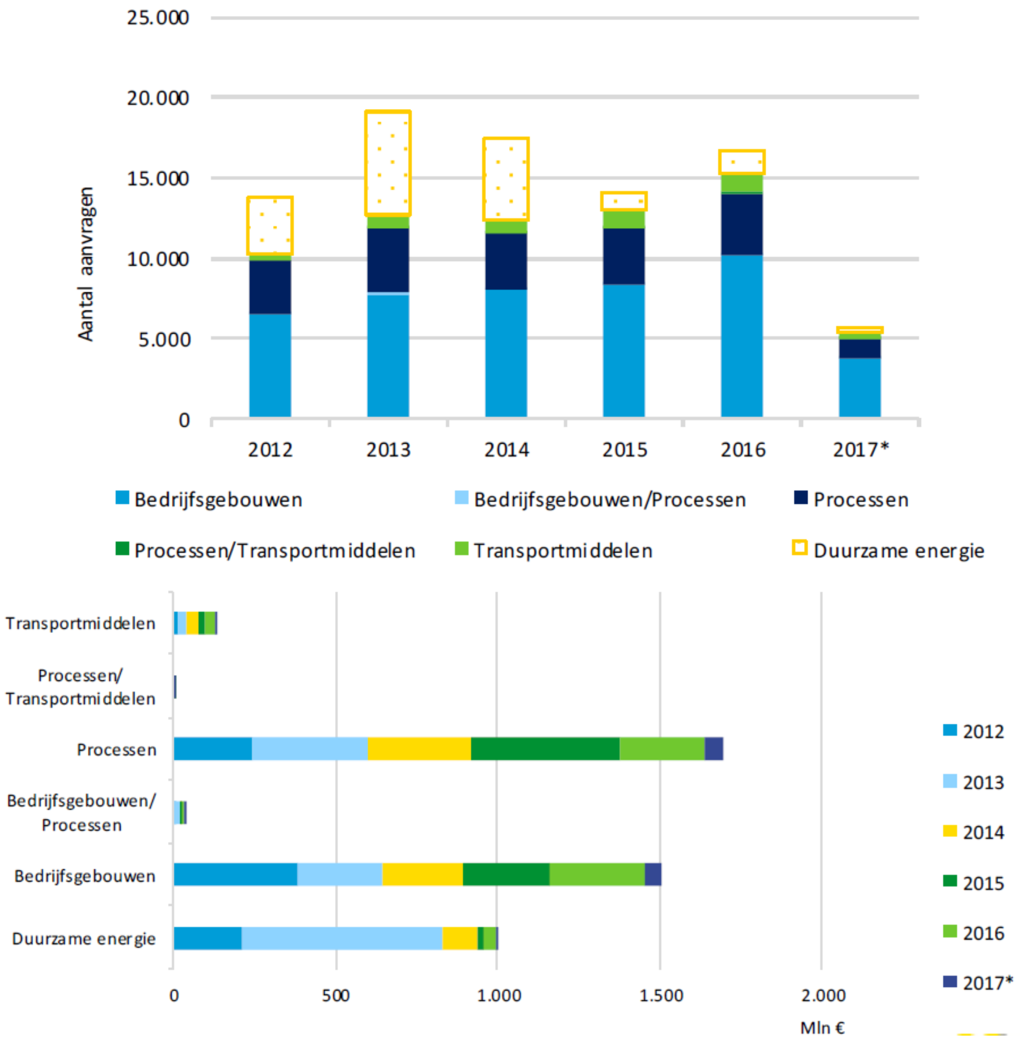
EVALUATION OF THE ENERGY INVESTMENT ALLOWANCE IN THE NETHERLANDS (CE DELFT, 2018)
The Energy Investment Allowance is a policy aimed to improve energy efficiency in companies and industry. It is a fiscal instrument, allowing applicants to subtract investment costs from taxes. Eligible technologies are listed by the government, allowing only innovative technologies to increase energy efficiency. There is a mandatory evaluation of the policy every 5 years.

[5] Difference-in-differences method compares the changes in outcomes over time between a population enrolled in a programme (the treatment group) and a population not enrolled (the comparison group).

Different types of evidence are used to evaluate this policy: literature review, data and indicator analysis, stakeholder interviews and case studies. The evaluation is focused on effectiveness and efficiency of the policy.

With respect to the data and indicator analysis, substantial attention is given to output indicators: for example, the number of applications, number of applications with positive decision, investments in million EUR split per year and type of investment and technology. This provides a valuable data source to track or monitor the output of the policy over different years in a detailed manner. It is also a starting point for the impact assessment of the policy on gross and net energy savings.

Figure 10. Number of applications for an Energy Investment Allowances (top) and investments (in mEUR) (bottom) (CE Delft, 2018).



Often several indicators can be selected to evaluate a policy or measure. It should be noted that selecting more indicators is not necessarily better. Some indicators can be more time consuming (and costly) to collect and analyse or have a higher degree of uncertainty (see section 'Uncertainty'). When selecting indicators, consider the data and data sources that are already available (see section 'Monitoring'). In selecting indicators, different **conceptual approaches** can be used to ensure that the data from the indicator results in useful information, for example:

- SMART: meaning indicators should be specific (clearly relate to the intervention), measurable (can be quantified, tested and verified), attainable, relevant (a valid measure of the result), and timely (sufficiently frequent and long to capture the time-lag between output delivery and the expected change in result and impact indicators).
- RACER: Relevant (closely linked to the objectives); accepted (the role and responsibilities for the indicator need to be well defined); credible, unambiguous and easy to interpret; easy to monitor (data collection should be possible at low cost); and robust against manipulation (EC, 2017a).

Example of result indicators

EVALUATION OF THE ENERGY INVESTMENT ALLOWANCE IN THE NETHERLANDS (CE DELFT, 2018)
The evaluators calculated the gross and net energy savings achieved by the policy. This concerned the energy savings that occurs by commissioning the energy-efficient technology that is supported by the policy. Hereto, portfolio analysis is applied to estimate the yearly gross energy savings effect.

For the twenty most popular technologies, the Dutch agency RVO annually calculates a savings figure of primary energy per euro invested per year (Nm3 natural gas/euro per year). These twenty technologies represent more than 60% of all applications. The expected annual gross energy savings of the 20 most popular technologies is calculated by multiplying the savings figure with the requested investments. Nm^3 of natural gas is expressed as MJ and TJ based on the calorific value of natural gas in the Netherlands, namely 31,65 MJ/Nm^3. The energy savings are also expressed in terms of avoided CO_2 emissions. A default emission factor of 1,77 kg CO_2/Nm^3 is applied.

Table 8. Total annual gross energy and CO_2 savings from the Energy Investment Allowances (CE Delft, 2018).

Jaar	Besparingen TJ (finaal)	Besparingen CO_2 (Mton)
2012	5.438	0,30
2013	5.573	0,31
2014	6.116	0,34
2015	8.640	0,48
2016	10.821	0,61
2017*	1.742	0,1
Totaal	**38.329**	**2,14**

(2) Collection of data

Data collection is often closely linked to existing monitoring schemes (see section 'Monitoring performance data and new data collection'). Often evaluators depend on monitoring systems which are indicator based. If these are not put in place during the design of the programme, it will be too late to create such systems later on. It is therefore of added value to already consider the type of indicators useful for evaluation purposes in **setting-up a monitoring system**. A system of indicators has also more chance of functioning when the suppliers and users of the information have been involved in its creation (EVALSED, 2013).

(3) Transforming data into indicators

In some cases, the collected data can already be in a format that is useful as an indicator. In the example above, the output indicators used for the evaluation of the Energy Investment Allowance are the same as the data that is monitored (CE Delft, 2018). The only steps that are taken is how the data is grouped and presented.

Often, however, relevant and meaningful indicators cannot be measured directly, but have to be calculated using a combination of monitored data, assumptions and other information sources. This can be done using **bottom-up and top-down methods**. Methodologies to do this can be very different depending on the type of indicator, sectors involved, instrument type, etc. Focusing on indicators about greenhouse gas emission savings, these indicators represent changes in emissions and removals that are the result of an intervention. There are already detailed guidelines available on how to assess energy savings from energy efficiency policies and measures (e.g. EMEEES, 2009; MultEE, 2016), while these are less common for other sectors, such as agricultural policies and measures. There are also resources, such as WRI (2015), with practical guidance on how to calculate emission savings with bottom-up and top-down methods.

The steps involved in calculating changes in greenhouse gas emissions and removals from a policy include:

- **Define the boundary**: policy interventions can have multiple direct and indirect, intended and unintended effects on greenhouse gas emissions. It is important in policy evaluation to acknowledge these effects. However, all direct and indirect effects do not have to be included in the greenhouse gas emission savings indicator. Quantification of some effects might be difficult due to missing information and data and/or can have little added value if the effect is minor;
- **Determine the baseline**: the baseline represents the conditions that most likely would have occurred in the absence of the policy. There are several options to select a baseline:
    - the condition or situation in a reference year;
    - the most likely technology that would have been used as alternative;
    - a modelled counterfactual scenario.
- **Calculate the emission savings**: by comparing the baseline with the observed scenario, greenhouse gas emission savings can be estimated. When bottom-up methods are used, consider and, if possible, quantify rebound and free-rider effects (see also sections on 'counterfactual analysis', 'rebound effect' and 'free rider effect').

Depending on some of the choices and assumptions used, the results of the indicators can be very different. There are methods and tools to give insight in the uncertainty around the results of greenhouse gas savings indicators, such as sensitivity analysis (see section on 'uncertainty analysis'). In the BOXES below, examples are given on how data can be translated into indicators to evaluate a climate policy or measure with bottom-up methods.

Examples on how monitored data can be transformed into indicators in a bottom-up manner

BETTER ENERGY HOMES (BEH) IN IRELAND (EPATEE, 2017)

Description policy measure	The objective of the Better Energy Homes scheme is to improve energy efficiency of residential buildings. It provides direct grants (approximately 30% of the total investment costs) to homeowners or landlords to upgrade their dwellings by insulating ceiling/attic or walls, or by installing heating controls, high efficiency boiler upgrades and solar heating systems. The monitoring and evaluation of the scheme is structured in two ways: • a regular monitoring of each application for a grant (amount of grants approved, number and type of actions carried out, energy savings, CO_2 emissions avoided and jobs supported). Standardized values of energy savings were defined by type of action (in kWh/year.m²) based on the results for 1500 dwellings using simplified engineering calculations; • complementary ex-post studies, for example, to estimate the impact of the scheme.
Data	The ex-post evaluation of actions (in 2009) assessed how much energy savings were realised by the scheme and how this differed from the technical savings potential forecast when the BEH scheme was set up. The evaluation was based on the analysis of the indicator metered gas consumption in a "participants" group (210 homeowners who invested in actions with a grant) and a control group (153.928 households with similar dwellings but who did not participate in the scheme). Data was monitored before and after investment.
Method	The statistical approach used to evaluate the energy savings was based on a difference-in-difference method, comparing pre- and post-intervention heating consumption for both groups. The use of a control group ensures that the energy savings evaluated were related to the improvements carried out by the homeowners, and not to other factors unrelated to the policy (such as differences in climatic conditions in different years).
Rebound effect	The effect was based on the comparison between modelled and metered energy consumption (billing analysis).
Free rider effect	A default value of 18% based on the evaluation of the Energy Efficiency Commitment in UK (2002-2005) was used for the analysis, but this is not included in the regular monitoring of the energy savings.
Interaction	Overlap with other policies was avoided by allocating energy savings per action to only one given policy.
Emission factor	Direct CO_2 emission factors (weighted for the different fuels saved) were applied. The emission factor for electricity during the action lifetime is adjusted to take into account the change in fuel mix used in electricity generation and efficiency improvements in electricity generation.
Result	The ex-post evaluation showed final energy savings of about 21% for the participants on average compared to the control group. However, the ex-post savings were 36% lower than the average technical savings potential. This may be due to the effects of behavioral changes (direct and indirect rebound effects), poor initial estimates of achievable savings (technical savings) and poorly performing equipment and potential inefficiencies in the systems installed.

FRENCH BONUS-MALUS SCHEME (NAVIGANT, 2018), (FRANCE, 2015)

Description policy measure	The Bonus-Malus scheme was implemented by France in January 2008. It introduced one of the first CO_2-based fees and rebates system for new cars. The overall objective is to lower average CO_2 emissions of passenger vehicles from 176g to 130g CO_2 per kilometer in 2020. Since its inception there have been several adjustments to the Bonus-Malus. The system

	started with a fee for cars with CO_2 emissions above 160 g/km, and a rebate for cars with CO_2 emissions below 130 g/km. These limits have decreased over the years. In 2020 an incremental fee is imposed on cars with CO_2 emissions above 110 g/km (138 g/km based on the latest testing methods), while only electric and fuel cell vehicles are eligible for a rebate.
Data	The ex post evaluation looked at data between 2008 and early 2015 and is included in the biennial reporting of France on greenhouse gas policies and measures (France, 2015). The average CO_2 emissions per km of new vehicles in France were monitored and known from 2003 to early 2015. This means two distinct trends can be compared: • the trend between 2003 and end 2007, prior to the implementation of the bonus-malus, corresponding with an average annual decrease by -0.9%; • the trend since 2008, post bonus-malus, with an average annual decrease of -3.5%.
Method	A counterfactual scenario (without measures) was made under the hypothesis that the trend observed before the implementation of the bonus-malus would continue also from 2008 to 2014. To estimate the emission savings from this policy, the average emission factors for new passenger vehicles were multiplied with the number of registrations and the average km driven by cars. These data were available for the period 2008-2013 from the statistics office (Service de l'observation et des statistiques).
Rebound effect	The evaluation recognized that the bonus-malus led to an increase in sales of new vehicles at the start of the implementation period. Which is one type of rebound effect as the production and circulation of these new vehicles implied an increase in emissions. This was not taken into account, because it was assumed that this was only a transient effect caused by the reaction of consumers to the induced price signal. Another, more important rebound effect, is the more intensive use of more economical vehicles. There was not sufficient reliable data to take this into account for this evaluation and it was assumed that the bonus-malus scheme did not lead to an increase in journeys made by car.
Free rider effect	The counterfactual accounts for a continued improvement of the efficiency of new cars, however it cannot be ruled out that a part of the increase in efficient vehicle sales would have happened without the bonus-malus scheme.
Interaction	There is an overlap with EU regulation 443/2009 setting emission performance standards for new passenger cars. The impact of both have not been split, so reported impact includes all policies that affect the efficiency of new vehicles.
Emission factor	Average emission factors of new cars was monitored in France. In the evaluation, it is recognized that there is a gap between official test results and actual performance of CO_2 emissions per km. This was accounted for in the analysis by applying a correction factor based on a study by ICCT (2014).
Result	In 2013, the total avoided emission savings from cars since the start of the measure in 2008 was estimated to be 1,6 Mtonne CO_2.

(4) Interpretation and analysis of indicators

The correct interpretation of indicator data is a very critical step and this requires a thorough understanding of the context and conditions to assess the causal link between the intervention and changes in the indicator. Understanding causality requires three steps (BetterEvaluation, 2020).

- The first step is to check that the **data are consistent with what would be expected**, if the intervention were contributing to producing the observed changes. The *intervention logic model* will be an important guide to assess whether the changes in indicator values matches with how the intervention was expected to work.
- The second is to develop an estimate of what would have happened without the intervention and compare that to the findings of what happened with the intervention. There are several methods that can be used:
  - **Statistical analysis** can be a quantitative approach to establish causality. It does require more resources as additional data might have to be collected and used on potential confounding factors. There are different approaches possible, such as difference-in-difference comparison (see example Better Energy Home from Ireland in BOX) and regression analysis (the latter discussed in more detail in section 4.4).
  - **Counterfactual modelling** uses modelling techniques to construct a counterfactual or baseline scenario. While modelling is typically associated with ex ante impact assessments, they can also be used for ex post evaluations. Models can be used to construct a counterfactual that quantifies what would have happened without the intervention. By comparing the counterfactual scenario with the observed indicators, the impact of the intervention can be analysed (see also the Better Regulation Toolbox #62, EC, 2017b).

  For some indicators, such as for example greenhouse gas emission reduction indicators, this can already be an integral part of the quantification of the indicator itself.
- Identify **other factors that might have caused the changes** in indicator data and see if it is possible to rule them out. To do this, different sources of evidence that addresses the same evaluation question can be collected and combined (triangulation of evidence). This can surpass the quantitative/qualitative divide: quantitative data on indicators can provide overviews and a comparative perspective, while qualitative data (such as stakeholder surveys and interviews) is able to capture subtleties, people's experience and judgements. So quantitative data from indicators can be combined with the qualitative results from stakeholder surveys or interviews.

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Medium - High	Data requirements are high, but a pragmatic approach where indicators are selected with already available data in mind, reduces efforts considerably. When setting-up a monitoring plan, indicators useful for evaluation should already be considered.
Complexity	Low - Medium	In most cases indicators can be relatively easily calculated, if not directly measurable. The complexity in any indicator analysis is to establish the causal link between the intervention and changes in the indicators. For this either analytical techniques (e.g. regression analysis) or triangulation with supporting evidence (e.g. surveys) is needed.
Usefulness	Medium - High	In most evaluations indicator analysis is a pivotal piece of evidence to quantify the impact of the measure. At its best it also can provide a causal link between observed trends and the implementation of the policy or measure.
Resources	Medium	The resources needed, depend on the data availability. If based on existing monitored data, few extra resources are needed for the analysis of the data.
Evaluation criteria	High	Effectiveness and efficiency can be evaluated, but indicators can also be useful when evaluating relevance and coherence, if appropriate indicators are used.
Communication / visualisation of results	High	Indicators can easily be communicated and visualised. An important consideration is that this could also result in spurious conclusions, for example when there is a correlation between a selected indicator and the intervention, but no causation.

Data sources

Data sources for indicator analysis will be very context dependent, coming from national data sources. On energy efficiency, the project Odyssee/Mure[6] collects, aggregates, processes and publishes national statistics and indicators on energy efficiency across all European countries.

Related topics

- Counterfactual analysis
- Monitoring performance data and new data collection
- Regression analysis
- Multi-criteria analysis
- Rebound effect
- Free rider effect
- Uncertainty

[6] https://www.odyssee-mure.eu/

Case studies

- **Case study #5:** Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904

Want to know more?

- (BetterEvaluation, 2020) BetterEvaluation – Sharing information to improve evaluation. Retrieved April 2020, https://www.betterevaluation.org/en/rainbow_framework/understand_causes
- (CE Delft, 2018) Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904
- (EC, 2017a) Better Regulation Toolbox – Tool #41: Monitoring arrangements and indicator, European Commission, 2017, https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-41_en_0.pdf
- (EC, 2017b) Better Regulation Toolbox – Tool #62: Use of analytical models and methods, European Commission, 2017,. https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-62_en_0.pdf
- (EMEEES, 2009) Measuring and reporting energy savings for the Energy Services Directive – how it can be done, EMEEES project supported by IEE, 2009, http://www.evaluate-energy-savings.eu/emeees/en/home/index.php
- (EPATEE, 2017) Better Energy Homes scheme (Ireland). Case study prepared by IEECP (Broc J.S) for the EPATEE project, funded by the European Union's Horizon 2020 programme, 2017, https://www.epatee-toolbox.eu/wp-content/uploads/2018/10/epatee_case_study_ireland_better_energy_homes_ok.pdf
- (EVALSED, 2013) Evalsed Sourcebook: Method and Techniques, European Commission, 2013, https://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf
- (France, 2015) Rapport de la France - En application de l'article 13.1 du règlement 525/2013 relatif à un mécanisme pour la surveillance et la déclaration des émissions de gaz à effet de serre, Ministère de l'Ecologie, du Développement durable et de l'Energie de la France, 2015, https://www.ecologique-solidaire.gouv.fr/sites/default/files/2015_France_MMR_Art13.pdf
- (ICCT, 2014) From laboratory to road - A 2014 update of official and "real-world" fuel consumption and CO2 values for passengers cars in Europe, ICCT, 2014, https://theicct.org/sites/default/files/publications/ICCT_LaboratoryToRoad_2014_Report_English.pdf
- (MultEE, 2016) Document with general formulae of bottom-up methods to assess the impact of energy efficiency measures, MultEE project supported by European Union's Horizon 2020 research and innovation programme, 2016, https://multee.eu/system/files/D2.1_Document%20with%20general%20formulae%20of%20bottom-up%20methods.pdf
- (Navigant, 2018) Bonus-Malus Vehicle Incentive System in France https://www.euki.de/wp-content/uploads/2019/09/20180827_FR_Bonus-Malus_Study.pdf
- (UKERC and IRENA, 2014) Evaluating Renewable Energy Policy: A Review of Criteria and Indicators for Assessment, IRENA and UKERC, 2014, https://www.irena.org/documentdownloads/publications/evaluating_re_policy.pdf
- (WRI, 2015) Monitoring implementation and effects of GHG mitigation policies: steps to develop performance indicators – Working paper, WRI, 2015, https://wriorg.s3.amazonaws.com/s3fs-public/Monitoring_Implementation_and_Effects_of_GHG_Mitigation_Policies.pdf

4.3 Cost effectiveness analysis and cost benefit analysis

Description

For ex-post policy appraisal, the question of whether a policy objective has been achieved in the most cost-effective way, can be answered by means of a **cost effectiveness analysis (CEA)** which relates the costs of a measure to the achieved physical effects (i.e. euro per tonne of CO_2 emissions reduced) (Ecologic, 2005). As indicated in the table below, it allows to compare policy in terms of effectiveness in achieving a single desired outcome for a given level of cost. The type of costs taken into account are focused more on social or economic costs (i.e. costs to society), rather than financial costs (i.e. cash outlays of organisations). To assess the value of the policy measure the costs are considered that accrue to taxpayers, participants, competing organisations or any other groups that are affected by the policy measure (Cellini et al., 2010).

Table 9. Comparison of cost effectiveness analysis and cost benefit analysis (WRI, 2014).

Method	Purpose	Advantages	Disadvantages
Cost-effectiveness analysis (CEA)	To compare policy options to determine which is most effective in achieving a single desired outcome for a given level of cost (such as GHG reduced per dollar), or which option achieves a given objective for the least cost	Simple method to compare policy effectiveness based on GHG emissions reduced per unit of money spent Useful when benefits cannot be calculated or are uncertain	Does not consider wider benefits of the policy/action other than a single measure of effectiveness (such as GHG reduction)
Cost-benefit analysis (CBA)	To compare policy options to determine which has the greatest net benefit to society (the difference between their total social benefits and total social costs); or to analyze a single policy or action to determine whether its total benefits to society exceed its costs	Assesses broader benefits of a policy beyond a single measure of effectiveness (which may include environmental, social, and economic benefits)	Difficult to monetize non-economic benefits and determine appropriate discount rates; can underestimate non-economic benefits

Cost-benefit analysis (CBA) allows to compare the economic efficiency implications of a policy or measure. The social benefits from the policy are contrasted with associated costs based on a common denominator, namely money. By comparing costs and benefits in these monetary terms, a CBA provides an assessment of whether a policy option was worth implementing (i.e. whether the benefits outweigh the costs) from societal perspective (Ecologic, 2005). Economic policy appraisal faces difficulties since some resources, such as environmental ones, are difficult to price in money terms: many goods and services provided by ecosystems – such as clean air, biodiversity – are not traded on a market, and therefore no market price is available which reflects their economic value. The benefits can be estimated through the use of valuation studies, e.g. eliciting people's willingness to pay for a particular environmental good (Ecologic, 2005). The classification of benefits is not as well-developed as for costs. However, they can be direct or indirect meaning that they can affect stakeholders targeted by the initiative or go beyond the target group and even become overall societal benefits (EC, 2017). The benefits may comprise broader environmental, social and economic outcomes of a policy, rather than greenhouse gas or energy savings effects only, as listed in the table below.
For example, the positive environmental impacts of renewables are the avoided pollution from fossil fuel generation due to the displacement of generation by coal, oil or gas burning to generate electricity. The environmental effects of pollution from fossil fuel combustion in power stations include (Finance Sweden, 2012), for instance:

- Increased mortality and morbidity due to higher concentrations of particulates, linked to chest and heart problems;
- Ecological effects on water quality due to acidification;
- Ecological effects on heathlands due to sulphur and nitrogen deposition;
- Damages to agricultural crops, particularly where SO_x and NO_x react to form low-level ozone;
- Impacts on historic buildings;
- CO_2 emissions contributing to global climate change.

Table 10. Cost benefit analysis and examples of non-GHG benefits (WRI, 2014).

Category	Examples of non-GHG effects	
Environmental effects	• Air quality and air pollution (such as particular matter, ozone, carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen oxides (NO_x), lead, and mercury) • Water quality, water pollution, and water scarcity • Ozone depletion • Waste	• Toxic chemical/pollutants • Biodiversity/wildlife loss • Loss or degradation of ecosystem services • Deforestation and forest degradation • Loss of top soil • Loss or degradation of natural resources • Energy use
Social effects	• Public health • Quality of life • Gender equality • Traffic congestion	• Road safety • Walkability • Access to energy, thermal comfort, fuel poverty • Stakeholder participation in policy-making processes
Economic effects	• Employment and job creation • Productivity (such as agricultural yield) • Prices of goods and services (such as decreased energy prices) • Cost savings (such as decreased fuel costs) • Overall economic activity (such as GDP)	• Household income • Poverty reduction • New business/investment opportunities • Energy security/independence • Imports and exports • Inflation • Budget surplus/deficit

CEA differs from CBA which compares the benefits of a policy to its costs, and therefore requires two (or more) parameters, namely measure(s) of benefits and measure(s) of costs. In comparison to CBA, CEA is a rather simple methodology and useful when the policy benefits cannot be quantified in monetary terms (WRI, 2014).

When to use it?

It is most useful to analyse **cost effectiveness** when the policy objectives are clearly identified and defined, ideally connected to a quantified target and a clear baseline. As mentioned, it is also useful in cases where major outcomes are either intangible or otherwise difficult to monetize. On the other hand, the main difficulty is that it provides no value for the output or effects, leaving that to the subjective judgment of the policymaker. In general, CEA may provide a good starting point by requiring the evaluator to identify the most important outcome and relate that outcome to the social costs of the policy option (Cellini et al., 2010).

On the other hand, **cost benefit analysis** is most useful when analysing a single policy to determine whether the total benefits to society exceed the costs or when comparing alternative policies to analyse which option achieves the greatest benefit to society. The major difficulty with CBA is to monetize all costs and benefits. Therefore, although CBA is a profound technique for policy evaluation, it is often difficult and time consuming (Cellini et al., 2010).

How to use it?

In short, the **cost effectiveness** of a policy is calculated by dividing the annualised costs of the policy by a quantified measure of the physical effect: the effects of a policy can be both reduced pressures (e.g. the least cost option to reduce CO_2 emissions) or avoided impacts (e.g. cheapest way to keep global warming below 2°), the latter usually more difficult to assess (Ecologic, 2005).

Carrying out a **cost benefit analysis** requires valuing - as far as possible - in monetary terms both private and external effects of a policy. The valuation of private effects, which occur in markets and for which market prices therefore exist, is comparatively straightforward. For external effects, such as environmental or climate costs and benefits, market prices do not exist and alternative methods must be used to infer the value that different parts of society attach to the effects of a policy (EEA, 2019).

What are the steps to be taken for cost effectiveness analysis and cost benefit analysis (Cellini, 2010)?

(1) Set the framework for the analysis and define the baseline or benchmark

A policy will generate costs and benefits for the society as a whole. The first step to assess these impacts is to consider what would happen if the policy or project was not carried out. This default action is known as the “do nothing” option and provides the benchmark or counterfactual. Evaluators will appraise costs and benefits of the policy relative to their counterfactual.

In the evaluation of the Dutch long term agreement on energy efficiency MJA3 (Ecorys, 2013) – which is described in more detail in case #4 - the additionality of effects and costs on top of autonomous evolutions was determined by comparing Dutch involved industrial sectors to European averages on energy savings from similar sectors. Given the uncertainty of results, the Dutch assessment was complemented with survey results among MJA3 participating sectors to get a better, but still rough, estimation of realized savings without the voluntary agreement MJA3.

(2) Identify and categorize costs and benefits that should be recognized

Even though all costs and benefits cannot be known for certain, the analyst should make a reasonable effort to identify those that were significantly induced by the policy. The most appropriate choice will depend on several factors including the nature of the policy measure and the availability of data (EC, 2017). Costs and benefits may include only the **direct costs or benefits of implementing the policy**, such as compliance costs for regulatory policies (i.e. those costs incurred by businesses and other parties in undertaking the actions necessary to comply with the new regulatory requirements). It is often useful to analyse compliance costs on the basis of their individual components, such as charges, administrative costs and substantive compliance costs (EC, 2017). It may also include **broader indirect costs or benefits** to other members of society (e.g. increased prices for goods or services, reduced energy costs from energy savings, changes in economic activity, multipliers and spillovers). CBA should also include a wide range of social, economic and environmental benefits (WRI, 2014; Cellini et al., 2010), as illustrated in the EXAMPLES below.

Country examples: categories of costs & benefits

In the CEA of the DUTCH LONG TERM AGREEMENT ON ENERGY EFFICIENCY MJA3 (Ecorys, 2013), different direct cost types of implementing MJA3 were estimated (see case #4 for more detail):

- Implementation costs (or administrative costs): the costs that the government must incur to ensure compliance and/or implement the MJA3. These type of costs were monitored annually by the Dutch Agency;
- Administrative burdens are the costs for the companies or business to meet information obligations arising from the MJA3. It's about collecting, editing, registering, saving and making available the necessary information. An online survey allowed a rough estimate to be made of these type of costs by collecting the number of required hours of employee categories (secretary, staff and management) to comply with the obligations;
- Substantive compliance costs are the additional costs of companies to comply with the MJA3 with regard to behavior of persons and conditions (buildings, production processes or products/services) in companies with a view to safeguard public goals. The online survey allowed to make a rough estimate of these type of costs as well.

The CBA of the FRENCH FISCAL MEASURE ÉCOPASTILLE (i.e. eco-tax bonus-malus & super-bonus for new, private vehicles), monetized by econometric studies multiple (in)direct costs and benefits (CGDD, 2013) (see case #6 for more detail):

- Economic costs: loss of consumer surplus linked to changed consumer’s choice compared to their previous buying habits; and opportunity costs for public funds which are both derived from econometric regression studies;

- Environmental costs related to the effects of local air pollution;
- Economic benefits generated by fuel savings over the lifetime of new passenger vehicles;
- Environmental benefits or CO_2 emission reductions.

Concerning ex-post evaluations of energy efficiency policies within Europe, broader effects or benefits of an energy efficiency instrument (such as energy efficiency obligation schemes EEOs) are illustrated in Figure 11. Despite this diversity of benefits, most evaluations that are currently carried out focus on one benefit only, namely bill savings, which is often compared to the cost of energy efficiency policy measures. A more comprehensive analysis would need to incorporate a wider range of benefits from policymaker's, individual and societal perspective (RAP, 2016).

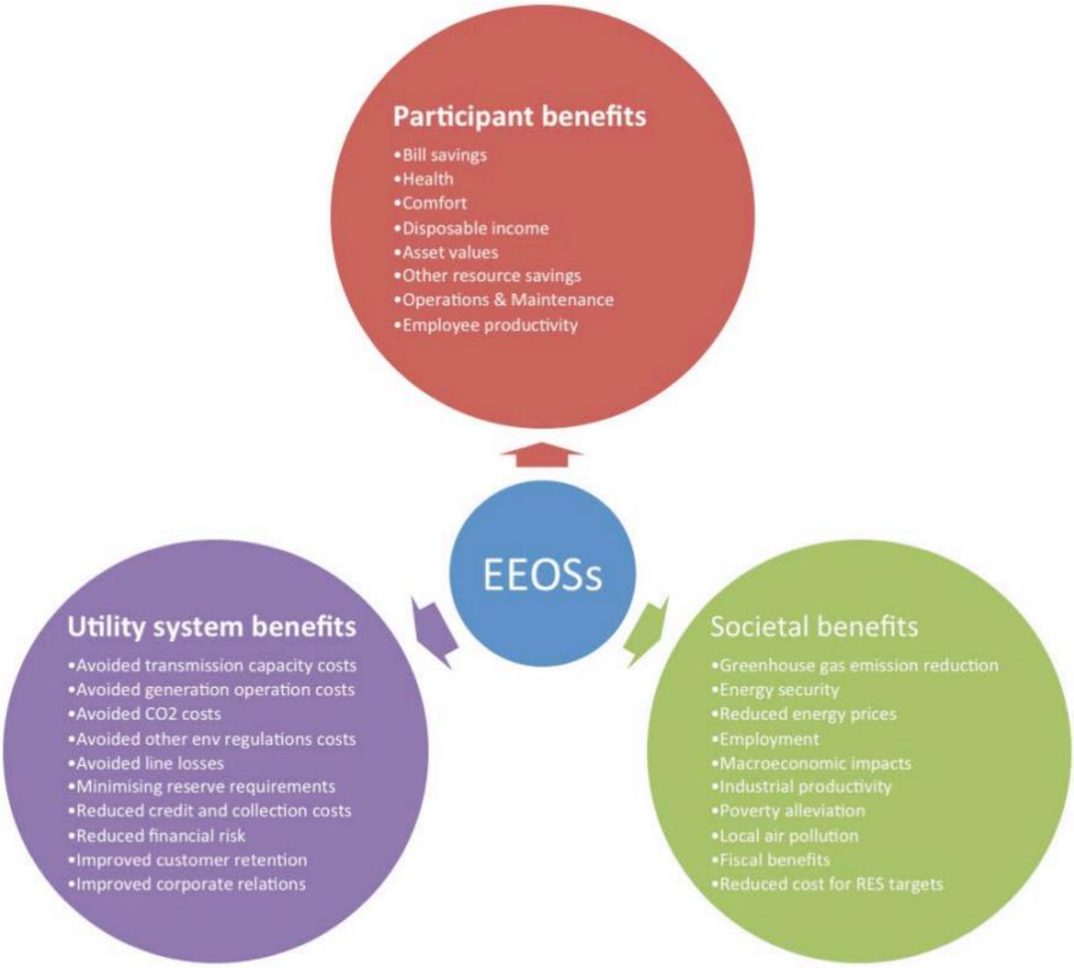


Figure 11. Multiple benefits of energy efficiency policies, namely energy efficiency obligation schemes (EEOs) (RAP, 2016).

(3) Project costs and benefits over the (relevant) life of the policy measure, if applicable

In this step, a time period is defined that is sufficient to capture the significant costs and benefits of the policy, as well as how costs and benefits changed over time. For an ex post analysis, much may be known if actual costs and outcomes have already been reported or monitored regularly by involved parties (cf. monitoring by Dutch Agency for MJA3). It may help to consider whether costs and benefits occur only one time, accruing in the first year, or whether they are recurring costs or benefits that occur every year (Cellini et al., 2010).

(4) Monetize the identified costs

In this step, the identified costs are expressed in monetary terms. Different methodological approaches can be used to estimate the costs ex-post, and will depend on several factors (e.g. data availability, type of policy measure), as illustrated in the BOX below on monetizing costs and benefits. For each cost monetized, it is necessary to describe its definition, how it is measured and any assumptions made in the estimations. The assumptions are preferably subjected to a **sensitivity analysis** to determine to what extent the outcome of the analysis is dependent on these values (Cellini et al., 2010). Preferably flag the limitations of any estimated result and take these into account when comparing the multiple policy options (EC, 2017).

(5) Quantify benefits in terms of units of effects (CEA) or monetize benefits (CBA)

For **cost effectiveness analysis**, the effects are the quantifiable outcomes central to the policy's objectives. In this context, it is the total net change (i.e. additional change in comparison to the benchmark) of greenhouse gas emissions resulting from the policy measure. In the CEA of the Dutch long term agreement on energy efficiency MJA3, the main effects comprised the additional, yearly energy savings resulting from the MJA3 covenant (WRI, 2014).

The benefits in case of **cost benefit analysis** involves assigning a monetary value as a proxy to represent benefits for social, economic and environmental impacts that may not have an explicit economic or monetary value. CBA is dependent on the assumption that the value of non-economic impacts can be represented by the value that individuals are willing to pay to preserve or avoid damages. However, some benefits may be intangible, uncertain, subjective, or controversial to monetize (WRI, 2014). Again, it is important to describe the definitions of the monetized benefits, how they are measured and any assumptions made. Preferably, a sensitivity analysis is performed to understand how the outcome of the CBA is controlled by these assumptions.

How can costs and benefits be quantified?

Effects on existing markets can be monetized with reference to market prices. Where there is no market, there is of course no market price, but other techniques are available that can be used to quantify the effects (PBL, 2013). Table 11 synthesizes indicators and methodologies used for the quantification of different effects of climate change policies and measures. For some of the categories identified (e.g. employment effects), limited examples of valuation techniques are as yet available, and for others (e.g. value of reduced mortality), some significant uncertainties remain. In those cases, and when monetization is controversial, reporting results in physical units is often advisable and can be combined with alternative assessment techniques, such as multicriteria analysis (Urge-Vorsatz et al., 2014).

As different methodologies apply to measuring the various types of costs and benefits, EEA published a user-friendly overview of these types of costs and benefits as well as the methods for assessing them. This way, EEA wants to guide Member States in their policies and measures reporting and to enhance comparability of the measures' impacts reported by Member states. The compilation can be found on EIONET MMR (EEA, 2019): http://cdr.eionet.europa.eu/help/mmr/PaMs_Cost-benefits%20methodologies_climate%20mitigation_EEA.xlsx.

Table 11. Indicators and methodologies used for the monetisation of different effects of climate mitigation policies and measures, separated by physical and monetary metrics (Urge-Vorsatz et al., 2014).

Category of effects	Subcategory of effects	Physical indicator	Monetary indicator	Appraisal method
Health benefits	Outdoor air pollution related Indoor air pollution related Energy poverty related Outdoor noise related Transport and traffic related Heat island related	Avoided cases Avoided hospital admissions Restricted activity days Years lived with disability Disability-adjusted life years (DALYs) Quality-adjusted life years Years of life lost	Avoided costs approach: cost of illness (cost per avoided case) Willingness to pay (WTP) for avoided case or death: value of a lost year and value of a statistical life (VSL)	Revealed preferences: avoided costs approach Stated preferences approach (contingent valuation)
Energy poverty and distributional effects	Access to modern energy services	Additional kWh of quality energy (e.g. electricity) consumed Households with modern energy services (e.g. connected to the electricity grid)	WTP for an additional unit of quality energy (e.g. cost per kWh) or for having access to electricity (cost per household)	Consumer surplus estimation through stated preferences method (contingent valuation)
	Affordability of energy services	Decreased energy demand (e.g. kWh)	Per unit cost of energy (e.g. cost per kWh)	Energy prices
Comfort and living conditions	Thermal comfort	Increased indoor temperatures Increased percent of floor area heated	Forgone energy cost savings	Energy prices
	Exposure to external noise	Decibels (dBs) of external noise avoided	WTP to reduce exposure to external noise (e.g., cost per dB) Increase in the rental or sale price of properties (cost, percent)	Stated preferences (contingent valuation) Hedonic pricing

Category of effects	Subcategory of effects	Physical indicator	Monetary indicator	Appraisal method
Provision of ecosystem services		Hectares (ha) of ecosystem or units of ecosystem service flow (e.g. number of recreational visitors per year)	Cost per ha of ecosystem or unit of ecosystem service flow per year (e.g. cost per ha per year)	Market prices, stated preferences, and revealed preferences Benefit transfer and meta-analytical techniques
Damage to building materials		Frequency of cleaning and maintenance of buildings	WTP for avoiding damage to building materials	Avoided cost approach (cleaning and restoration) Stated preferences (e.g. contingent valuation)
Productivity	Performance of individuals and organizations	Increase in labor productivity	Per unit labor costs	Market price of labor
	Crop yields	Increase in crop yields (percent)	Cost per unit of agricultural produce (e.g. cost per tonne)	Avoided cost and price of agricultural products
Energy security		Units of imported energy avoided (e.g. oil barrels)	Cost per unit of imported energy (e.g. cost per oil barrel) WTP to secure the energy supply (e.g. cost per MWh)	Estimation of the macroeconomic external costs of energy imports Stated preferences (contingent valuation)
Macroeconomic effects		Percentage points of additional gross domestic product growth (%) Additional full-time equivalent (FTE) positions created	Monetary units per additional employment created (e.g., cost per FTE) Shadow price of labor costs in social cost-benefit analysis	Input-output (I/O) analysis, computable general equilibrium (CGE) models and macroeconomic models Analytical methods Opportunity costs of labor and public expenditures Stated preferences (choice experiment) Shadow pricing of labor costs

(6) Discount costs and benefits to obtain present values

This step is about converting all monetary values of CEA and CBA to their present value or their equivalent value at the beginning of the policy measure (year 1). Rather than an actual interest rate, in CEA and CBA a **social discount rate** is applied to calculate the present value of costs and benefits. The social discount rate reflects society's preference for consumption today over consumption in the future (Cellini et al., 2010). The choice of an appropriate discount rate is important, however, social discount rates can vary widely (for example, from 0% to over 10%), depending on how to address equity concerns with respect to future generations, which are not accounted for in national interest rates or typical discount rates. Present value is calculated as follows (WRI, 2014):

$$PV = V_y / (1+r)^t$$

where PV = present value; V_y = value in a particular year; r = discount rate; and t =number of years from present.

(7) Calculation of a cost effectiveness ratio (for CEA) or a net present value (for CBA)

In the next step, the calculated present values for costs and benefits are represented as a cost effectiveness ratio (CEA) or as net present values (CVA) estimated for the analysis period.

Table 12. Calculation of cost effectiveness ratio (CEA) and net present value (CBA) (WRI, 2014).

Calculate cost-effectiveness	Calculate net present value
CEA results in a ratio of costs to effectiveness, as follows: $\text{cost effectiveness} = \frac{PV(c)}{\text{effectiveness}} = \frac{\sum_{t=0}^{n} \frac{C_t}{(1+r)^t}}{\text{net reduction in t } CO_2e}$ C = costs, t = year, n = analysis period	Once present values for costs and benefits are calculated, the result of the CBA is represented as the net present value (NPV) of all benefits and costs, representing the net social benefit: $NPV = PV(B) - PV(C)$ $NPV = \sum_{t=0}^{n} \frac{B_t}{(1+r)^t} - \sum_{t=0}^{n} \frac{C_t}{(1+r)^t}$ B = benefits, C = costs, t = year, n = analysis period

For climate change policy, it is to be expected that a primary objective or outcome is to reduce greenhouse gas emissions. In this case, it would be appropriate to consider the cost effectiveness ratio in terms of the average cost of saving each tonne of CO_2 (equivalent). The NPV is the sum of all monetized costs and benefits, discounted to the base year chosen, and reflects the valuation of changes in the outcome, for example, changes in the traded and nontraded GHG emissions resulting from the policy measure. If the NPV is positive, the policy is estimated to provide a net monetized benefit, and if the NPV is negative, then the policy is estimated to result in an overall monetised cost to society (BEIS, 2013).

(8) Perform a sensitivity analysis by varying the assumptions to estimate costs & outcomes

The sensitivity analysis means recalculating the NPV when the values of certain key parameters are changed (e.g. discount rate). A more sophisticated approach is to assume that each parameter follows a distribution and to then take multiple values from this distribution for each variable. For each draw, a NPV can be calculated, leading to distribution of NPVs (Finance Sweden, 2012). In the EXAMPLE on the social cost of CO_2 (SC-CO2) developed by the US EPA, it is illustrated that the discount rate has a strong impact on the resulting SC-CO2.

Example on sensitivity analysis of discount rates

SOCIAL COST OF CARBON ACCORDING TO US EPA (2016)
EPA and other federal agencies use estimates of the social cost of carbon (SC-CO2) to value the climate impacts of regulatory actions. The SC-CO2 is a measure, in dollars, of the long-term damage done by a tonne of CO_2 emissions in a given year. This dollar figure also represents the value of damages avoided for a small emission reduction (i.e. the benefit of a CO_2 reduction).

The SC-CO2 is meant to be a comprehensive estimate of climate change damages and therefore includes, among other things, changes in net agricultural productivity, human health, property damages from increased flood risk and changes in energy system costs, such as reduced costs for heating and increased costs for air conditioning. However, it does not currently include all important damages. The integrated assessment models first estimate damages occurring after the emission release and into the future, often as far out as the year 2300. The models then discount the value of those damages over the entire time span back to present value to arrive at the SC-CO2. For example, the SC-CO2 for the year 2020 represents the present value of climate change damages that occur between the years 2020 and 2300; these damages are associated with the release of one tonne of carbon dioxide in the year 2020.

One of the most important factors influencing SC-CO2 estimates is the discount rate. A large portion of climate change damages are expected to occur many decades into the future and the present value of those damages (the value at present of damages that occur in the future) is highly dependent on the discount rate. SC-CO2 estimates based on several discount rates are therefore included in the social cost, because the literature shows that the SC-CO2 is highly sensitive to the discount rate and because no consensus exists on the appropriate rate to use for analyses spanning multiple generations.

The table below summarizes four SC-CO2 estimates in multiple years for different discount rates (5%, 3% and 2.5%) indicating the sensitivity of the social costs for this key parameter. The four SC-CO2 estimates are: $14, $46, $68, and $138 per metric tonne of CO_2 emissions in the year 2025 (in 2007 dollars).

Table 13. Social cost of CO_2 as estimated by EPA, 2015-2050 (SC-CO2 are dollar-year ($2007) and emissions-year specific) (EPA, 2016).

	Discount Rate and Statistic			
Year	5% Average	**3% Average**	2.5% Average	High Impact (3% 95th percentile)
2015	$11	**$36**	$56	$105
2020	$12	**$42**	$62	$123
2025	$14	**$46**	$68	$138
2030	$16	**$50**	$73	$152
2035	$18	**$55**	$78	$168
2040	$21	**$60**	$84	$183
2045	$23	**$64**	$89	$197
2050	$26	**$69**	$95	$212

Advantages/disadvantages

Characteristics	Cost effectiveness analysis		Cost benefit analysis	
	Score	Explanation	Score	Explanation
Data requirements	Medium	For an ex post analysis, much may be known in cases where actual costs and outcomes have been reported or monitored annually by involved parties.	High	The major difficulty is to monetize all costs and benefits.
Complexity	Low - Medium	The actual calculations of cost effectiveness are rather straightforward, and Excel may suffice. The complexity will depend on several factors including the nature of the policy measure and the data availability, as this will determine the way of identifying the benchmark and the type of costs & effects of the assessment.	Low-Medium	The actual calculations of CBA are rather straightforward (NPV), and Excel may suffice. The complexity will depend on several factors including the nature of the policy measure and the data availability, as this will determine the way of identifying the benchmark and the type of costs & benefits of the assessment.
Usefulness	Medium	An appropriate CEA allows to evaluate whether a policy objective has been achieved in the most cost-effective way, which relates the costs of a measure to the achieved physical effects.	High	By comparing costs and benefits in monetary terms, a CBA provides an assessment of whether a policy option is worth implementing (i.e. whether the benefits outweigh the costs) from a societal perspective.
Resources	Low-Medium	The resources needed depend on the data availability and type of policy measure: extra resources are needed for the collection of new evidence (e.g. survey and interviews of stakeholders) and various experts supporting the understanding of benchmark and type of costs & effects.	High	The major difficulty with CBA is to monetize all costs and benefits. Therefore, CBA is often difficult and time consuming
Evaluation criteria	Medium	Evaluation criteria effectiveness and efficiency can be estimated.	Medium	Evaluation criteria effectiveness and efficiency can be estimated.
Communication / visualisation of results	Medium	The resulting cost effectiveness can be communicated and visualized in an understandable manner. The uncertainties in relation to costs and outcomes, should be reported and communicated in a transparent way, to avoid misinterpretation of results.	Medium	The resulting NPV can be communicated and visualized in an understandable manner. The uncertainties in relation to costs and benefits, should be reported and communicated in a transparent way, to avoid misinterpretation of results.

Related topics

Counterfactual analysis
Methodologies for collection of evidence
Rebound effect
Free rider effect
Uncertainty

Case studies of relevant examples of the approach

- **Case study #4: CEA:** Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3) - Eindrapport van de door Ecorys uitgevoerde Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3), Ecorys, 2013, https://www.rijksoverheid.nl/documenten/rapporten/2013/04/10/evaluatie-meerjarenafspraak-energie-efficientie-2008-2020-mja3
- **Case study #5: CEA:** Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904
- **Case study #6: CBA :** Évaluation économique du dispositif d'écopastille sur la période 2008-2012, Commissariat Général au Développement Durable (CGDD), 2013, http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0078/Temis-0078465/20744.pdf

Want to know more?

- (BEIS, 2019) Valuation of Energy Use and Greenhouse Gas (GHG) Emissions: Supplementary Guidance to the HM Treasury Green Book on Appraisal and Evaluation in Central Government, Department for Business, Energy and Industrial Strategy BEIS United Kingdom, 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/794737/valuation-of-energy-use-and-greenhouse-gas-emissions-for-appraisal-2018.pdf
- (Cellini et al., 2010) Cost Effectiveness and Cost-Benefit Analysis - In Handbook of Practical Program Evaluation - 3rd ed. edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, Authors chapter: Cellini, Stephanie R., and James E. Kee, 2010, http://www.blancopeck.net/HandbookProgramEvaluation.pdf
- (CGDD, 2013) Évaluation économique du dispositif d'écopastille sur la période 2008-2012, Commissariat Général au Développement Durable (CGDD), 2013, http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0078/Temis-0078465/20744.pdf
- (EC, 2017) Better Regulation Toolbox – Tool #59. Methods to assess costs and benefits, European Commission, 2017, https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-59_en_0.pdf
- (Ecologic, 2005) Cost-effectiveness of environmental policies - An inventory of applied ex-post evaluation studies with a focus on methodologies, guidelines and good practice, Ecologic, 2005, https://www.ecologic.eu/sites/files/project/2013/1731_Cost-effectiveness_conclusions.pdf
- (Ecorys, 2013) Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3) - Eindrapport van de door Ecorys uitgevoerde Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3), Ecorys, 2013, https://www.rijksoverheid.nl/documenten/rapporten/2013/04/10/evaluatie-meerjarenafspraak-energie-efficientie-2008-2020-mja3
- (EEA, 2019) MMR Article 13 - Methodologies for Cost-Benefit Analysis, EEA and Environment Agency Austria, PBL, December 2019, http://cdr.eionet.europa.eu/help/mmr/PaMs_Cost-benefits%20methodologies_climate%20mitigation_EEA.xlsx
- (EPA, 2016) EPA Fact Sheet: Social Cost of Carbon, US EPA, 2016, https://www.epa.gov/sites/production/files/2016-12/documents/social_cost_of_carbon_fact_sheet.pdf
- (Finance Sweden, 2012) The Costs and Benefits of Renewable Energy in Scotland, report for the Expert Group on Environmental Studies, Ministry of Finance Sweden, 2012, http://www.sou.gov.se/wp-content/uploads/2014/05/2012_5-The-Costs-and-Benefits-of-Renewable-Energy-in-Scotland.pdf
- (PBL, 2013) General guidance for cost-benefit analysis, PBL, 2013, https://www.pbl.nl/sites/default/files/downloads/pbl-cpb-2015-general-guidance-for-cost-benefit-analysis_01512.pdf

- (RAP, 2016) Costs and Benefits of Energy Efficiency Obligation Schemes, RAP, April 2016, https://ec.europa.eu/energy/sites/ener/files/documents/final_report_on_study_on_costs_and_benefits_of_eeos_0.pdf
- (Urge-Vorsatz et al., 2014) Measuring the Co-Benefits of Climate Change Mitigation, Urge-Vorsatz D., Tirado Herrero S., Dubash N.K. and Lecocq F., Annual Review of Environment and Resources Volume 39, 2014, https://www.annualreviews.org/doi/pdf/10.1146/annurev-environ-031312-125456
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

4.4 Regression analysis

What is it?

Regression analysis is a statistical method to investigate the relationship between two or more variables. It does this by analysing how one independent variable affects another dependent variable, when all the other independent variables are fixed.

When to use it?

Regression analysis is a useful method to estimate impacts of a policy in the case (1) where the policy effect can be characterised by a specific variable (e.g. product emissions standards are expected to lead to changes in the greenhouse gas emissions arising from the use of the respective product) and (2) when good **data is available on the trends for this variable** as well as for the other variables which may have an effect on it (e.g. sales of the product, usage of the product, emission factor i.e. GHG/unit output). However, in the case where the policy cannot be characterised so easily, or where data on the different variables is lacking, a regression analysis may be less useful as a tool.

A regression analysis can also be useful in **identifying outliers** from expected results. For example, it can show that in certain applications the policy is effective in achieving its objective, while also identifying applications of the policy where the results are different or less effective. This may then indicate other important variables that were not initially expected to be important. Using a product example, as above, a regression analysis might indicate that certain models of product may have a different effect which is an apparent outlier to the trend of other products.

How to use it?

Regression analysis is based on single or multiple equations using historical data to estimate, for instance, behavioural (and unobservable) relationships that cannot be directly measured. The exact specification of regression equations varies highly by activity and sector affected by the policy or measure. In terms of procedure, the following ingredients are crucial: First, **expert knowledge and judgement** is needed to set-up a specification, to conduct the estimation and interpret the results. Secondly, the actual estimation and interpretation requires expertise in econometric and statistical analysis (theory and practice) including the testing of statistical parameters to assure quality and robustness of the results. Thirdly, data requirements go far beyond basic requirements **as time series data of sufficient lengths** for each variable is needed in order to conduct the analysis and ensure quality. Typical inputs are data sets which combine cases with and without the changes, so that the differences can be analysed compared to the counterfactual. These differences can be either over time (e.g. before and after introduction the measure), over sector (e.g. those inside and outside the coverage of the measure), or in different geographical areas (e.g. inside and outside a member state).

As there are a number of different ways in which a regression analysis can be applied, a specific example is used – an ex-post evaluation of regulations on CO_2 emissions from cars and light duty (LDV) vehicles (DG Climate Action, 2015) – to illustrate the approach.

Example on regression analysis

Ex-post evaluation of regulations on CO_2 emissions from cars and light duty vehicles (DG Climate Action, 2015)

The policies being evaluated in this case study consist of two separate regulations:

- Regulation (EC) No 443/2009 also called passenger car CO_2 Regulation;
- Regulation (EU) No 510/2011 also called the Light Duty Vehicle (LDV) CO_2 Regulation.

Both regulations aim to reduce GHG emissions from road transport by setting mandatory fleet-based CO_2 reduction targets in terms of gCO_2/km for new cars and for new LDVs respectively. The regulations oblige manufacturers to adhere to these minimum standards for all new cars and LDVs sold in the EU market. For each type of vehicle, a utility parameter is set to account for differences in vehicle types, while ensuring emission reductions targets are set in line with the weight of vehicles.

Scope of the evaluation

The evaluation aimed to investigate both the effectiveness of the regulations in achieving their objectives as well as their wider societal, economic and environmental impacts. It therefore assessed the effectiveness of the regulations in terms of emission reductions achieved in the period from 1995 to 2014 in the EU vehicle market and compared this to the originally anticipated results. The evaluation questions for the study were as follows:

(1) Effectiveness:

- To what degree have the regulations contributed to achieving their targets and what are their weaknesses?
- To what extent have the regulations been more successful in achieving their objectives compared to the voluntary agreement on car CO_2 emissions?
- How do the effects of the regulations correspond to the objectives?

(2) Efficiency:

- Are the costs resulting from the implementation of the regulations proportional to the results that have been achieved?
- What are the major sources of inefficiencies? What steps could be taken to improve the efficiency of the regulations? Are there missing tools and/or actions to implement the regulations more efficiently?

The evaluation criteria of relevance, coherence as well as EU added value were assessed as well, but were not subject to regression analysis.

Evaluation methodology

In order to answer the evaluation questions relating to effectiveness and efficiency, a regression model was used to quantify the impact of the regulations on emission reductions from passenger cars and LDVs, while controlling for possible other factors that may have had an impact on emissions from new vehicles, e.g. technological developments. The key input data as well as the data sources used for the regression analysis are outlined in the table below.

Table 14. Overview of type of data and data sources used for regression analysis of car and LDV CO_2 regulations (DG Climate Action, 2015).

Type of data	Data source used
Annual new vehicle registrations by vehicle type and fuel type (e.g. petrol, diesel and alternative fuel vehicle)	ACEA Statistics EEA CO_2 Monitoring reports which provides data on CO_2 emissions as well as manufacturer name, type, version, make, mass, fuel type etc. ICCT – International Council on Clean Transportation
CO_2 performance and registrations data by average mass and mass distribution in form of bins for each country by manufacturer and fuel type	The Commission's Decision 1753/2000 Monitoring of CO_2 emissions database from 2000- 2009
Annual and lifetime mileage by fuel type	UK's MOT database (Ministry of Transport) of periodic technical inspection data for vehicles. It has been assumed that the figures are representative of the wider European fleet (equivalent comprehensive data for other countries do not currently exist).
Vehicle survival rates by age	UK's MOT database of periodic technical inspection data and, in the absence of equivalent datasets from other countries, it has been assumed that these data are representative of the wider EU.
Well-to-tank emission factors	JEC Consortium study on well-to-wheels analysis of fuels and powertrains (JEC, 2014)
EU-average pre-tax and post-tax fuel prices per year	Fuel prices covering the period of interest were obtained from European Commission's EU Oil Bulletin 2006-2014 (DG ENER, 2014). The EU Oil Bulletin provides detailed prices covering each week in any given year of interest. The weekly figures were averaged to provide annual average prices for the 2006-2013 time period, and were also adjusted for inflation so that all prices were presented in a common reporting year (i.e. 2014 prices).
Data on market penetration of CO_2 abatement technologies for cars and LCVs	Previous study by the European Commission specifically on market penetration
Data on the unit costs of CO_2 abatement technologies for cars and LCVs	Analysis carried out for ICCT by the consulting firm FEV (FEV, 2013), and from other sources including research by the National Research Council of the US National Academies (National Academies, 2011)

Using the data collected as outlined in the table above, the evaluation assessed the average specific emissions of the new European car and LDV fleet for different years. The result of this thereby showed whether the objective of the regulations was achieved and to what extent.

A regression analysis was used to investigate to what extent the regulations had driven this change in emissions in the car and LDV fleet. In order to do this, the effects of other parameters that are independent from the regulations, but that could also impact on emissions, were taken out of the study or adjusted so that the only remaining change in CO_2 emissions can be attributed to the regulations. Therefore, the following types of **control variables** were defined to control for other variables being independent from the regulations*:*

(1) *Time-specific variables:* all variables that are specific to the time period of the policy that is studied. These variables can allow for the study to control for changes in the dependent variable that cannot be explained by other variables except for the time period chosen. In the vehicle CO_2 regulations study, a time trend was used instead of choosing individual variables, to show an autonomous improvement over time irrespective of the regulations. This autonomous trend was then taken out of the CO_2 impacts to adjust for this time-specific impact.

(2) *Outcome variables:* all variables that can impact on the outcome studied (e.g. CO_2 emissions of new cars) that are not a consequence of the regulations directly. For example, the study showed that manufacturers had reduced the mass of cars in order to meet the regulation standards in a different way than intended. Therefore, in the study the average mass of cars was controlled for to avoid any bias of this change.

(3) *Omitted variables:* these are variables that are correlated in some way both with the policy introduction and the results (e.g. CO_2 reduction), but do not show a causal relationship. In the study of the vehicle CO_2 regulations, shifts in consumer preferences were categorized as omitted variables. This is because if consumer preferences have shifted towards lower emission cars independent of the regulations, this effect needs to be controlled for in the results to ensure it is not attributed to the effectiveness of the regulations.

(4) *Anticipated variables:* often in the case of GHG reduction policies, the measures and their introduction are announced months or years in advance of the point in time when the policy goes into effect. This can already impact on emissions either by affected entities already adapting to incoming legislation or by investing heavily in high emission activities before the regulation comes into place. Therefore, it may be necessary to include an anticipation variable that can measure the impact in the period between the announcement of the policy and the actual implementation start date.

After appropriately selecting a set of control variables using the categorization as outlined above, the impacts of each of these variables can be eliminated from the overall CO_2 reduction result. In the case of the vehicle CO_2 regulations study, the following formula was used for both before and after the introduction of the policy:

Average gCO2/km = Fixed CO2 value expected + time trend + changes expect due to other factors + adjustment for anticipation factor.

Evaluation outcome

The regression analysis then produces a set of results as given in the example below for the vehicle regulations study. The table below shows in each row a variable that could impact on emission reductions, e.g. the introduction of the regulations (Policy2009) or autonomous improvement over time (time). The actual emission reductions that can be attributed to each variable are indicated in the values presented. Underneath each value the robust standard errors are presented in parentheses, which indicate the significance level (i.e. the probability that the estimated effect is discarded, e.g. a p-value of 0.05 demonstrates a 5% chance that the conclusion of emission reductions attributable to the variable is false).

The overall results indicate that "*an annual reduction of 3.5 gCO2/km is attributable to the regulations while autonomous improvement is around 1.6 gCO2/km per year*". In addition, the inclusion of an anticipation factor does not have a significant impact on the results. Moreover, the results show that an increase in the average share of diesel cars has a very significant impact on emission reductions. However, during the timeframe of this evaluation, the share has not changed a lot, so this factor is less important in the overall results.

Table 15. Example of results from regression analysis (DG Climate Action, 2015).

	With anticipation	Without anticipation
Anticipation	-0.524 (1.034)	
Policy2009 * Time	-3.501*** (0.450)	-3.409*** (0.346)
Time	-1.594*** (0.346)	-1.684*** (0.252)
Registrations	-0.004 (0.003)	-0.004 (0.003)
Share of diesel	-27.672*** (7.183)	-27.701*** (7.165)
Policy2009	23.402*** (2.825)	23.096*** (2.669)
Constant	187.769*** (3.668)	188.067*** (3.601)
Observations	264	264

Robust standard errors in parenthesis
** p<0.1, ** p<0.05, *** p<0.01*

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	High	The results of a regression analysis can depend a lot on the quality and availability of data, in particular the relationship of each data type with the result measures. In addition, the number of control variables can also improve the result of a regression analysis. A regression analysis therefore has a high reliance on data both in qualitative and quantitative terms.
Complexity	High	As regression analysis needs to adjust for any other impact than the policy on emissions, it has a high complexity to avoid results being skewed by other impacts, e.g. technological development, behaviour change etc. This adds a high complexity to using this methodology in addition to the econometric modelling and statistical skills required to carry out a regression analysis.
Usefulness	Medium	Often in the case of GHG policies there is no control group available and instead the study relies on a 'before' and 'after' comparison. This means the choice of control variables is important, but caution with the results is needed as there may be variables that were not adjusted for.
Resources	Medium	Regression analysis can be time-consuming due to the need for a lot of data (time series) that needs to be collected and analysed for its impacts and the use of model software.
Evaluation criteria	Medium	Effectiveness and efficiency: regression analysis is suitable to evaluate quantitative criteria that aim to estimate how their change has been impacted by the policy (e.g. CO_2 reduction, energy savings, cost savings etc.)
Communication / visualisation of results	Medium	Results can be presented through the use of graphs and tables that can visualise distributions and attribution of impacts to specific elements. However, results may be hard to interpret due to their highly quantitative nature and careful explanation is needed which factors have been taken into account and which not.

Tools

Various software programmes can be used for regression analysis, such as MS Excel, SPSS, Stata, R or Python.

Data sources

As outlined above, regression analysis heavily relies on high quality, time series data of sufficient length to ensure that all variables that could impact the results are considered.

Related topics

Methodologies for collection of evidence
Indicator analysis

Assessing policy interactions
Counterfactual analysis
Rebound effect
Uncertainty

Want to know more?

- (DG Climate Action, 2010) Ex-post quantification of the effects and costs of policies and measures CLIMA.A.3/SER/2010/0005, Öko-Institut, Cambridge Econometrics, AMEC and TNO under authority of European Commission – DG Climate Action, 2010, https://ec.europa.eu/clima/sites/clima/files/strategies/progress/monitoring/docs/report_expost_quantification_en.pdf,
- (DG Climate Action, 2015) Evaluation of Regulations 443/2009 and 510/2011 on CO2 emissions from light-duty vehicles Final Report. Study contract no. 071201/2013/664487/ETU/CLIMA.C.2, Ricardo AEA and TEPR under authority of European Commisson – DG Climate Action, 2015, https://ec.europa.eu/clima/sites/clima/files/transport/vehicles/docs/evaluation_ldv_co2_regs_en.pdf
- (DG ENER, 2014) Market Observatory Oil bulletin, time series data, European Commission, Directorate-General for Energy, 2014.
- (Harvard Business Review, 2015) A refresher on regression analysis. Harvard Business Review by Amy Gallo, 2015, https://hbr.org/2015/11/a-refresher-on-regression-analysis
- (Fujiwara et al., 2019) The practice of climate change policy evaluations in the European Union and its member states: results from a meta-analysis, Fujiwara, N., van Asselt, H., Böβner, S. et al., Sustain Earth 2 – 9, 2019, https://doi.org/10.1186/s42055-019-0015-8
- (JEC, 2014) Well-to-wheels analysis of future automotive fuels and powertrains in European context – JRC technical reports, JEC, 2014, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC85329/wtw_report_v4a%20march%202014_final.pdf

4.5 Decomposition analysis

Description

Decomposition analysis can be used to quantify the influence of key factors on a variable of interest (for example CO_2 emissions). Such key factors can also be called drivers.

Example: Kaya identity

One famous identity describing the relationship between such drivers and the variable of interest is the Kaya identity (IPCC, 2014) which reads as follows:

*CO2 = population * GDP / population * energy / GDP * CO2 / energy*

The **variable** of interest in this case is CO_2 emissions. The **drivers** are: population, wealth (GDP / population), energy intensity of GDP (energy / GDP) and emission intensity (CO2 / energy).

While CO_2 emissions are typically measured in Mt CO_2-eq., all the drivers are measured in other units. Using decomposition analysis, each of these drivers' impact on the variable of interest is "translated" into the unit in which the variable is measured (i.e. into Mt CO_2-eq). This has an added value because otherwise each driver's impact would not be directly visible.

As a consequence, decomposition analysis can help to understand why a variable of interest developed as it did, taking into account relevant drivers.

When to use it?

Decomposition analysis can be used when one wants to arrive at quantitative estimates on how various **key factors (drivers) influence greenhouse gas emissions**. In the scope of ex-post evaluation, one may be interested in how various drivers contributed to the development of total greenhouse gas emissions or to specific emissions (such as CO_2, SO_x, CH_4), either over the whole economy or in different sectors (such as buildings, transport, agriculture). These drivers do not directly depict policies and measures. The **effects of policies will be indirectly visible** through the changes in the drivers that they can be associated with.

Example on how effects of polices and measures become visible in decomposition results

An example of how effects of policies and measures become indirectly visible via decomposition results is depicted in Figure 12. The results of a decomposition of passenger road transport, hint towards the effect the **introduction of an environmental tax** (starting in 1999) had in Germany: the results indicate that a change in consumer choice on where to refuel, led to emission reductions by the driver "refuelling behaviour". After the introduction of the environmental tax, a significant number of consumers chose to refuel abroad where this particular tax was not applied. Before 1999, refuelling behaviour led to emission increases (the tax was still absent), while afterwards it contributed to emission reductions (see rose segment in the figure below). Before 1999 this segment was in the positive area (i.e. driving emissions) and after 1999 in the negative area (i.e. reducing emissions).

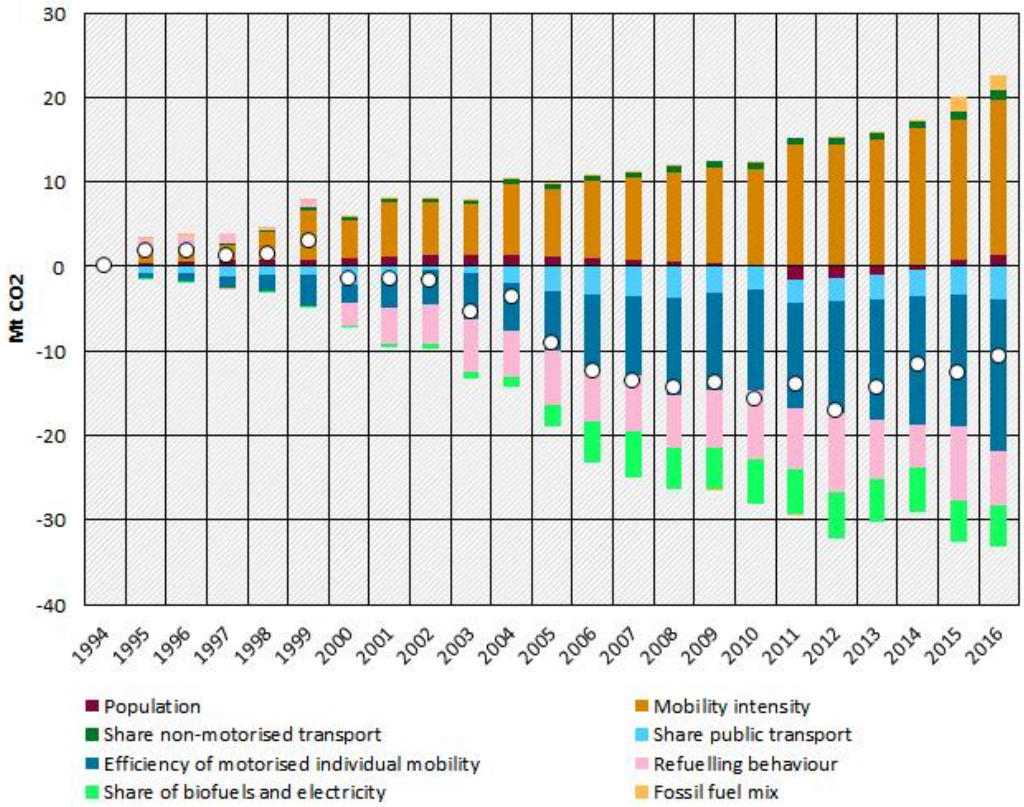


Figure 12. Example visualisation of decomposition results: decomposition of passenger road transport emissions in Germany vs. 1994 (translated from UBA, 2018, p.44).

Decomposition analysis can be used to analyse cumulated changes compared to a "reference year", (e.g. versus 1990) or year-on-year changes (such as 1991 vs. 1990; 1992 vs. 1991 and so on). Applying the first would emphasize for each year the effects compared to a "reference year" and thus on the effects as they cumulate (this is depicted in the figure above). This may be an interesting approach to learn how effects of policies and measures unfold over time and can likely be used for ex-post policy evaluation. The latter would show which drivers explain emission changes compared to each previous year. Such an approach may also be interesting when one wants to gather insights about effects that may be subject to short-term fluctuations.

How to use it?

The following steps are needed to complete an ex-post decomposition analysis:

(1) Determination of governing function

Once the scope of the analysis has been identified - for example total greenhouse gas emissions - one needs to define an identity, the so-called governing function, which relates the variable of interest to its relevant drivers (see for example Ang 2004, 2005).

The drivers used in this identity have to be relevant and independent from one another. The governing function and the relationship between the drivers is true by definition and based on expert knowledge.

Only as many relevant drivers will need to be chosen that can sufficiently explain the development of the variable of interest. Consulting the literature may aid decision making into which drivers are relevant to the issue in question. As the governing function is an identity and true by definition one needs to make sure to only include relevant drivers.

Example of governing function

One example is set out below (EEA, 2019), where the first line shows the identity, and the second line the names of the drivers. Such a governing function is where the decomposition process starts.

Variables	*GHG* =	pop	·	GDP/ population	·	primary energy/ GDP	·	energy-related GHG / primary energy	·	GHG/ energy-related GHG
Driver names	=	*population*		*wealth*		*energy intensity of GDP*		*carbon intensity energy sectors*		*carbon intensity non-energy sectors*

(2) Choice of decomposition method.

Once the governing function has been set-up in step (1), one needs to choose which decomposition method to apply for the analysis. There is a wealth of methods available that exhibit various characteristics. One of the most important distinguishing features is whether the method allows a complete decomposition or also produces a residual term (i.e. the rest that cannot be explained by the other drivers). Ang (2004) provides an overview of various methods suitable in the context of energy, which can also guide in the context of Effort Sharing. The following diagram from this publication highlights the various methods available for decomposition analysis.

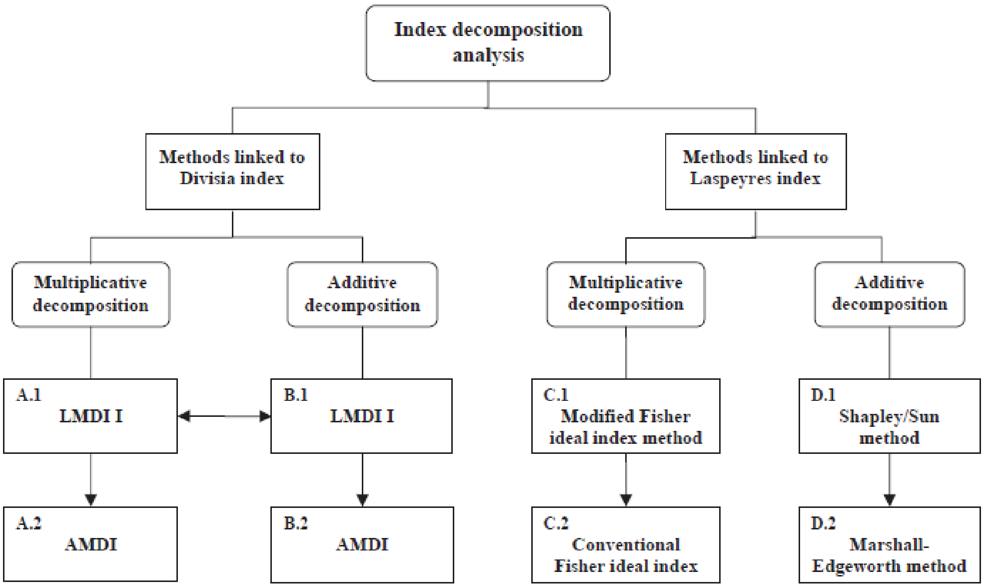


Figure 13. Recommended methods for energy index decomposition analysis (Ang, 2004).

According to (Ang, 2004), decompositions (in the energy field) can essentially be divided into two approaches: methods on the basis of a division index as well as on the basis of the Laspeyres index [7]. (Ang, 2004) evaluates the different options according to their theoretical basis, the adaptability to different questions, the user friendliness and the complexity of interpretation of the results.

Methods that meet these criteria, generate no or very small residuals and can deal with fluctuating, negative or near zero values as input. Both the **additive and the multiplicative log Mean Divisia Index (LMDI 1)** meet many of the most important requirements and are thus recommended by him for most applications. The broad use of this method in various publications underpins its relevance. See for example: (EEA, 2018), (UBA, 2018), (EEA, 2017), (ICF, 2016), (Förster et al, 2013).

(3) Construction of the "decomposition tool"

Once the governing function and decomposition method have been determined, one needs to set up a tool that conducts the necessary computations. Most of the time, such a tool can be created with Excel functionality and does not require knowledge of programming languages or the use of databases. For the LMDI method, for example, (Ang, 2005) provides all formulas needed for the construction of such a tool.

[7] Divisa indexes are built along the weighted sum of logarithmic growth rates. The weights are the components' shares in total value. Laspeyres methods measure percentage change of a group of items over time. They use weights based on values from the base year.

Example decomposition tool

Assuming one would want to apply the additive LMDI decomposition method to the governing function from step (1), the change over time would be depicted as follows: the change in GHG over time is equal to the sum of the changes of the drivers:

Variables	*GHG =*	pop	·	GDP/ population	·	primary energy/ GDP	·	energy-related GHG/ primary energy	·	GHG/ energy-related GHG
Variables for decomposition	*GHG =*	*X1*	·	*X2*	·	*X3*	·	*X4*	·	*X5*
Depiction of change	***ΔGHG = GHG_t - GHG_0***	***ΔX1***	**+**	***ΔX2***	**+**	***ΔX3***	**+**	***ΔX4***	+	***ΔX5***

The tool would need to be populated with the following formulas to decompose the change over time and calculate the contributions of each of the drivers (X1-X5) to the overall change:

$$\Delta GHG = \sum_{i}^{5} \frac{GHGt - GHG0}{\ln GHGt - \ln GHG0} \ln(\frac{Xit}{Xi0})$$

$$= \frac{GHG_t - GHG_0}{\ln GHG_t - \ln GHG_0} \ln(\frac{X1_t}{X1_0}) + \frac{GHG_t - GHG_0}{\ln GHG_t - \ln GHG_0} \ln(\frac{X2_t}{X2_0}) + \frac{GHG_t - GHG_0}{\ln GHG_t - \ln GHG_0} \ln(\frac{X3_t}{X3_0}) + \frac{GHG_t - GHG_0}{\ln GHG_t - \ln GHG_0} \ln(\frac{X4_t}{X4_0}) + \frac{GHG_t - GHG_0}{\ln GHG_t - \ln GHG_0} \ln(\frac{X5_t}{X5_0})$$

Once the above formulas have been implemented into the tool, it is advised to populate the tool with **dummy data** first to make sure the functionality has been correctly set up, to make sure there is no residual left (i.e. the sum of changes of all drivers is equal to the change of the variable of interest). It may also be useful to not only include the results quantitatively, but also construct a figure of the results, such as the one on passenger road transport decomposition in Figure 12 where the circles depict the change in the decomposed variable from 0-t. The value of this circle is equal to the sum of the values in the corresponding bar, as each segment of the bar depicts the value for one driver's contribution to emission change between 0 and t. The figure below explains how this is reflected in the visualisation.

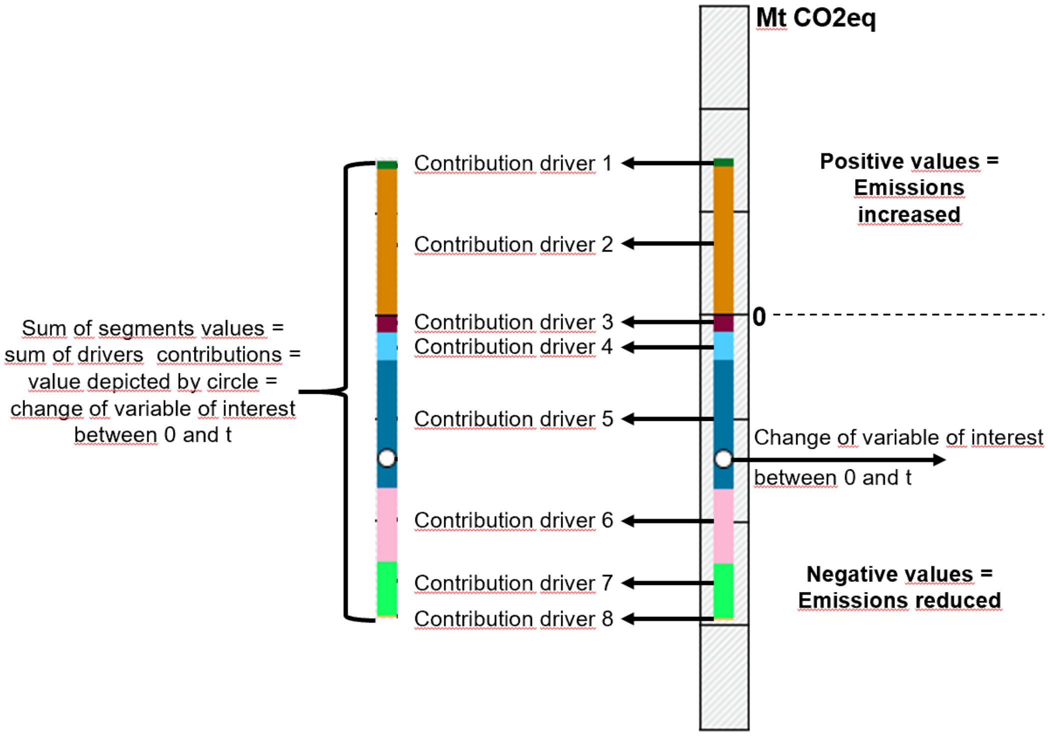


Figure 14. Logic of visualising decomposition results (Source: own illustration).

(4) Data gathering and population of the tool

The data required as inputs to the decomposition depends on the research question and scope of the analysis. Let us come back to the example in Figure 12, and assume the research question is to understand which drivers can explain the cumulated CO_2 emission changes in the passenger road transport sector between 1994 and 2016. The tool would need to be populated with data for the year 1994 and for the year 2016. The input data for such a decomposition would be (according to the approach taken in UBA, 2018): population, passenger transport (total), passenger transport (motorised), motorised individual passenger transport (MIT), final energy use MIT, total fuel sales for MIT, fossil fuel sales for MIT and CO_2 emissions MIT. In this case the cumulated change *between these two years, 1994 and 2016*, would then be attributed to the various drivers according to the chosen methodology (i.e. 2016 (last) bar in Figure 12). Depending on the scope of the analysis, the years in between may also be of interest and be populated with data. In Figure 12 this is the case, so it also informs on how emission driving/reducing effects *cumulate over time* between 1994 and 2016.

(5) Analysis of results

After the tool has been populated with data and results are available, one will need to **interpret these carefully**. The values of the drivers inform how much they have contributed to the overall change observed. For example, we see immediately from the results that the share of biofuels and electricity in road passenger transport (green segment in Figure 12) contributed to emission reductions. However, drivers in a governing function **do not directly depict a policy or measure**. Their values however may change due to policies and measures. This link will need to be made by the analyst using expert knowledge and consulting background information.

Staying with the example from Figure 12: after the introduction of the environmental tax, a significant number of consumers chose to refuel abroad where this particular tax was not applied. Before 1999, refuelling behaviour positively impacted emissions (the tax was still absent), while afterwards it contributed to emission reductions (see rose segment in Figure 12). Before 1999 this segment was in the positive area (i.e. increasing emissions) and after 1999 in the negative area (i.e. reducing emissions).

While for some drivers it may be straightforward to attribute the result to a specific policy and measure, for others it may not. Often, more than one policy and measure have been introduced within the timeframe of analysis. Thus, drivers will also include overlapping effects of policies and measures that cannot be disentangled.

Autonomous changes may also contribute to emission driving or reducing effects. For example, colder than usual winters may lead to increased need for heating and the use of less efficient power plants. This may lead to a fuel mix heavier on fossil fuel than in average years. Thus, the associated driver may lead to increasing emissions in some years, of a year-to-year analysis.

Advantages/disadvantages
The governing function to be developed is an identity based on expert knowledge (and literature and data availability). It is thus, by definition, true. Therefore, it is important to include only factors that are relevant and independent from one another (see also EEA, 2019). To analyse explicitly with decomposition analysis how a policy or measure has contributed to emission changes is challenging. Policies and measures themselves cannot be depicted by a driver, rather they impact drivers. Often, many policies and measures are present and their effects may overlap. Effects of one policy and measure may manifest in various drivers, too. Thus, interpretation of decomposition analysis results needs to take place carefully.

Characteristics	Score	Explanation
Data requirements	Medium – High	Depending on complexity of governing function
Complexity	Medium	Excel will suffice, but setup and defining governing function may be complex
Usefulness	Low- Medium	Effects can only be indirectly attributed to specific policies and measures; For some sectors, circumstances may allow for medium usefulness. E.g. if only one policy has been introduced or policy has no overlapping effects, or results clearly indicate link to a specific policy (such as changing refuelling behaviour as in the example given above).
Resources	Low-Medium	The resources needed depend on the complexity of the governing function and whether various experts are needed for setting up the case under question. Software choice may also drive resource needs.
Evaluation criteria	Low	Effectiveness can only be indirectly attributed to specific policies and measures.
Communication / visualisation of results	High	Results can be visualised in an appealing and easily understandable manner. Decomposition methods that do not yield a residual are specifically useful for communication of results.

Tools

At the time of writing the authors are not aware of an openly available tool to conduct decomposition analyses in a flexible and adjustable manner. The authors suggest consulting Ang (2005). This publication highlights the formulas needed to set up an LDMI decomposition analysis. The functionality for these formulas can be implemented in excel (this was done, for example, for all decomposition analyses published in (UBA, 2018)).

Data sources

Data sources are very specific depending on the decomposition question. The two case studies below may help to inform which data proves useful in these cases. Under *Want to know more* (UBA,2018) and (ICF, 2016) also provide such insights.

Related topics

Regression analysis
Assessing policy interactions

Case studies

- **Case study #2:** (Åström et al., 2017) The impact of Swedish SO2 policy instruments on SO2 emissions 1990–2012, Åström S., Yaramenka K., Mawdsleya I., Danielsson H., Grennfelt P., Gernerb A., Ekvalla T., Ahlgren E.O, Environmental Science & Policy Volume 77, November 2017, pp 32-39, https://doi.org/10.1016/j.envsci.2017.07.014
- **Case study #3:** (Reuter et al., 2019) Applying ex post index decomposition analysis to final energy consumption for evaluating European energy efficiency policies and targets, Reuter M., Patel M.K., Eichhammer W., Energy efficiency 12 (2019), No.5, pp.1329-1357, https://link.springer.com/content/pdf/10.1007%2Fs12053-018-09772-w.pdf

Want to know more?

- (Ang B.W., 2004) Decomposition analysis for policymaking in energy: which is the preferred method? Energy Policy 32 (9), S. 1131–1139, https://doi.org/10.1016/S0301-4215(03)00076-4
- (Ang B.W., 2005) Decomposition analysis: a practical guide, Energy Policy 33 (2005) 867–871, https://doi.org/10.1016/j.enpol.2003.10.010
- (EEA, 2019) Assessing GHG emission trends using decomposition analysis. EIONET climate change mitigation, 28 May 2019, EEA.
- (EEA, 2018) Trends and drivers in greenhouse gas emissions in the EU in 2016, EEA, 2018, https://www.eea.europa.eu/publications/trends-and-drivers-in-greenhouse
- (EEA, 2017) Analysis of key trends and drivers in greenhouse gas emissions in the EU between 1990 and 2015, EEA, 2017, https://www.eea.europa.eu/publications/analysis-of-key-trends-and
- (Förster et al, 2013) European energy efficiency and decarbonization strategies beyond 2030 – a sectoral multi-model decomposition, Förster H. et al., 2013, Climate Change Economics, https://www.researchgate.net/publication/259443348_European_energy_efficiency_and_decarbonization_strategies_beyond_2030_A_sectoral_multi-model_decomposition
- (ICF, 2016) Decomposition analysis of the changes in GHG emissions in the EU and Member States, ICF, 2016, https://ec.europa.eu/clima/sites/clima/files/strategies/progress/docs/dca_report_en.pdf
- (IPCC, 2014) Drivers, Trends and Mitigation, In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer O., Pichs-Madruga R., Sokona Y., Farahani E., Kadner S., Seyboth K., Adler A., Baum I., Brunner S., Eickemeier P., Kriemann B., Savolainen J., Schlömer S., von Stechow C., Zwickel T. and Minx J.c. (eds.)], Cambridge University Press, Cambridge, 2014, https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_summary-for-policymakers.pdf
- (UBA, 2018) (GERMAN): Komponentenzerlegung energiebedingter Emissionen, UBA, 2018 https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2018-07-02_climate-change_15-2018_treibhausgasemissionen_teilbericht3.pdf

4.6 Multi-criteria analysis

Description
A multi-criteria analysis (MCA) is a method that can be used to support a decision-making process when various options are being evaluated. In particular, an MCA compares alternative options on the basis of a range of different factors and from that process it allows decision-makers to select the best performing action (UK DCLG, 2009). Factors that are assessed typically include costs and benefits of each option as well as its ability to achieve various environmental, social and economic objectives (Urge-Vorsatz et al., 2014). An MCA allows for each factor to be measured in a different way and with different units. It can also include a weighting of each factor so that there is a set method for evaluation of the different indicators as well as a method for ranking the options (WRI, 2020).

When to use it?
MCAs are especially useful in cases where policy options may have different environmental and social impacts that are measured with different units. In this case an MCA can provide **a method for comparing different indicators and ranking the options** while providing a transparent rationale for a policy evaluation.

As an MCA uses a wide variety of assessment criteria, it is likely to also include values that are representing **stakeholder views** rather than objective data. By specifying how these will be measured and contributing to the overall assessment of options, an MCA can make these *"otherwise implicit [subjective judgments] explicit*" (Urge-Vorsatz et al., 2014). An MCA can therefore be useful in a situation where a transparent process is needed to assess a variety of stakeholder views. An involvement of stakeholders in the methodology setting process can show the sensitivity of results based on stakeholder perspectives (Dubash et al., 2013). This can be useful in the case where a wide range of stakeholders are involved and mostly qualitative data is available to assess the performance of policies. The stakeholders' involvement in the full process of an MCA can also ensure that its outcomes are accepted by a wide range of stakeholders.

It should be noted, however, that due to the nature of MCAs to assess different options, **MCAs are more typically used for ex-ante identification of suitable policy options** rather than for ex-post evaluations, where a single option has already been implemented. However, an MCA carried out to select a policy option could be revisited as part of an ex-post evaluation to see how the initial assessment may change when updated performance data of the policy is used and changes in circumstances are reflected in the assessment.

How to use it?

An MCA is typically developed by a number of steps, as outlined below (UK DCLG, 2009):

(1) The *context* for the decision is defined. In the case of an ex-post evaluation this includes the policy that is evaluated, the context in which the policy was implemented, and the stakeholders involved.
(2) The *possible policy options* that will be assessed and appraised are defined. In the case of an ex-post evaluation this may include different adaptations of the policy that is being assessed, e.g. its application in different sectors or otherwise a comparison of policies with similar objectives.
(3) The *assessment criteria* are defined. This will form a list of all the factors that will need to be assessed as well as the units in which they will be defined. This can result in a diverse list of factors to consider. For each assessment criteria a corresponding measurement scale will need to be defined that is expressed in a quantitative way. For example, the assessment criteria to what extent the policy option provides an incentive to invest can be scored on a scale from 0-5, whereby 0

represents the case that no incentive is provided and 5 a very high incentive is provided (Spyridaki et al., 2016).

(4) *Method for scoring and weighting* is developed. This is a process whereby a set methodology is defined by which the different assessment criteria can be combined into a ranking of options. This may include weighting of some assessment criteria, e.g. multiplying the results of some assessment criteria to ensure these have a more or less significant impact on the overall score and thus on the ranking of options.

(5) *Scoring of each of the options* listed against each of the assessment criteria. In most MCAs the scoring may be based on predicted outcomes of policy actions. However, in the case of an ex-post evaluation this may include actual performance data and stakeholder views on the performance of the policy in light of each assessment criteria.

(6) *Results of the scoring* will be added up following the methodology that was created as part of step 4. If relevant, the results can be assessed through a sensitivity analysis to show the dependence of the result on the variation of one of the scores or weighting in the method. The final results will show a ranking of the options considered.

While best practice may be to set a methodology for scoring and weighting in advance, the sensitivity analysis as part of step 6 may be used to adapt the weighting methodology.

Best practice examples show that the results are more meaningful to stakeholders in the case where stakeholders are involved in each of the steps outlined above. This can increase the validity of judgments made in the scoring process and create political consensus around the final results of the MCA (UK DCLG, 2009).

Advantages/disadvantages

Characteristics	Score	Explanation
Data requirements	Low-Medium	The data requirements of an MCA is directly related to the assessment criteria that are included. As it is possible to include assessment criteria with different units, the criteria can be designed based on the availability of data and its format. An MCA can therefore still provide meaningful results with low data availability, although more data may improve its validity.
Complexity	Low	The complexity of an MCA is low as a methodology can be designed that acknowledges the differences in data available.
Usefulness	Low	MCAs are typically used for ex-ante identification of suitable policy options, rather than for ex-post evaluations. Its applicability to ex-post evaluations may be limited as there are not always different policy options to be assessed.
Resources	Low	An MCA method can be designed based on the data and resources available. The general method of scoring, weighting and ranking is quick, and no additional resources are required.
Evaluation criteria	Medium	An MCA can evaluate the efficiency of policy options through the measurement of costs and benefits, as well as its effectiveness in terms of other assessment criteria aimed at identifying impacts related to the objectives. However, as indicated above, an MCA is designed to rank options and can therefore only be used in case different policies (or aspects of policies) are compared.
Communication / visualisation of results	High	Results of an MCA can be presented in tables that show the ranking of each option as well as their score against each assessment criteria. Colour coding may be used to show the variation of scoring for each option.

Tools

There are no specific tools for an MCA. A MS Excel spreadsheet may be used to sum and weight the scoring for each option and develop a table that ranks each of the options based on their scores.

Data sources

Data sources for an MCA can come from a variety of sources depending on the type of assessment criteria that are being considered, e.g. costs, emission reduction data, social impacts etc. Due to this high variability, no official data sources can be mentioned for MCAs.

Related topics

Methodologies for collection of evidence
Monitoring performance data and new data collection
Indicator analysis

Case studies

- **Case study #8:** Evaluating public policy instruments in the Greek building sector, Spyridaki N., Banaka S. and Flamos A., Energy Policy 88:528-543, 2016, https://www.sciencedirect.com/science/article/pii/S030142151530183X

Want to know more?

- (UK DCLG, 2009) Multi-criteria analysis: a manual, Department for Communities and Local Government UK DCLG, 2009, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/191506/Mult-crisis_analysis_a_manual.pdf
- (Dubash, et al., 2013) Indian climate change policy: Exploring a co-benefits based approach, Dubash N., Raghunandan D., Sant G. and Sreenivas, A., Economic and Political Weekly. 48. 47-61, 2013, https://www.researchgate.net/publication/283176798_Indian_climate_change_policy_Exploring_a_co-benefits_based_approach.
- (Spyridaki et al., 2016) Evaluating public policy instruments in the Greek building sector, Spyridaki N., Banaka S. and Flamos A., Energy Policy 88:528-543, 2016, https://www.sciencedirect.com/science/article/pii/S030142151530183X
- (Urge-Vorsatz et al., 2014) Measuring the Co-Benefits of Climate Change Mitigation, Urge-Vorsatz D., Tirado Herrero S., Dubash N.K. and Lecocq F., Annual Review of Environment and Resources Volume 39, 2014, https://www.annualreviews.org/doi/pdf/10.1146/annurev-environ-031312-125456
- (WRI, 2020) Greenhouse Gas Protocol – Policy and Action Standard. An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions, Wold Resources Institute WRI, 2020, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

5 Key evaluation issues and methodological challenges

Having considered evaluation approaches in the previous chapters, this chapter describes typical evaluation issues or challenges that Member States face and how they can be tackled. Some of these are priority evaluation issues which are specific to greenhouse gas policy evaluation (e.g. ETS versus non-ETS emissions, emission factors). Other issues are common to any domain of policy evaluation (e.g. policy interaction, consistency between ex-post assessment and projections, rebound effects). In the next paragraphs, guidance is given on both types of priority issue, illustrated by Member States' experiences from the bibliography.

5.1 Assessing policy interactions

Description

A specific policy or measure does not operate in a vacuum. They are implemented in a complex setting with pre-existing, multiple climate, environmental, energy and other policies. Policy interactions occur and are relevant whenever a targeted stakeholder acts differently in response to an intervention, either positive or negative, when confronted with these multiple other policies. In practice, the effectiveness and efficiency of almost all interventions will be affected, one way or another, by other policies and measures. In fact, these interactions can be and are being used to good effect to combine different actions that reinforce each other and increase the impact of these policy mixes beyond the sum of the impact of the individual actions.

Interactions can be **classified into three types**:

- Overlapping: In this type of interaction, the effects of policies and measures overlap and the combined impact of the group is less than the sum of the impacts of the individual policies and measures;
- Reinforcing: In this type of interaction, the effectiveness of a single intervention is increased by the existence of other policies and measures. The sum of the impacts of the individual policies and measures is smaller than the impact of the group. An example could be information and awareness raising campaigns that have a positive effect on the number of applications for an economic incentive;
- Neutral or independent: In this case, policies and measures do not have an influence on the impact of one another, even though they are related. The sum of the impacts of the individual policies and measures equals the impact of the group.

Note that policies might interact with one another in complex ways, having overlapping, reinforcing and/or neutral interactions at the same time.

Quantification and analysis of interactions between different policies and measures is also important for the evaluation of coherence of a policy. Evaluating coherence is about assessing the nature of these interactions.

In this section, we will look into how the interactions can be analysed and taken into account in the quantification of the outcomes of policies and measures.

How to address this?

(1) Assessing whether to evaluate a single or a group of interventions

Considering that interactions of different policy instruments are sometimes purposefully used to increase the effectiveness and efficiency, it does not always makes sense from an evaluation perspective, to try to split the outcomes or impact. It could therefore sometimes be better and more coherent to evaluate a group of policies and measures. This is very context dependent, so a first step is to assess whether to evaluate a single or a group of policies and measures. In some cases, evaluating a group of policies and measures results in a more coherent and transparent outcome than evaluating single policies and measures. It reduces the risk of double counting and already integrates the impact of the interactions the interventions have one another. Accessibility of data might also be a consideration for evaluation a group of policies and measures. In other cases, however, it can be very informative to understand and to quantify the role of single interventions within a mix of interacting policy instruments.

Table 16. Criteria for determining whether to assess policies or actions individually or as a package (Source: WRI, 2014).

Criteria	Questions	Guidance
Objectives and use of results	Do the end-users of the assessment results want to know the impact of individual policies/actions, for example, to inform choices on which individual policies/actions to implement or continue supporting?	If "Yes" then undertake an individual assessment
Significant interactions	Are there significant (major or moderate) interactions between the identified policies/actions, either overlapping or reinforcing, that will be difficult to estimate if policies/actions are assessed individually?	If "Yes" then consider assessing a package of policies/actions
Feasibility	Will the assessment be manageable if a package of policies/actions is assessed? Is data available for the package of policies/actions?	If "No" then undertake an individual assessment
	For ex-post assessments, is it possible to disaggregate the observed impacts of interacting policies/actions?	If "No" then consider assessing a package of policies/actions

Addressing interactions, even when analysing a group of policies and measures, is important as there could always be other policies and measures outside that group that have an interacting effect. Most Member States' examples of studying policy interactions involve the interaction of renewable energy and climate policies at EU level (e.g. between the EU ETS and the Renewable Energy Directive).

Example of grouped evaluation

EVALUATION OF ENERGY INVESTMENT MEASURES (DIALOGIC, 2017)

This study evaluated three incentives in the Netherlands to promote energy efficiency and renewable energy. The evaluation applies quantitative (such as econometric analysis) and qualitative methods (such as interviews and focus groups). Interviews and portfolio-analysis were used to assess the overlap between the three measures. These measures are not clearly positioned in relation to each other and this manifests itself in practice. Interviewees mentioned for example that projects could be eligible for different support measures.

(2) Mapping the policy interactions

Approaches typically start with listing all policies and measures that have a potential interaction effect on the impact of the intervention or policy that is being evaluated.

Secondly, the nature of the interactions can be further explored and analysed. To assist this, (WRI, 2014) suggests a **policy interaction matrix** can be made, which is a visual way to understand the interactions between combinations of policies. A separate matrix could be developed for each relevant parameter used to evaluate the policy. For each combination, a qualitative determination is made of whether the net interaction of policies is likely to be neutral, overlapping, or reinforcing with respect to the parameter. For each combination of policies, the general magnitude of the interaction should be recorded, for example major, moderate, or minor. This assessment should be based on **expert judgment**, published studies of similar combinations of policies/actions, or consultations with relevant experts. The type of instrument can already be an important determinant for whether an interaction will be neutral, overlapping, or reinforcing, as illustrated in the example below on energy efficiency policy instruments.

Table 17. Interaction effects of energy efficiency policy types (Source: Rosenow et al., 2016).

	energy or CO2 taxes	grants	loans	on-bill finance	tax rebates	Regulations	voluntary agreements	standards and norms	energy labelling schemes	information, advice, billing feedback, smart metering
Energy Efficiency Obligations	+	-	-	-	-	0	-	+	+	+
energy or CO2 taxes		+	+	+	+	+	+	+	+	+
Grants			-	-	-	0	0	+	+	+
Loans				-	-	0	0	+	+	+
on-bill finance					-	0	0	+	+	+
tax rebates						0	0	+	+	+
regulations							-	+	+	+
voluntary agreements								+	+	+
standards and norms									+	+
energy labelling schemes										+

+: complementary (savings from combination of policy A and policy B > than sum of savings policy A and policy B)

0: neutral (savings from combination of policy A and policy B > than sum of savings policy A and policy B)

-: overlapping (savings from combination of policy A and policy B < than sum of savings policy A and policy B)

(3) Quantitative and qualitative methods to assess interactions

In bottom-up calculations, often the most pragmatic approach is to allocate savings that could have originated from more than one policy or measure to only one. While this is a very rudimentary approach, it avoids double counting. This was the case for the policy *Energy Efficiency Agreement for Industries* in Finland, that overlaps with the Energy Audit Programme and where the monitoring database of the two schemes was used to remove double entries (EPATEE, 2017). Obviously, this approach is only possible when the overlap between the different interventions is complete and when the main aim is to understand the combined effects. It is not appropriate to understand the effectiveness of a single policy or measure.

Models that can deal with combinations of different policies have been applied to assess and quantify the impact of these interactions. These models can be either bottom-up or top-down models. An example is the study of CO_2-based tax incentives for new cars in the Netherlands, which applied a model to determine the effect of car tax rate on average CO_2 emissions (Kok, 2015). To differentiate the impact of two tax incentives (the company car tax and the vehicle purchase tax) different model runs were applied, with and without these taxes.

Example of models used to assess interactions

SIX YEARS OF CO_2-BASED TAX INCENTIVES FOR NEW PASSENGER CARS IN THE NETHERLANDS: IMPACTS ON PURCHASING BEHAVIOR TRENDS AND CO_2 EFFECTIVENESS (KOK, 2015)

This study examined tax changes in the Netherlands and assessed the impacts on consumer purchasing behavior as well as the CO_2 effectiveness of the tax incentives for low-carbon cars. To determine whether changes in car purchasing behavior, tax revenues and average CO_2 emissions can be explained by the gradual implementation of tax reforms and incentives, different types of model scenarios were developed. The study looked at three different fiscal instruments: the purchase tax, road tax and company car tax.

One aspect of the study was to explore to what extent different instruments contributed to the overall fiscal policy effect. The effect of the combined impacts of fiscal policies on average CO_2 emissions was already assessed based on modeling. The road tax was assumed to have no impact, while it was expected that the company car tax would have the biggest impact. Focus was therefore given to the company car tax and its impact on the average CO_2 emissions of cars in 2013 in relation to the total impact of Dutch tax incentives. To do this, the model was first used to determine the effect of a flat company car tax rate in 2013 (counterfactual scenario); and second, the actual CO_2-differentiated tax rate in 2013 (policy scenario). The flat tax rate resulted in an increase of the average CO_2 emissions of new passenger cars sold in the Netherlands of 9 g/km compared to the CO_2-differentiated tax rate. This corresponded with 70% of the total tax policy effect, which was 13 g/km in 2013.

Multicriteria evaluation has been applied in energy and climate policy impact evaluations as well. To address the interaction of policy instruments, it is investigated how the outcomes of the interactions between different interventions affect a number of criteria and variables (Oikonomou et al., 2007). In the Interact project, a methodology was used to assess and evaluate the appropriateness of different policy combinations using a multicriteria framework. The performance of policy options was assessed against a non-weighted set of standard criteria[8], using a simple numerical scale from 1= poor to 5=good.

Policy **theory evaluation** has been one of the main contributors in current research investigating interaction among policy instruments. Theory-based policy evaluation establishes a rational theory on how a policy instrument was intended to work, and therefore also accounts for its interrelationships with other policy instruments in the policy mix (Oikonomou et al., 2007). This was used in the Netherlands to evaluate 20 energy efficiency policies and measures (Harmelink et al., 2007). Drawing up a policy theory in practice included documenting all implicit and explicit assumptions in the policy implementation process and mapping the cause–impact relationship, including the relationship with other policy instruments.

[8] In this case, the criteria were: (1) environmental effectiveness: defined as the likelihood of the policy achieving a specific environmental objective; (2) static economic efficiency: defined as the potential to minimise the direct costs of meeting an environmental objective in the short term; (3) dynamic economic efficiency: defined as the potential to promote technological innovation; (4) administrative simplicity: defined as the administrative burden on both the target group and the implementing organisations; (5) equity: defined as fairness in burden sharing between the target group and other groups; (6) political acceptability: defined as the acceptability of the proposal by key groups in the economy.

Related topics

Intervention logic model
Counterfactual analysis
Systematic literature review
Interviews
Uncertainty

Case studies of approaches to address the issue

Given the low availability, no specific case study describing policy interactions in detail is included, however the short examples introduced above clarify current Member States' practices.

Want to know more?

- (CARISMA, 2016) Effects of Interactions between EU Climate and Energy Policies. Working Document Series No3, CARISMA, 2016, https://www.i4ce.org/wp-core/wp-content/uploads/2017/06/CARISMA-Working-Document-3-Policy-Interactions.pdf
- (Dialogic, 2017) Beleidsevaluatie Energie-innovatieregelingen, Dialogic, 2017, https://www.dialogic.nl/wp-content/uploads/2018/07/Eindrapport-Evaluatie-Energie-innovatieregelingen.pdf)
- (EPATEE, 2017) Energy Efficiency Agreement for Industries in Finland. Case study prepared by Motiva (Gynther L., Suomi U.) for the EPATEE project - funded by the European Union's Horizon 2020 programme, 2017, https://epatee.eu/sites/default/files/epatee_case_study_finland_energy_efficiency_agreement_for_industries_ok_0.pdf
- (Harmelink et al., 2007) Theory-based policy evaluation of 20 energy efficiency instruments, Harmelink M. et al., Energy Efficiency 1:131–148, 2007, http://harmelinkconsulting.nl/files/2015-09/harmelinkconsulting-c684629d6b049dc5320eba2fad1e60eb-theory-based-policy-evaluation-of-20-ene.pdf
- (Kok , 2015) Six years of CO2-based tax incentives for new passenger cars in The Netherlands: Impacts on purchasing behaviour trends and CO2 effectiveness', Transportation Research part A, (77) 137-153, Kok R., 2015, https://www.sciencedirect.com/science/article/pii/S0965856415000877
- (Rosenow et al., 2016) Energy efficiency and the policy mix, Rosenow J. et al., Building Research and Information, 44: 562-574, 2016, https://www.researchgate.net/publication/289540320_Energy_efficiency_and_the_policy_mix
- (Oikonomou et al., 2007) A framework on interactions of climate and energy policy instruments, Oikonomou V. & Jepma C.J., Mitigation and Adaptation Strategies for Global Change, 13:131–156, 2007, https://link.springer.com/article/10.1007/s11027-007-9082-9
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

5.2 Rebound effect

Description

Sustainable development requires a new approach to resource management. Increasing resource efficiency therefore becomes a central element of environmental programmes. However, actual greenhouse gas emission reductions are often lower than anticipated due to rebound effects. The rebound effect (or take-back effect) can be defined as the reduction in expected gains from an intervention that increases the efficiency of resource use, because of behavioral or other systemic responses. As a result, the theoretical impact an intervention could have is smaller than observed.

Rebound effects can be split into:

- **Direct rebound effects**: occur when a decrease in the cost of using a product results in an increased use of the product. Direct rebound effects have been described extensively for the transport sector and for residential heating.

  For example: More efficient internal combustion engines make it possible to build more economical vehicles. Direct rebound effects occur when the engines become more powerful or when the vehicle is driven more frequently.

- **Indirect rebound effect**: occur when a decrease in the cost of using a product results in increased use of other products or expenditure.

  For example: More efficient vehicles enable cost savings and the money saved could be used for an additional holiday flight. Part of the fuel savings in car traffic is thus offset by the additional fuel consumption in air traffic.

  Other relevant indirect rebound effects apply to producers where energy efficiency improvements led to changes in demand and productions. These are sometimes also referred to as embodied effects.

  For example: The manufacture and installation of energy efficient equipment, such as insulation materials, is associated with GHG emissions at different stages of the supply chain including production and travel.

- **Macro-economic rebound effect**: occurs when the initial savings from an intervention, result in a stimulated demand of the whole economy.

  For example: More efficient cars reduce the cost per kilometer driven making the use of cars more attractive. As a result, more households may buy cars and public transport may become less busy. Lower travel costs can also make a single-family home in the country more attractive. This would directly increase the distances travelled and can lead to an increase in living space, which in turn leads to higher energy consumption for heating.

The rebound effect can have a temporal dimension as well and a differentiation can be made between short-term and long-term rebound effects. Rebound effects can occur through a variety of **mechanisms** (Fisch & Grießhammer, 2013):

- Income effects: through efficiency measures money is saved which can be used to increase the use of the more efficient good (direct rebound) or of other goods (indirect rebound);
- Substitution effect: the price of the resource is lower due to the efficiency measure, which leads to the resource being used more intensively and effectively substituting other resources;
- Psychological effects: the efficiency measures produce a “green conscience” and in turn the same or other goods are used more;
- Technological rebound: the price reduction of a resource allows new technologies that require this resource to emerge which were previously not economically viable;
- Consumer accumulation: new, more efficient technologies are used additionally instead of replacing less efficient technologies;
- Macroeconomic or emergent effects: cumulative effects and interactions can be seen on the macroeconomic scale, which are hard to define methodologically and relate back to specific efficiency measures.

The principles of the rebound effects in relation to the mechanisms that produce them and the type of rebound effects that can be observed, are summarized in Figure 15 below.

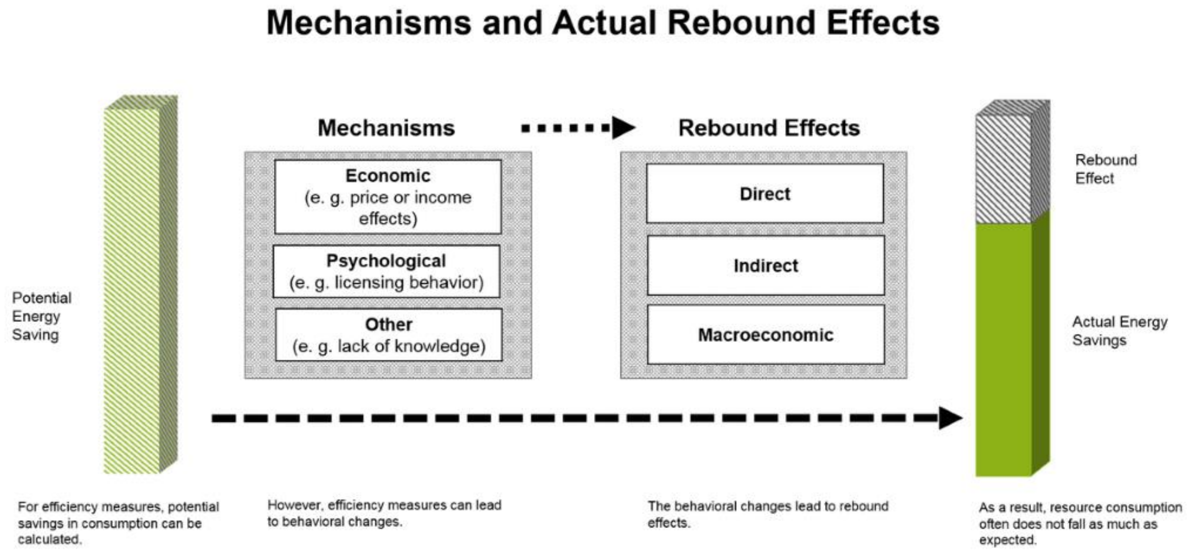


Figure 15. Schematic representation of the rebound effect (Source: EE Rebound, 2020).

Rebound effects can be very significant in certain sectors, reducing the total impact of an intervention. Evaluations that do not include rebound effects thus could **overestimate the impact of an intervention** on energy savings or avoided greenhouse gas emissions. Rebound effects have most often been quantified in relation to **energy efficiency improvements in households and in transport**, but also have been shown in interventions targeting waste and food. Determining the size of rebound effects is difficult, but existing studies show that *direct rebound effects* for energy use in households are between 10-30 %. Industrial production has direct rebound effects of around 15 % and in energy intensive industries between 20-60 %. Studies *concerning indirect and macroeconomic rebound effects* are rare and vary greatly. Sources quote rebound effects anywhere between 15-100 % (Fischer & Grießhammer, 2013). Rebound effects are not, however, necessarily negative. Consumer behavior can also change in a way that further resource savings are achieved. Such **sufficiency** (when within the same area) or **spill-over** (in other areas) effects are the opposite of direct or indirect rebound effects (EE Rebound, 2020). For example, if the purchase of a more efficient washing machine leads to an increased awareness of energy-efficient washing and machines are thus loaded better or washed at lower temperatures, this would be an example of sufficiency. Spill-over effects occur, for example, when purchasing a more economic showerhead leads to a better understanding of water efficiency and the purchase of water-saving fittings for the washbasin.

How to address this?

Several studies have quantified the rebound effect. These studies show that the size of the rebound effect is very **context dependent**, not only with respect to the sector and instrument type, but also to national circumstances (e.g. rebound effects are higher in lower income countries).

Direct rebound effects are easier to define and measure, because they are related to the demand for a specific product or service. In contrast, *indirect rebound effects* are more difficult to determine, because data on all resource demand from an individual or a household needs to be collected. Rebound *effects related to the whole economy* are extremely difficult to capture and there are few studies that undertake such an evaluation. The study by (Maxwell et al, 2011) highlights various methods of evaluating the rebound effect based on a literature review of existing studies. Evaluations of **rebound effects preferably take into account different types of rebound effects** and the various levels (micro- or macroeconomic) on which they can be observed (Table 18). Examples of these methodologies are briefly presented in the BOX below.

Table 18. Methodological approaches to measure rebound effects (Source: Maxwell et al, 2011).

Rebound type	Method of analysis
Direct	Micro-economic modelling of households/producers, including estimating price elasticities, income elasticities, etc.
Indirect	Micro-econometric/Macro-econometric modelling of households/producers: estimation of cross-price substitution elasticities (impact of a change in the price of one factor/good on the demand of the other factor/good)
Economy-wide*	Macro-economic models (often estimate behavioral relationships within an input-model (IO) structure) or Computable General Equilibrium (CGE) models

*Note: Economy-wide rebound is often measured jointly with indirect rebound.

It is advised to consider the rebound effect on multiple scales and to analyse it from different perspectives to provide a comprehensive analysis of rebound effects. Focusing only on direct rebound effects, or limiting the sectoral focus, gives an incomplete picture of magnitude of the rebound effect and could lead to a misinterpretation of efficiency gains (Madlener & Turner, 2016).

Examples of estimating the different types of rebound effects

(1) Direct Rebound Effect (Sorrell, 2007)

In order to estimate the direct rebound effect, most studies rely on secondary data sources that include information on, for instance, energy demand and energy efficiency. This data can take a number of forms (e.g. cross-sectional, time-series) and apply to different levels of aggregation (e.g. household, region, country). These studies use **econometric techniques to estimate elasticities**, namely the percentage change in one variable following the percentage change in another, when all other variables are hold constant. For the example of **energy efficiency**, the direct rebound effect may be estimated from one of two energy efficiency elasticities, depending upon the data availability:

- E1 the elasticity of the demand for *energy* with respect to *energy efficiency*;
- E2 the elasticity of the demand for *useful work* with respect to *energy efficiency*.

Most data sets, however, provide only limited data and variation in energy efficiency, meaning that estimates for E1 and E1 have a large variance or uncertainty. Instead, most studies estimate the rebound effect from one of three price elasticities:

- E3 the elasticity of the demand for *useful work* with respect to the *price of useful work*;
- E4 the elasticity of the demand for *useful work* with respect to the *price of energy*;
- E5 the elasticity of the demand for *energy* with respect to the *price of energy*.

In the case of personal automotive transport, the above elasticities could correspond to:

- E1 the elasticity of the demand for *motor-fuel* (for passenger cars) with respect to *kilometres per litre*;
- E2 the elasticity of the demand for *vehicle kilometres* with respect to *kilometres per litre*;
- E3 the elasticity of the demand for *vehicle kilometres* with respect to the *cost per kilometre*;
- E4 the elasticity of the demand for *vehicle kilometres* with respect to the *price of motor-fuel*; and,
- E5 the elasticity of the demand for *motor-fuel* with respect to the *price of motor-fuel*.

Estimates of E1, E2 and E3 require data on energy efficiency for the relevant energy service, while estimates of E4 and E5 require data on energy prices. Generally, the latter tends to be both more available and more accurate than the former. Similarly, estimates of E2, E3 and E4 require data on the demand for useful work, while estimates of E1 and E5 require data on the demand for energy. Again, the latter tends to be both more available and more accurate than the former.

Small and van Dender (2005) estimate the **direct rebound effect for personal automotive transport**. They use aggregate data on vehicle numbers, fuel efficiency, gasoline consumption, vehicle miles travelled and other variables for 50 US states and the District of Columbia covering the period 1961 to 2001. They use an econometric model explaining the amount of travel by passenger cars as a function of the cost per mile and other variables. By employing simultaneous equations for vehicle numbers, average fuel efficiency and vehicle miles travelled, they are able to treat fuel efficiency as endogenous: i.e. more fuel-efficient cars may encourage more driving, while the expectation of more driving may encourage the purchase of more fuel-efficient cars.

(2) Indirect Rebound Effect (Maxwell et al, 2011)

Indirect rebound effects are more difficult to estimate, because more variables and uncertainty are involved. One approach is to estimate these on a micro-economic (household) level, which can be done by estimating **cross-price elasticities**. This involves comparing price elasticities of different products and services to provide insight into how the price of one product can change consumption of another.

Another aspect of the indirect rebound effect is the **embodied effect**. This refers to the additional impact on GHG emissions through the manufacture or installation of, for instance, energy efficient equipment. Chitnis et al (2012) estimate the embodied effect using a life cycle analysis (LCA). They calculate the GHG emissions incurred in the manufacturing and supply of energy efficiency equipment. In their study, they examine seven measures that improve the energy efficiency of UK dwellings. The methodology is based on estimates of the income elasticity and GHG intensity of 16 categories of household goods and services. The embodied emissions are assigned to the year in which the measures are installed and divided by the total number of dwellings. This gives the average per-household embodied emissions for a measure. The study also estimate the average per-household embodied emissions of the relevant alternative. The difference between these two estimates is the embodied effect of the energy efficiency improvement. The study estimate that the embodied effect accounts for 10 – 67% of the rebound effect for different types of energy efficiency measures. While they can be relatively small for the measures considered in the study, the embodied effect should not be ignored when estimating rebound effects.

(3) Economy-Wide Rebound Effect (Maxwell et al, 2011; Hanley et al, 2010)

The economic effect more generally represents a sum of direct and indirect rebound effects. This will be influenced by a large number of variables, and is thus very difficult to estimate through econometric analysis of secondary data. **Computable General Equilibrium (CGE)** can be an approach to study the economic wide effects. CGE models are widely used in the estimation of the impacts of energy and climate policy. Modelling frameworks and associated benchmark data are also already available to the CGE models. They reflect structural and behavioral characteristics of economies and can demonstrate the impact of individual measures.

Hanley et al (2010) use a regional, energy-economy-environment computable general equilibrium (CGE) model to estimate what rebound effects can be expected from energy efficiency measures. Their CGE modelling framework is parameterized on data from Scotland and pulls together data from a variety of databases including the Social Accounting Matrix (SAM), an Input-Output team at the Scottish Executive, and the environmental (CO_2 and fuel use) database for Scotland. The rebound effect is estimated by identifying the impact of energy efficiency stimuli on the total use of energy. If total energy use falls proportionally by less than the increase in efficiency, there is rebound. The extent of rebound can be calculated by comparing the scale of any reduction (or increase) in energy consumption with the scale of the efficiency stimulus. Energy consumption is estimated using the following indicators:

- Scottish energy consumption: total use of electricity (gigawatt hours) and total use of non-electricity energy (tons of oil equivalents);

- Share of electricity generated in Scotland using renewable sources (share of total electricity output, in gigawatt hours, from the renewable source sectors).

Energy or resource efficiency is indicated by the ratio of GDP per unit of energy; a rise in this ratio indicates an improvement in the sustainability of economic development. Additionally, carbon intensity (as GDP per unit of CO_2 emissions) is considered.
The CGE model indicates that improvements in energy efficiency in production sectors generates rebound effects, where energy use increases in response to efficiency gains and the ratio of GDP to CO_2 emissions falls.

Tools
There are no specific tools for assessment of rebound effects.

Data sources
No official data sources. Data needs to be obtained from a variety of sources depending on what kind of methodological approach is taken.

Related topics

Assessing policy interactions
Uncertainty

Case studies of approaches to address the issue
A comprehensive list of case studies that illustrate the rebound effect both to demonstrate its existence and magnitude, how these are measured, and associated limitations, can be found in Maxwell et al (2011). Similarly, an assessment report from Sorrell (2007) provides an extensive overview of rebound evaluation studies by sector and method in relation to energy efficiency.

Want to know more?

- (BEIS, 2019) Valuation of Energy Use and Greenhouse Gas – Background Information, Department for Business, Energy & Industrial Strategy UK, April 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/794738/background-documentation-guidance-on-valuation-of-energy-use-and-greenhouse-gas-emissions.pdf
- (Chitnis et al., 2012) Estimating direct and indirect rebound effects for UK households, SLRG Working paper 02-12, Chitnis M., Sorrell S., Druckman A., Firth S. and Jackson T., 2012, https://www.sustainablelifestyles.ac.uk/sites/default/files/publicationsdocs/slrg_working_paper_01-12.pdf
- (EE Rebound, 2020) What's the Rebound Effect? EE-Rebound Project, 2020, https://www.ee-rebound.de/englisch/rebound-effect/what-s-the-rebound-effect/
- (Fisch & Grießhammer, 2013) Mehr als nur weniger – Suffizienz: Begriff, Begründung, Potenziale, Öko-Institut Working Paper, Fish C. & Grießhammer R., 2013, www.oeko.de/oekodoc/1836/2013-505-de.pdf
- (Hanley et al., 2009) Do increases in energy efficiency improve environmental quality and sustainability?, Hanley N., McGregor P., Swales K. and Turner K., Ecological Economics Vol. 68 pp. 692-709, 2009, https://www.sciencedirect.com/science/article/pii/S0921800908002589.
- (Madlener et al., 2016) After 35 Years of Rebound Research in Economics: Where Do We Stand? In Santarius et al. (eds.), Madlener R. & Turner K., Rethinking Climate and Energy Politics, pp. 17-36. DOI: 10.1007/978-3-319-38807-6_2, https://link.springer.com/chapter/10.1007/978-3-319-38807-6_2
- (Maxwell et al., 2011) Addressing the Rebound Effect, a report for the European Commission DG Environment, Maxwell D., Owen P., McAndrew. L., Muehmel K. and Neubauer A., 26 April 2011, https://ec.europa.eu/environment/eussd/pdf/rebound_effect_report.pdf
- (Small and Van Dender, 2005) A study to evaluate the effect of reduced greenhouse gas emissions on vehicle miles travelled, Small K.A. and Van Dender K., prepared for the State of California Air Resources Board, the California Environment Protection Agency and the California Energy Commission, 2005, https://www.socsci.uci.edu/~ksmall/CARB%20Final%20Report.pdf
- (Sorrell, 2007) The Rebound Effect: an assessment of the evidence for economy-wide energy savings from improved energy efficiency, Sorrell S., a report produced by the Sussex Energy Group for the Technology and Policy Assessment function of the UK Energy Research Centre, October 2007, http://www.ukerc.ac.uk/publications/the-rebound-effect-an-assessment-of-the-evidence-for-economy-wide-energy-savings-from-improved-energy-efficiency.html

5.3 Free rider effect

Description

Policies and measures, in particular those that provide financial support, might lead to free-rider behavior. Economic agents make use of the financial support even though they would have taken the same investment or change in behavior without the support. Such free-rider effect can be full, i.e. the supported activity would have happened in the exact same way without the funding, or partial, i.e. the supported activity would have partly taken place, would have happened later or in a slightly different way without the funding (Fraunhofer ISI, 2018).

Free-rider effects reduce the effectiveness and efficiency of an intervention. They might be looked at from different perspectives: (1) the number of supported actions, meaning that a percentage of activities would have happened the same way or similarly or later without the support, (2) the costs (subsidies) that accrue for activities that would have taken place without the subsidy, and (3) the impact that would have occurred even without the subsidy. In some cases – if timing of a subsidy is anticipated and foreseeable – it might even lead to the effect that an activity is postponed until the subsidy is released. In these cases, anticipation brings delays in action and adverse effects.

To comply with the Better Regulation goal for policies and measures to achieve objectives and bring benefits at minimum costs, it is important to take free-rider effects into account and deduct free-rider effects from gross impacts, so to **report net effects**.

How to address this?

Free-rider effects can be seen within the larger context of establishing the attribution (or causal relationship) of an intervention and observed changes. In an ideal case, evaluators would know what would have happened if the intervention had not occurred (counterfactual scenario). It would then be possible to attribute the changes to the intervention. In reality, however, the counterfactual is often unknown and the impact of other factors (such as free-riding, but also of other interventions) is a challenge to be disentangled. Changes might be attributed to an intervention while in fact they might (partially) result from other interventions, structural changes, changes in underlying parameters (e.g. energy prices, scale effects, GDP growth) or free-riding etc. These factors are also affected by the intervention under consideration, as markets, technologies, and behaviors have been influenced by energy and climate policy over time, making it difficult to discern the free-rider effect. It is therefore considered important to clearly state which of these additional effects have been accounted for when assessing impacts of an intervention. A common approach is to distinguish gross effects (no correction for other influencing factors) from net effects (after correction for other factors) and clearly indicate which effects are included within the net assessment. Evaluating a mix of supplementary policies surrounding an intervention may also provide insights into free-riders that overlap across interventions.

Generally, the counterfactual might be established by comparing activities of players that received the intervention with those that did not, by comparing activities of players before and after receiving the intervention or by designing a hypothetical behavior without the intervention. In the specific case of free-riding, several methods can be applied to assess free-rider effects. These are summarized in Figure 16.

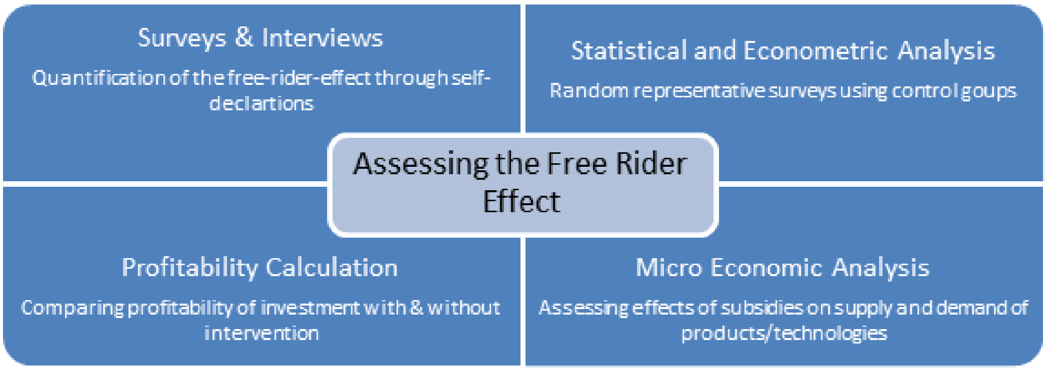


Figure 16. Assessing free-rider effects (based on Fraunhofer ISI, 2018).

(1) Surveys - interviews
Recipients can be asked whether they would have taken action even without the intervention, and if so to what extent, in what way and at what point in time. This would lead to a possible quantification of the free-rider-effect which would then be deducted from the total (gross) impact. As indicated in the section 'Surveys', care must be taken to the survey design, framing and formulation of questions to limit bias.

(2) Statistical and econometric analysis
The required data can be collected from different sources, such as, statistical sources, company reports, official household surveys or specific designed representative surveys to distinguish the action by the group of recipients from a control group that did not receive the intervention. Statistical methods based on these data can then be used to compare the changes in each group and attribute the impacts to the intervention (Wade et al., 2015). For example, when evaluating the impact of a subsidy programme on the installation of solar water heaters, a random survey can be taken to identify households that have installed solar water heaters and establish two groups: one group that made use of the subsidy programme and one that did not. Using statistical methods, the evaluation can then disentangle whether the subsidy programme itself led to statistically significant higher rates of solar water heater installation or whether an increase in installations was due to other circumstances. Similarly, econometric time series analysis can be used to understand whether the increase in installations is following a trend or whether additional installations compared to the trend were installed because of the subsidy programme. If the number of installations is following a trend, it can be deducted that the subsidy is subject to free-rider effects.

(3) Profitability calculation
Free-rider effects might be assessed by comparing the profitability (rate of return, payback time) of an investment with and without the intervention. If an investment is profitable even without the intervention, making use of the support might be considered free-riding. This concept for estimating the free rider impact was used in the Dutch evaluation of the financial instrument, Energy Investment Allowance EIA (case #5)

Example of free rider assessment using profitability calculations

DUTCH ENERGY INVESTMENT ALLOWANCE (EIA) 2012-2017 (CE DELFT, 2018)
The Dutch evaluation on the energy investment allowance scheme attempted to gain insights on potential free riding by asking participants through a survey what they would have done had the scheme not been implemented (Method 1). Unfortunately, the sample size was too small to draw conclusions. Instead they used a different method and looked at the payback times with and without the scheme. If the investment pays off already without the scheme, then the use of the scheme can be considered free riding. Applying this method, the study assessed free riding to be 29% (+/- 14%). The part of energy

savings achieved through free riding was then deducted from the overall gross energy savings to arrive at net savings.

It is important to note though that other (non-economic) barriers might exist that prevent a climate-friendly investment to be taken up and financial support policies might help overcome these barriers. For example, a municipality might consider investment in efficient street lighting with a high profitability. It can be argued that the investment pays off in a reasonable time (or with a reasonable rate of return) and thus no financial support is needed. However, due to overall capital constraints and other more urgent investment needs within the municipality, they might not be able to prioritize the investment in efficient street lighting. A financial support programme can help overcome this barrier and thus induce positive impacts that would not have happened otherwise.

(4) Micro economic analysis

Micro economic analysis can be applied to assess the effects of a financial incentive, such as a subsidy, on the supply/demand of products/technologies. It can be done using a detailed micro economic model or by using estimates for price elasticities of demand and/or supply to derive how demand/supply reacts in response to a change in price (costs) of a product or technologies due to the incentive. Low price elasticities (close to zero) imply rather rigid reactions, meaning that demand/supply of products is not very price reactive but rather depends on other factors. As a consequence, the incentive might lead to free rider effects because reactions would have happened independent of price or costs and the subsidy is just taken along. Higher elasticities imply rather flexible reactions, thus response to a change in price is rather high and an incentive can therefore set a positive signal for demand to rise. Fraunhofer ISI (2018) illustrates this with an example for a rebate for households who implement energy efficiency measures. They show that a substantial higher amount is paid for rebates that result in no additional energy savings in case of rather rigid elasticities, because the rebate makes no difference for households that are insensitive to costs. On the other hand, households with highly elastic demand for energy efficiency measures are more sensitive to costs and are more likely to implement additional measures based on the rebate.

Tools

There are no specific tools for assessment of free rider effects.

Data sources

No official data sources. Data needs to be obtained for individual aspects through surveys or other data sources as described above

Related topics

- Counterfactual analysis
- Surveys
- Interviews
- Assessing policy interactions
- Rebound effect
- Uncertainty

Case studies of approaches to address the issue

- **Case study #5:** Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904

- **Case study #7:** Free-riding on tax credits CIDD for home insulation in France: an econometric assessment using panel data, Nauleau M., Energy Economics vol. 46 pp. 78-92, 2014, https://doi.org/10.1016/j.eneco.2014.08.011

Want to know more?

- (ACEEE, 2019) Evaluation, Measurement, & Verification, ACEEE, consulted on November 2019, https://aceee.org/sector/state-policy/toolkit/emv
- (CE Delft, 2018) Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904
- (EPATEE, 2018a) Evaluating net energy saving - EPATEE topical case study illustrated with examples, EPATEE, 2018, https://www.epatee-toolbox.eu/wp-content/uploads/2018/10/epatee_topical_case_study_evaluating_net_energy_savings.pdf
- (EPATEE, 2018b) Specific guidance on evaluating Financial incentives on energy efficiency in Industry using the Unit consumption method, EPATEE, 2018, https://www.epatee-toolbox.eu/wp-content/uploads/2018/10/PSMC-17-1.pdf
- (EPATEE, 2018c) Terminology and typologies used in the EPATEE case studies, EPATEE, 2018, https://epatee.eu/sites/default/files/epatee_terminology_and_typologies_used_in_the_case_studies_v2.pdf
- (Fraunhofer ISI, 2018) How relevant are free-rider effects for target achievement? - updated version – IEPPEC conference, Fraunhofer ISI, 2018, https://energy-evaluation.org/wp-content/uploads/2019/06/2018-breitschopf-paper-vienna.pdf
- (Nauleau, 2014) Free-riding on tax credits for home insulation in France. An econometric assessment using panel data, Nauleau M., Energy Economics 46 pp. 78–92, 2014, https://www.sciencedirect.com/science/article/pii/S0140988314001923
- (Olsthoorn et al., 2017) Free riding and rebates for residential energy efficiency upgrades: A multi-country contingent valuation experiment, Olsthoorn, M. et al., In Energy Economics Vol. 68 Supplement 1, 2017, https://doi.org/10.1016/j.eneco.2018.01.007
- (Wade et al., 2015) Evaluation good practices: is 'good enough' better than 'perfect'? In ECEE summer study proceedings 2015 (8-099-15-Wade), Wade J. & Eyre N., 2015, https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2015/8-monitoring-and-evaluation-building-confidence-and-enhancing-practices/evaluation-good-practice-is-8216good-enough8217-better-than-8216perfect8217/2015/8-099-15_Wade.pdf/

5.4 Uncertainty

Description

Analysing and reducing uncertainty is an important element of policy evaluation, even though for ex-post analysis some sources of uncertainty are not as relevant than for ex ante assessments, as statistics on the policy impact might be available. Important **sources of uncertainty** for ex post evaluations are (identification of problems) (OECD, 2014):

- the counterfactual scenario, describing what would have happened without the policy or measure;
- available statistics and data might not be completely fit for purpose for an evaluation so additional assumptions might need to be taken, increasing the level of uncertainty;
- the outcome of any method for collecting information, either quantitative or qualitative, will inherently have a level of uncertainty;
- indirect effects, such as rebound or free-rider effects are often very uncertain, but could have a significant effect on the effectiveness and efficiency of a policy or measure.

Uncertainty is relevant for both quantitative and qualitative evidence. While for quantitative evidence, uncertainty could also be expressed numerically, for qualitative evidence other approaches might be needed[9]. Assessing uncertainty is often overlooked in policy evaluation. It is nevertheless crucial to gain **a better understanding of the robustness** of the findings and conclusions.

[9] Assessing uncertainty qualitatively can be done by characterizing the level of confidence of the results. This level of confidence can be based on the (1) quantity and quality of the available evidence, and (2) the degree of agreement of the evidence (WRU, 2015).

How to address this?

Two aspects are important while dealing with uncertainty for ex-post evaluations of policies: (1) quantifying the uncertainty of the evaluation results, and (2) reducing the uncertainty to improve the robustness of the evaluation findings. Both approaches are explained.

(1) Uncertainty analysis: How to quantify uncertainty?

According to the Better Regulation guidelines, the influence of key variables on the result could be investigated by a **sensitivity analysis**. These key variables should be allowed to vary in order to test the robustness of the final result and should be linked to the drivers of the problem identified. Possible ways to approach the problem of uncertainty analysis are (OECD, 2014):

- Worst/best case scenario analysis: this requires adopting all the most conservative and all the least conservative values for the variables used. It thus gives the two outer limits between which the actual/real result will lay. However, this does not give any indication of the probability;
- Partial sensitivity analysis: this is changing only some of the assumptions, but not others, namely for those key risk factors and underlying assumptions that are expected to be most uncertain;
- Monte Carlo sensitivity analysis: this is a more sophisticated technique that entails the creation of a probability distribution around key assumptions. While this is a more robust approach to sensitivity analysis, care needs to be taken in adopting reasonable and justified assumptions about the probability distributions. This type of analysis normally takes the form of a random sampling process to approximate the expected values and the variability inherent in the assumptions which are expressed as probability distributions for the most sensitive and uncertain parameters (risk variables). It is a computer-aided methodology through which many possible scenarios are generated by a random selection of input values from the specified probability distributions;
- Qualitative discussion: if the robustness of the basic assumptions cannot be examined numerically, a qualitative discussion on the appropriateness of each assumption can help readers to gauge the reliability of the results.

Monte Carlo analysis is not used often in policy evaluations, because it requires additional information not only on key parameters but also requires quantitative information on their variability. It is also a method that is more effective in more complicated assessments with many different key variables. In their evaluation of national policies and measures, the Belgian Federal government used Monte Carlo analysis in the evaluation of the tax incentive to promote energy efficiency in households (FPS Environment, 2017). Because information was only available on the number of applications for a tax reduction, assumptions had to be made on how this was split over the ten eligible technologies (e.g. condensing gas boiler, roof insulation, and PV) with different energy savings and emission reductions. In addition to this, energy savings and emission reductions per technology were also variable. A Monte Carlo analysis was done to estimate the uncertainty of the outcome. The results showed that for the most recent historic years 2011-2015, the 95% uncertainty range was approximately 10% of the average emission reduction over the period 2004-2015. This was the only policy measure for which a Monte Carlo analysis was done, uncertainty for other measures was assessed by partial sensitivity analysis. A French example of partial sensitivity analysis is explained in the BOX below.

Example of partial sensitivity analysis

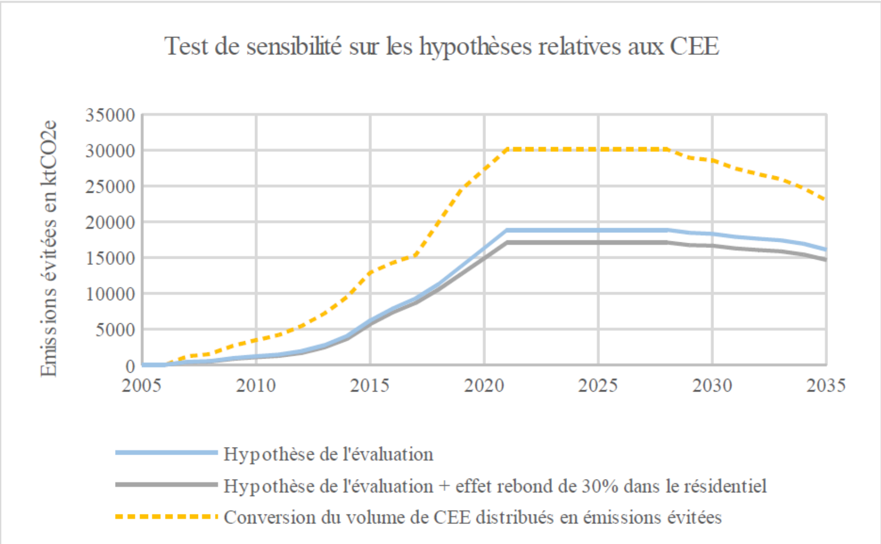
EVALUATION OF THE ENERGY EFFICIENCY CERTIFICATE SCHEME IN FRANCE (FRANCE, 2017)

When assessing the impact of the Energy Efficiency Certificate Scheme, it was recognised that there is a high level of uncertainty on the impact of the measures. This was related to the fact that detailed statistics were not available (uncertainty related to expected savings), the counterfactual scenario was unclear (especially the additionality of the measure compared to other measures) and indirect effects (most importantly the rebound effect) could not be quantified. Therefore a partial sensitivity analysis was done, performing the same calculation according to three different scenarios:

- One calculation corresponding with a best case scenario, assumed that all certificates resulted in avoided GHG emissions attributable to the measure;
- One calculation, the scenario selected as most likely, assumed that only 20% of energy and emission savings for the residential sector could be attributed to the Energy Efficiency Certificates, while the remaining part was attributed to other measures which provided a larger financial incentive. For other sectors, 100% of savings were allocated to the Energy Efficiency Certificates;
- One calculation based on the one above, but assuming a rebound effect of 30%.

This French case shows how a simple sensitivity analysis could be done and represented. The results give an insight into the uncertainty and how conservative the selected scenario is compared to alternatives.

Figure 17. Results of partial sensitivity analysis of French Energy Efficiency Certificate Scheme (France, 2017).



The data requirements of worst/best case scenario or partial sensitivity analysis are moderate and require only boundaries on all or most variables considered in the estimate. On the other hand, Monte Carlo analysis requires a far more in-depth understanding of the uncertainty linked to driving factors, such as probability distributions.

(2) How to reduce uncertainty and increase robustness of the findings of policy evaluations

The robustness of an evaluation is determined by the extent to which the design of the evaluation, the method of data collection and the analysis of the data lead to replicable answers to the evaluation questions posed (IOB, 2009). It relates to the extent to which results are replicable, i.e. that repeating the evaluation will produce the same conclusions and outcome. In order to do this, uncertainty needs to be reduced as much as possible and reported on transparently. There are several ways to increase robustness and reduce uncertainty of evaluations:

- Use different sources of information to address the same evaluation criteria and questions. This is often called **triangulation**. Combining different quantitative and/or qualitative methods to answer a specific evaluation question could result in: converging results that lead to the same conclusions, increasing the validity through verification (i.e. complementary results that supplement the individual results), and divergent/contradictory results that underscore the high uncertainty of a finding and can lead to new and better explanations;
- In the case of **own data collection** (e.g. surveys, monitoring), ensure that the selected approach is **sufficiently robust** so that results can withstand scrutiny. This means avoiding bias, having

sufficiently large and representative samples, etc. This is part of setting up a good methodology, and reducing uncertainty and increasing robustness is an essential part in that;

- Evaluate and assess quantitatively or qualitatively the uncertainty of the results of the policy evaluation and report on this transparently in the evaluation.

Related topics

Counterfactual analysis
Methodologies for collecting evidence
Rebound effect
Free-rider effect

Case studies of approaches to address the issue

Given the low availability, no specific case study describing in detail how to deal with uncertainty is included, however the examples introduced above clarify current Member States' practices.

Want to know more?

- (FPS Environment, 2017) Development of impact assessment methods and technical support for policies and measures carried out within the framework of the federal climate policy, Federal Public Service FPS Environment, 2017, https://www.klimaat.be/files/3315/0537/7367/Evaluation_federal_PAMs_July_2017_corr.pdf
- (France, 2017) Rapport de la France En application de l'article 13.1 du règlement n° 525/2013 relatif à un mécanisme pour la surveillance et la déclaration des émissions de gaz à effet de serre, France, Actualisation 2017, http://cdr.eionet.europa.eu/fr/eu/mmr/art04-13-14_lcds_pams_projections/pams/envwsc9fg/Report_2017_France_MMR_article_13_EN.pdf
- (HM Treasury, 2012a) Quality in policy impact evaluation: understanding the effects of policy from other influences (supplementary Magenta Book guidance), HM Treasury, 2012, https://www.gov.uk/government/publications/the-magenta-book
- (HM Treasury, 2012b) Quality in qualitative evaluation: a framework for assessing research evidence (supplementary Magenta Book guidance), HM Treasury, 2012, https://www.gov.uk/government/publications/the-magenta-book
- (IOB, 2009) Evaluation policy and guidelines for evaluations, IOB - Policy and Operations Evaluation Department – Dutch Ministry of Foreign Affairs, 2009, https://www.oecd.org/dac/evaluation/iob-evaluation-policy-and-guidelines-for-evaluations.pdf
- (OECD, 2014), OECD Regulatory Compliance Cost Assessment Guidance, OECD Publishing, 2014, http://www.oecd.org/gov/regulatory-policy/compliance-costs.htm
- (WRI, 2015) Monitoring implementation and effects of GHG mitigation policies: steps to develop performance indicators – Working paper, WRI, 2015, https://wriorg.s3.amazonaws.com/s3fs-public/Monitoring_Implementation_and_Effects_of_GHG_Mitigation_Policies.pdf

5.5 Emission factors

Description

A critical step when assessing the impact of a policy or measure on avoided greenhouse gas emissions is defining the appropriate emission factor to use. These emission factors translate changes in activities resulting from the policy intervention into changes in greenhouse gas emissions. This allows comparison of the impact among different interventions, irrespective of the instrument type or sector, and comparison of the intervention with the total or sectoral greenhouse gas emissions.

The unit of the numerator of the emission factor will be expressed as a weight or volume of greenhouse gases emitted, but the denominator can be very diverse. For example, emission factors may be expressed in terms of **energy output** (such as kg CO_2 emitted per liter of diesel consumed) or in terms of **physical output** (such as kg CO_{2eq} emitted per tonne of steel or cement produced) (WRI, 2015). Where emission factors are not readily available for the collected or monitored activity data, additional calculation steps will be needed based on relevant statistics or assumptions.

As with other steps in the calculation of the impact of a measure, the selection of the emission factor can have an important impact on the overall result and therefore needs to be selected carefully. There are several considerations evaluators need to take into account when selecting or determining the emission factors:

- *Greenhouse gases*: emission factors will depend on the greenhouse gases that are included in the assessment. Interventions often have an impact on more than one greenhouse gas, albeit that one greenhouse gas might be affected more than others. Evaluators can therefore opt to include all, most or only one greenhouse gas in their assessment. Therefore, it is important to **report transparently** or to highlight what is included or not.
- *Scope:* there are several ways to define the scope of the analysis and their respective emission factors. Does the evaluation only cover direct emissions or also indirect emissions from downstream or upstream effects, for example, and does the evaluation only cover domestic effects or also effects on emissions abroad?
- *Counterfactual scenario:* the counterfactual scenario is relevant for policy evaluation in general and thus can also be relevant for selecting the emission factor. More specifically, if the emission factor is expressed in physical output. The most appropriate emission factor is affected by the most likely scenario in case of no policy. For example in the transport sector, depending on the type of intervention and the most likely counterfactual scenario, the most appropriate emission factor could be the average emission factor of all cars in the stock or only the emission factor of new cars available on the market.
- *Time:* emission factors can be selected that are static or that are dynamic in time. The selection of which is more appropriate depends on the counterfactual scenario.

How to address this?

(1) Emission factors in terms of energy output

For all policies and measures that directly affect fossil fuel combustion, emission factors are available for each fuel type. **IPCC emission factors** are included in the 2006 Guidelines (IPCC, 2006), EU emission factors for different fuel types are also included in the EU ETS directive (Commission Regulation (EU) No 601/2012[10]), and national emission factors are included in the **National Inventory Report**. The difficulty here is often assessing how reductions or changes in energy consumption, relate to changes in the consumption of *specific* fuels.

Emissions factors used for policy evaluation are calculated relative to the **counterfactual scenario**. Therefore, the assumptions made with respect to the counterfactual will have a profound impact upon the emissions factor and the impact of the policy on emission savings. This is particularly an issue for the evaluation of savings associated with electricity production or consumption (see also AEA et al., 2009 for a more elaborate discussion). Emission factors for electricity are relevant in different cases: when renewable electricity or combined heat and power replaces conventional electricity generation; when energy efficiency reduces demand of electricity and in cases where mitigation measures increase electricity consumption (e.g. electric vehicles and heat pumps). For each of these cases different approaches might be appropriate to **determine the emission factor of electricity**. The following approaches might be considered (AEA et al, 2009):

- *Average emission factors:* EU or country specific average emission factors are the most straightforward way to calculate and use emission factors for electricity. The average may include all generation technologies or only a subset of available technologies. Sometimes must-run technologies, such as nuclear power, combined heat and power, renewables or electricity generated with blast furnace gas are excluded from the average.

[10] Mostly consistent with IPCC 2006 emission factors.

- *Short term marginal emission factors:* average emission factors are a simplification of the more complex effects additional renewable capacity or increased/reduced electricity demand will have on the market. To assess more accurately what kind of generation is being displaced, the emission factor can be based on the marginal power plant. This approach assumes that the marginal conventional power plant along the merit order curve is affected, so the most expensive technologies are displaced first. The emission factor thus depends strongly on the characteristics of the existing generation mix. Moreover, emission factors tend to be different for different periods of the day, week or season. There are different ways to calculate these marginal emission factors, based on statistical approaches or dispatch modelling, but all require detailed data on production and associated emissions.
- *Built margin emission factor:* in some cases, it could also be argued that the intervention does not displace existing power plant production, but displaces investments in new capacity. In this case, emission factors of the most likely new alternative technology could be more appropriate to use.
- *Long term marginal emission factors*: especially for interventions that are running for a long period of time, policies and measures will have more systemic impacts on the power sector. These might not be captured well by using average or short-term marginal emission factors. In this case, a combination of built margin emission factors and short-term marginal emission factors might be the most accurate approach.

Electricity has to be transmitted through a grid from the generation point to the consumption point, which means that transport and distribution losses occur. These can represent between 5% to 15% of the electricity transmitted. Another factor that confounds the calculation of emission factors is import and export of electricity. Therefore, adjustments can be done to the emission factors to account for electricity trade. Such adjustments are based on the share of electricity that is imported or exported compared to the domestic supply.

(2) Emission factors in terms of physical output

Emission factors in terms of physical output are much more diverse. These are needed for evaluations covering all possible emission sources and not only for interventions that affect fossil fuel combustion. The most complete sources for relevant emission factors are the **IPCC guidelines or online database** (IPCC, 2006; IPCC, 2020) and the **National Inventory Reports**.

Data sources

The National Inventory Reports and the emission factors used for drafting the annual emission inventory, are good starting points, as this increases consistency between the calculated, avoided greenhouse gas emissions and the reported national emissions. However, this might not always be appropriate.

Source	Geographical scope	GHGs	Emissions
(IPCC, 2006), (IPCC, 2020)	NA	All	Direct emissions
National Inventory Reports	EU-MS	All	Direct emissions
(JRC, 2017)	EU-MS	CO2, CH4, N2O	Direct and indirect emissions
(UK, 2019)	UK	CO2, CH4, N2O	Direct and indirect emissions
(ADEME, 2014)	France	CO2, CH4, N2O	Direct and indirect emissions
(Umweltbundesamt, 2016)	Germany	CO2	Direct emissions from fossil fuel combustion
(RVO, 2019)	Netherlands	CO2	Direct emissions from fossil fuel combustion

Related topics

Counterfactual analysis
Indicator analysis
Assessing co-benefits

Case studies of approaches to address the issue

Given the low availability, no specific case study describing emission factors in detail is included, however the short examples introduced above clarify current Member States' practices.

Want to know more?

- (ADEME, 2014) Documentation des facteurs d'émissions de la Base Carbone, ADEME, 2014, https://www.bilans-ges.ademe.fr/static/documents/%5BBase%20Carbone%5D%20Documentation%20g%C3%A9n%C3%A9rale%20v11.0.pdf
- (AEA et al, 2009) Quantification of the effects on greenhouse gas emissions of policies and measures, AEA et al, 2009, https://ec.europa.eu/clima/sites/clima/files/strategies/2020/docs/ghgpam_method_121209_en.pdf
- (IEA, 2019) Emission factors 2019 – database documentation, IEA, 2019, http://wds.iea.org/wds/pdf/CO2KWH_Methodology.pdf
- (IPPC, 2006) 2006 IPCC Guidelines for National Greenhouse Gas Inventories, https://www.ipcc-nggip.iges.or.jp/public/2006gl/
- (IPCC, 2020) 2020 IPCC Guidelines for National Greenhouse Gas Inventories, https://www.ipcc-nggip.iges.or.jp/EFDB/main.php
- (JRC, 2017) Covenant of Mayors for Climate and Energy: Default emission factors for local emission inventories, JRC, 2017, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC107518/jrc_technical_reports_-_com_default_emission_factors-2017.pdf
- (UK, 2019) Greenhouse gas reporting: conversion factors 2019, UK Department for Business, Energy & Industrial Strategy, 2019, https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2019
- (RVO, 2019) The Netherlands: list of fuels and standard CO_2 emission factors - version of January 2019, RVO, 2019, https://english.rvo.nl/sites/default/files/2019/05/The%20Netherlands%20list%20of%20fuels%20version%20January%202019.pdf
- (Umweltbundesamt, 2016) CO_2 Emission Factors for Fossil Fuels, Umweltbundesambt, 2016, https://www.umweltbundesamt.de/sites/default/files/medien/1968/publikationen/co2_emission_factors_for_fossil_fuels_correction.pdf
- (WRI, 2015) Monitoring implementation and effects of GHG mitigation policies: steps to develop performance indicators – Working paper, WRI, 2015, https://wriorg.s3.amazonaws.com/s3fs-public/Monitoring_Implementation_and_Effects_of_GHG_Mitigation_Policies.pdf

5.6 Assessing co-benefits

Description

The IPCC first used the term "co-benefit" in its 3rd Assessment Report11: "*Co-benefits refer to multiple benefits in different fields resulting from one policy, strategy, or action plan. Co-beneficial approaches to climate change mitigation are those that also promote positive outcomes in other areas such as concerns relating to the environment (e.g. air quality management, health, agriculture, forestry, and biodiversity), energy (e.g. renewable energy, alternative fuels, and energy efficiency) and economics (e.g. long-term economic sustainability, industrial competitiveness, income distribution)*."

There is a broad range of collateral benefits that can be associated with climate change mitigation policies in addition to direct avoided climate impact benefits, as illustrated in Figure 18. These collateral benefits are called "co-benefits" of climate change mitigation policies (OECD, 2009). For instance, a reduction of methane (mainly arising from agriculture sector) would lead to a reduction in overall greenhouse gas concentration as well as a decrease in tropospheric ozone concentrations, which have an important warming effect alongside detrimental impacts on human health and crop yields (local pollution). In the medium run, the only benefits of greenhouse gas mitigation policies are the co-benefits, since the direct benefits (avoided damage from climate change) are expected to occur in the longer run. Therefore, the integration of multiple objectives or benefits in policies can strengthen the support for climate policies and increase their cost-effectiveness (OECD, 2009). The co-benefits approach is a

[11] IPCC, Climate Change 2001, Mitigation, https://www.ipcc.ch/site/assets/uploads/2018/03/WGIII_TAR_full_report.pdf

positive and constructive **'win–win' way to operationalize how economic, social, ecological and political aspects** can be integrated within the concept of sustainable development. Different types of co-benefits for these aspects are illustrated in Figure 18.

However, co-benefit assessments are seldom applied in ex-post, nor ex-ante evaluations (Urge-Vorsatz et al., 2014). Ideally, for each evaluation, a full range of costs and benefits of all impacts should be considered. In practice, cost-benefit and cost-effectiveness analyses of climate change mitigation measures and policies deal with rather a narrow range of direct consequences and leave aside a broader range of indirect consequences. When indirect consequences do get included in the analyses, these are typically limited to mainstream benefits, such as air pollution reduction and health impacts.

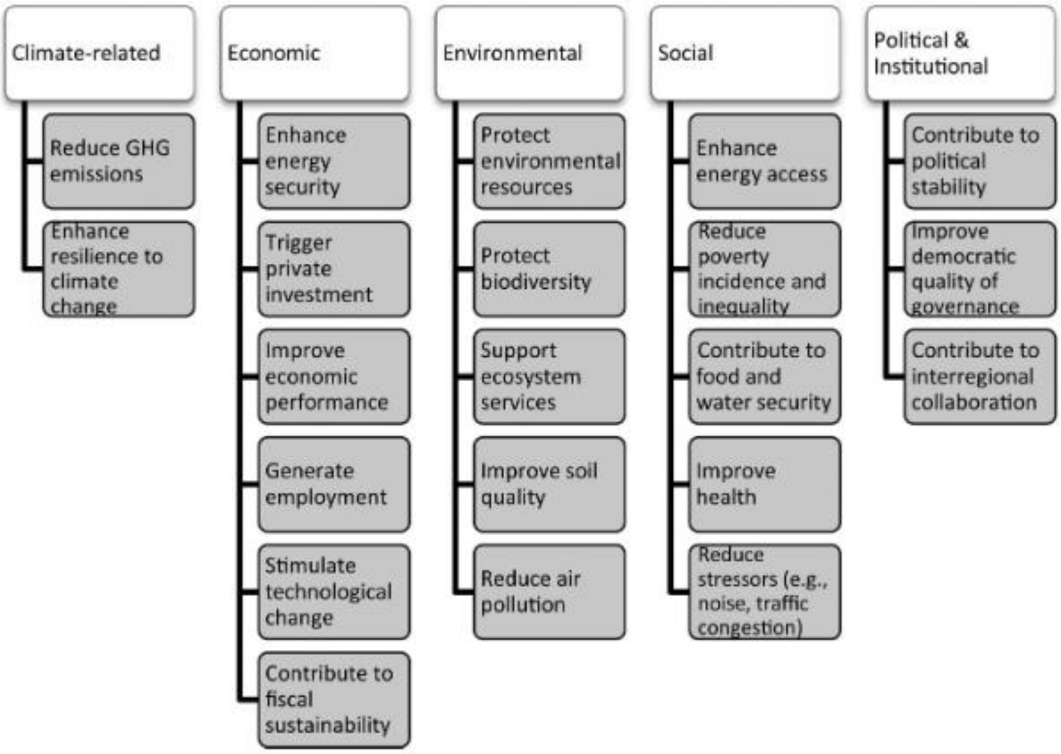


Figure 18. Type of co-benefits (Source: Mayrhofer et al., 2016).

How to address this?

There are a number of issues to consider in relation to the quantification of co-benefits (Urge-Vorsatz et al., 2014), among which:

- Context dependency: the quantification is extremely **context dependent**, or in other words, the sign and size of the impact of climate measures on welfare and other co-benefits depend strongly on local circumstances, how the policy is applied as well as the conditions under which the intervention took place. Therefore, it is difficult to estimate the impact of multiple co-benefits and to provide, generic, simplified methodologies for the assessment of co-benefits.
- Risk for double-counting: as some types of co-benefits overlap, special attention must be paid to **avoid double counting**. For example, improved air pollution resulting from investments in renewables or in energy efficiency affects household comfort, peoples' health, and workers' productivity; these three categories of co-benefits, at least partly, overlap.
- Complex dynamic relationships and feedback loops: evaluating distinct individual co-benefit categories may hide complex **dynamic relationships and feedback loops**. For instance, renewables and energy efficiency reduce air pollution, which decreases health care costs versus a baseline and may release public resources that can be invested or spent on alternative uses and,

as a consequence, further enhance employment or gross domestic product levels. In this regard, it must be pointed out that the overall positive and negative co-benefits of mitigation policies are seldom quantified or valued on a life cycle basis, with some environmental impacts left unaccounted for.

As illustrated in Figure 18, multiple types of co-benefits can be evaluated. The co-benefits of climate policies and measures are first quantified in physical units (e.g., avoided tons of pollutants released, life years saved, number of additional full-time jobs created, etc.). In a second step, these indicators can be translated in a monetary value (see section 'Cost Benefit Analysis'). For nonmarket goods and services, economic valuation methodologies are major concerns and shortcomings[12] related to the monetization, so it can be safer to use physical metrics for these co-benefits (Urge-Vorsatz et al., 2014). Table 19 synthesizes important **physical indicators used for the quantification** of different co-benefits of climate mitigation policies and measures. As indicated in section 'Indicator Analysis', it is of added value to already consider the type of indicators useful for evaluation purposes, while setting-up the monitoring system.

[12] Economic valuation methodologies have been criticized with arguments related to the commodification of ecosystem services or to the ethical implications of differences between the value of life in countries and regions with different income levels.

Table 19. Physical indicators used for the quantification of different co-benefits of climate mitigation policies and measures (Urge-Vorsatz et al., 2014).

Category of co-benefits	Subcategory of co-impacts	Physical indicators
Health benefits	Outdoor air pollution related Indoor air pollution related Energy poverty related Outdoor noise related Transport and traffic related Heat island related	Avoided cases Avoided hospital admissions Restricted activity days Years lived with disability Disability-adjusted life years (DALYs) Quality-adjusted life years Years of life lost
Energy poverty and distributional effects	Access to modern energy services	Additional kWh of quality energy (e.g. electricity) consumed Households with modern energy services (e.g. connected to the electricity grid)
	Affordability of energy services	Decreased energy demand (e.g. kWh)
Comfort and living conditions	Thermal comfort	Increased indoor temperatures Increased percent of floor area heated
	Exposure to external noise	Decibels (dBs) of external noise avoided
Provision of ecosystem services		Hectares (ha) of ecosystem or units of ecosystem service flow (e.g. number of recreational visitors per year)
Damage to building materials		Frequency of cleaning and maintenance of buildings
Productivity	Performance of individuals and organizations	Increase in labor productivity
	Crop yields	Increase in crop yields (percent)
Energy security		Units of imported energy avoided (e.g. oil barrels)
Macroeconomic effects		Percentage points of additional gross domestic product growth (%) Additional full-time equivalent (FTE) positions created

Data sources

In relation to **multiple benefits of energy efficiency**, the MB-EE (Multiple Benefits of Energy Efficiency) tool[13] is available as part of the ODYSSEE-MURE project, to represent a quantitative indicators approach to measuring multiple benefits. It aims to show the different aspects of energy efficiency beyond energy savings and give a more holistic view on its benefits. The MB-EEs are classified into three groups: environmental, economic, and social –related benefits. This indicator set is applied for 31 countries (EU28 plus Norway, Switzerland and Serbia) to provide a comprehensive tool of multiple benefits. The indicators considered have a broad database, which has been compiled from several sources having different temporal and spatial coverage (Odyssee-MURE, 2018). Therefore, the collected indicators represent different quality levels, which are divided into 3 categories, as presented in Table 20. The first group (A) has a good temporal and spatial coverage as well as a solid methodological basis (e.g. final energy savings, fossil fuels savings, impact on renewable targets, supplier diversity and import dependency). The methods are based on an excellent database directly from Odyssee or Eurostat. The second group (B) of indicators consists of those with limited spatial and temporal coverage, while still being based on a good methodological foundation, namely indicators based on InputOutput-analysis, such as GDP effects, employment effects and the effect on public budgets, as these only cover a few countries. The third group (C) consists of indicators which might have a good temporal or spatial coverage, but suffer from the need for simplification because of the lack of suitable data, while the method is still valid (e.g. indicators calculating the local and GHG emissions as well health and well-being). The potential future improvements for these indicators are methodological refinements that take into account temporal and spatial changes in the systems under consideration. For these improvements, however, detailed data sets are currently often lacking.

Table 20. Overview of the categorization of indicators depending on their methodological quality and databases (Odyssee-MURE, 2018).

Indicator	Description	Category
Energy savings	Top-down and bottom-up savings from ODYSSEE or MURE database	A
Saving of fossil fuels	Typical split of final energy per sector	A
Impacts on RES targets	Lowering of the gap to targets defined based on energy consumption	A
GHG savings	Avoided GHG emissions	B
Local air pollution	Avoided pollutants (SOx, NOx, PM2.5, PM10 and CO)	B
Alleviation of energy poverty	Changes in the share of energy cost in household income for low income households	C
Health and well-being	Avoided deaths linked to local emissions	C
Disposable household income	Changes in the share of energy cost in household income	B
Innovation impacts	Revealed Patent Advantage (RPA)	A
Competitiveness	Revealed Comparative Advantage (RCA)	B
Turnover of energy efficiency goods		C
Impact on GDP	Change in Gross Value Added	B
Employment effects	Additional full time equivalents in relevant branches	B
Impact on energy prices	Elasticity of fossil fuels in European Context	B
Public budgets	Additional income tax revenue	B
Industrial productivity	Change in productivity through lower energy cost as part of GVA	C
Asset value	Additional asset value of commercial buildings	C
Energy security (A)	Supplier diversity(HHI)	A
Energy security (B)	Import dependency	A
Impact on integration of renewables		C

[13] https://www.odyssee-mure.eu/data-tools/multiple-benefits-energy-efficiency.html

Related topics

Cost effectiveness analysis
Cost benefit analysis
Indicator analysis

Case studies of approaches to address the issue

Given the low availability, no specific case study describing co-benefits in detail is included, however the short examples introduced above clarify current Member States' practices.

Want to know more?

- (HM Treasury, 2018) The Green Book – Central Government Guidance on appraisal and evaluation, HM Treasury) , https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf
- (IASS Potsdam, 2017) Mobilizing the co-benefits of climate change mitigation - Connecting opportunities with interests in the new energy world of renewables – Working paper, IASS Potsdam, 2017, https://publications.iass-potsdam.de/rest/items/item_2348917_7/component/file_2666888/content
- (Mayrhofer et al., 2016) The science and politics of co-benefits in climate policy, Mayrhofer J.P., Gupta J., Environmental Science & Policy Volume 57 Pages 22-30, 2016, https://www.sciencedirect.com/science/article/pii/S1462901115301064
- (Ministry Environment Japan, 2009) Manual for Quantitative Evaluation of the Co-Benefits Approach to Climate Change Projects Version 1.0, Ministry of Environment Japan, June 2009, https://www.env.go.jp/en/earth/cc/manual_qecba.pdf
- (Odyssee-MURE, 2018) A comprehensive indicator set for measuring multiple benefits of energy efficiency, Reuter M., Patelb M.K., Eichhammera W., Lapillonned B. & Pollierd K., Odyssee-MURE, November 2018, https://www.odyssee-mure.eu/publications/archives/multiple-benefits-measure.pdf
- (OECD, 2009) Co-Benefits of Climate Change Mitigation Policies: Literature Review and New Results, OECD, Economics Department, OECD Economics Department Working Papers, 2009/01/01, https://www.researchgate.net/publication/46456924_Co-Benefits_of_Climate_Change_Mitigation_Policies_Literature_Review_and_New_Results/link/54c80a0b0cf289f0cecf5a5f/download
- (Urge-Vorsatz et al., 2014) Measuring the Co-Benefits of Climate Change Mitigation, Urge-Vorsatz D., Tirado Herrero S., Dubash N.K. and Lecocq F., Annual Review of Environment and Resources Volume 39, 2014, https://www.annualreviews.org/doi/pdf/10.1146/annurev-environ-031312-125456

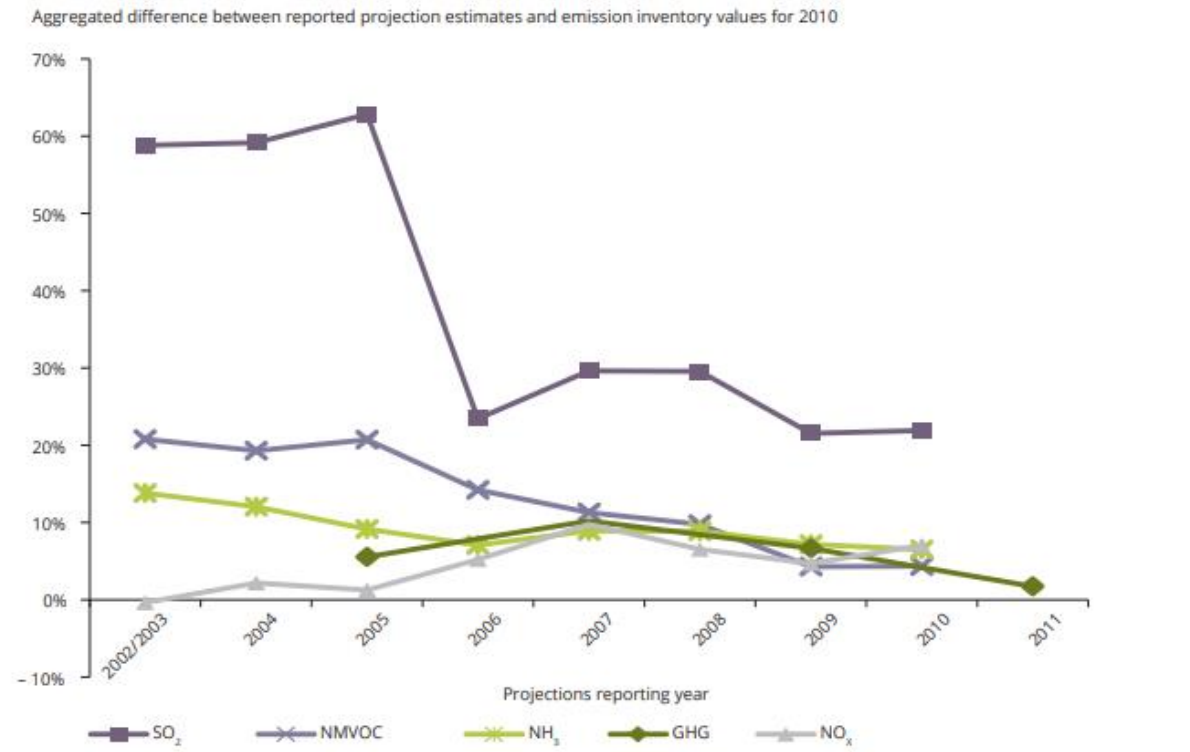
5.7 Consistency between ex post information and projections

Description

The findings from ex-post evaluations of climate policies can provide valuable insights on the effectiveness of those policies, as well as other similar policies, which can then be used to improve ex-ante projections of policy impacts in the future (see BOX). It is therefore important that results from ex-post policy studies are considered and integrated into the analytical work that underpins the development of emissions projections.

Aggregated underestimation or overestimation of projections reported by EU Member States

Analysis in the EEA's 'projections in hindsight' report highlights that there is a systematic trend that countries tend to overestimate their projected GHG emissions (EEA, 2015): the figure below shows the difference between reported 'projected' future emissions and reported 'actual' emissions for each year, with the GHG emissions in dark green. The EEA report identified several potential contributing factors to this trend, including limited or no reflection of new or changed policies in projections or of changes in effectiveness of implemented policies.



Note: 2010 GHG projections were reported biennially under the MMD, i.e. in 2005, 2007, 2008, 2009 and 2011, whereas air pollutions had to be reported annually under the NEC Directive. The average aggregated underestimation or overestimation does not include all 28 Member States for each year, but only for those countries that submitted national projections in that specific year. Croatia only joined the EU in 2013, and therefore did not report emission projections under the MMD or NEC Directive before, it is not included in this report.

Figure 19. The average aggregated underestimation or overestimation of projections reported by EU Member States (EEA, 2015).

However, there may be some **challenges to the effective integration** of the results from ex-post studies in ex-ante emissions projections, including:

(1) Ex-post impacts may be for a single policy instrument and derived using very disaggregated data (e.g. data on individual low carbon technologies). In contrast, projections may be derived at sectoral level based on more aggregated datasets. For example, the projections for the transport sector may take into account the influence of multiple drivers on road transport emissions (e.g. fuel prices, fuel efficiency, mix of vehicles) and policy packages (e.g. fuel taxes, road user tax, purchase taxes and subsidies) in an integrated way. However, the ex-post studies may look at one isolated element (e.g. effect of vehicle subsidies and infrastructure investments on the take up of electric vehicles).

(2) Ex-post impacts may be calculated on a consumption basis, whereas projections may be prepared on a production basis. For example, the greenhouse gas emissions savings that are associated with an energy efficiency policy may arise from a reduction of direct fossil fuel combustion in buildings and also from a reduction in electricity consumption. However, the projections may calculate emissions from the electricity production separately from emissions produced by direct combustion in buildings. As the ex-post evaluation data may not have the greenhouse gas emission savings separated per impact (e.g. buildings and electricity production), it may not be able to use its findings in the projections.

(3) Ex-post impacts may also include wider life-cycle emissions. For example, a policy to reduce emissions in the waste sector may deliver both *direct* emission savings from reduced waste to landfill, but also *indirect* savings associated with the increased use of recycling and reduced emissions from upstream industries. However, the projections may only account for direct

greenhouse gas emissions. It will therefore only be possible to use the findings from the ex-post study if these are disaggregated per category.

(4) Ex-post impacts will reflect the variables that were important historically but may omit key variables that will be important when projecting emissions in the future. For example, cooling demand will change as more buildings (will) incorporate air conditioning in the future.

How to address this?

To address the above challenges, the following principles should be considered when devising ex-post evaluation, and using the results from these evaluation in ex-ante projections:

(1) Align the **outputs** from the ex-post evaluations with the inputs to the projections modelling. This may include ensuring the scope of the emissions, and underlying activity data, are consistent with the inputs required by the projections model. It may also include aggregating/disaggregating the estimates, so that the scopes are the same. However, it should be noted that the scope for this may be limited due to methodological differences (e.g. projections are usually estimated top-down due to data availability, while ex-post evaluations typically use bottom-up data).

(2) Include some **past years** in the projections to enable the calibration of the policy impacts in the projections methodology with those estimated in the ex-post evaluation. If the projections include estimates of the policy impacts in past years, these estimates can be easily compared with those derived from ex-post studies. This can then inform how assumptions of projections may need to be modified to better reflect the policy impacts. For example, when evaluating its energy efficiency policies, the UK Government found that the actual impact of household insulation measures on energy consumption was less than the expected impact. This was found to be due in part to incorrect installation of the measures, but also in part due to the technical performance of the measures being less than expected. As a result of these findings the projected impact of the policy in the future was revised and scaled back to represent lower overall expected greenhouse gas savings (NAO, 2018).

(3) Align **assumptions on parameters** between ex-post and ex-ante assessments or correct for known differences. Some examples for specific sectors:

- Transport: compare fuel prices at time of ex-post analysis with assumptions for projections to highlight discrepancies. Use common sources for fuel price projections to ensure a consistent comparison, e.g. IEA World Outlook or Eurostat;
- Buildings: normalize past trends in weather data during ex-post analysis;
- Most sectors: growth assumptions as a function of GDP. Make sure that conditions under ex-post analysis are compared to wider trends assumed in projections in the case similar economic conditions are expected (e.g. no economic crises or are predicted to occur).

(4) Include **sensitivity analysis** in projections for factors that pose most uncertainty, i.e. to identify to what extent the results vary when the input parameters are changed from one extreme to another. These inputs could be economic, social and technological developments or other factors based on conditions identified in the ex-post studies. For example, different scenarios can be developed whereby both the highest level of innovation and technological development as observed in ex-post studies is assumed as well as the lowest level of innovation. The results of both scenarios could highlight how sensitive the outcomes of the study are in relation to the assumptions used. For the factors that are most sensitive, additional data may be collected from ex-post data to inform projections further.

(5) Harmonizing **time schedules** of different reporting requirements may also help with creating more consistency.

Tools

There are no specific tools for the consistency between ex-post information and projections.

Data sources

There are no official data sources for consistency between ex-post information and projections.

Related topics

Assessing policy interactions
Uncertainty

Case studies of approaches to address the issue

Given the low availability, no specific case study describing policy interactions in detail is included, however the short examples introduced above clarify current Member States' practices.

Want to know more?

- (BMWi, 2015) Zweiter Erfahrungsbericht zum Erneuerbare-Energien-Wärmegesetz, Bundesministerium für Wirtschaft und Energie (BMWi), 2015, https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/zweiter-erfahrungsbericht-erneuerbare-energien-waermegesetz.pdf?__blob=publicationFile&v=8
- (DG Environment, 2008) Assessment and improvement of methodologies used for GHG projections, DG Environment and VITO, Öko, IEEP, 2008, https://ec.europa.eu/clima/sites/clima/files/strategies/2020/docs/assessing_methodologies_for_ghg_projections_en.pdf
- (EEA, 2015) Projections in Hindsight: An assessment of past emission projections reported by Member States under EU air pollution and GHG legislation EEA Technical Report No 4/2015, European Environment Agency EEA, 2015, https://www.eea.europa.eu/publications/projections-in-hindsight
- (NAO, 2018) Programmes to reduce household energy consumption, National Audit Office, 2008, https://www.nao.org.uk/wp-content/uploads/2008/08/07081164.pdf
- (WRI, 2014) Policy and Action Standard - An accounting and reporting standard for estimating the greenhouse gas effects of policies and actions (GHG protocol), WRI, 2014, https://ghgprotocol.org/sites/default/files/standards/Policy%20and%20Action%20Standard.pdf

5.8 Splitting impacts between Effort Sharing and ETS

Description

For an appropriate description of the emission reduction effects of a policy or measure, whether these emission reductions are located under the European Emissions Trading System (EU ETS) or the Effort Sharing Decision or Regulation (ESD/ESR) is relevant. While emissions from activities included in the EU ETS are governed by the EU ETS legislation and subject to an EU-wide cap on emissions, those covered under the Effort Sharing legislation contribute to emission targets which are defined on national level: The Effort Sharing legislation sets annual emission trajectories for each Member State for the periods 2013-2020 (ESD; EU, 2009) and 2021-2030 (ESR; EU, 2018). These are translated into national annual emission allocations (AEAs) by implementing regulations. Member States should stay within the limits of their allocations. If they do not, they can make use of several flexibilities stipulated in the corresponding legislations, for example purchasing AEAs from other Member States. However, Member States generally try to avoid the purchase of AEAs from other Member States, as this is seen only as a last resort after the application of other flexibilities, like banking or borrowing.

The differentiation into emissions covered under the EU ETS and the Effort Sharing legislation means also that different actors are responsible to address reduction potentials from emission sources. In the coming years, Effort Sharing targets become more ambitious and thus the supply of AEAs from Member States with lower emissions than national targets decreases. With this, there is increasing interest in reducing emissions which count towards national emission targets under the Effort Sharing Legislation: As a result, the identification of policies and measures which effectively address emissions covered under the Effort Sharing legislation becomes a key interest of EU Member States.

Despite the importance, Member States face difficulties with the split into ETS and ESD emission reductions as it can sometimes only be an estimate, as the emission reductions cannot be clearly separated for some emission sources.

How to address this?

First, a general understanding of the coverage of the EU ETS and the Effort Sharing legislation needs to be achieved. To estimate the allocation of reduction effects under these policies, emissions and emissions reductions should be differentiated into emission source categories, following the greenhouse gas inventory logic. In a first step, the total ex-post emission reduction needs to be allocated to these source categories. With the knowledge about shares of emissions covered under both policies in these source categories, reduction effects can be differentiated accordingly.

General coverage and functioning of both policies

Emissions from large stationary installations, mostly from power and heat production and industrial installations, are covered by the EU ETS (EU, 2003). In addition, aviation emissions from flights between airports in the European Economic Area are covered under the EU ETS, too[14]. Emissions from these stationary installations and aircraft operators are reported annually and are published in the European Transaction Log (EUTL). Compliance takes place on an installation level one year after the emission took place.

Effort Sharing legislation covers emissions that are neither covered under the EU ETS nor related to the LULUCF sector. These emissions result from a broad range of activities, including road transport, energy consumption in buildings, agriculture (animals and soils), smaller industrial installations, smaller energy generation facilities and waste management. These emissions are not specified by emission source nor sector but are calculated from total GHG emissions without LULUCF emissions as reported in national GHG inventories, subtracting ETS emissions and domestic aviation emissions. Compliance takes place on a national level in the EUTL, several years after emissions took place.

[14] The legislation was designed to apply to all emissions from flights from, to and within the European Economic Area. The scope has been limited to flights **within the EEA** to support the development of a global measure by the International Civil Aviation Organization (ICAO). The next review of the ETS Directive should consider how to implement the global measure (CORSIA) in Union law through a revision of the EU ETS legislation. In the absence of a new amendment, the EU ETS would revert back to its original full scope from 2024.

Table 21. Emission source categories according to the reporting on GHG inventories: inclusion of ETS.

Emission source categories	Includes ETS emissions		
	CO_2	N2O	PFC
1. Energy			
A. Fuel combustion (sectoral approach)			
1. Energy industries	YES		
2. Manufacturing industries and construction	YES		
3. Transport	Domestic aviation and pipeline transport: YES Other transport: NO		
4. Other sectors	Residential: in general NO Commercial[15]: YES Agriculture/ Forestry / Fisheries[16]: YES		
5. Other			
B. Fugitive emissions from fuels	YES		
C. CO_2 transport and storage	YES		
2. Industrial processes and product use			
A. Mineral industry	YES		
B. Chemical industry	YES	YES	
C. Metal industry	YES		YES
D. Non-energy products from fuels and solvent use			
E. Electronic Industry			
F. Product uses as ODS substitutes			
G. Other product manufacture and use			
H. Other			
3. Agriculture			
4. Land use, land-use change and forestry			
5. Waste	In most Member States: NO		
6. Other			
Memo items:			
International bunkers			
Aviation	YES		
Navigation	Neither in ETS nor ESD/ESR		

[15] Depending on their size, following examples can be covered in the ETS: energy stations of universities, congress halls, big sport centers and swimming halls and power generation installation of hospitals which do not burn harmful waste.
[16] Depending on their size, big greenhouses can be, for example, covered in the ETS.

Allocation to emission source categories

The table above provides the different source category levels according to the reporting on GHG inventories. For each source category it is indicated whether ETS emissions occur, so it is clear whether a calculation of a split between ETS and Effort Sharing emissions is necessary after having calculated the ex-post emission reductions and having attributed them to a specific source category.

If a source category includes EU-ETS emissions, the Effort Sharing emissions share needs to be derived by taking the total GHG emissions (reported in greenhouse gas inventories for the source category) minus ETS emissions. It has to be noted that for the differentiation of emissions into ETS and Effort Sharing emissions, but also for the emission source categories from GHG inventories, often assumptions need to be applied. The GHG inventory does not inform about ETS and Effort Sharing emissions, nor informs the main activity type from installations covered under the ETS in an unambiguous way about their allocation to emission source categories. For national averages at the source category level, official estimates could be used for specific inventory years which are submitted via the reporting table on consistency between GHG inventory data and data from the emissions trading system (Article 14 of Commission Implementing Regulation (EU) 2020/1208).

Most ETS emissions occur from Energy Industries which are allocated in source categories 1.A.1 and 1.A.2. Other relevant ETS emissions are from industrial processes, allocated in sector 2. If the cell is empty in Table 21, this means that emissions allocated to these source categories are covered completely under the Effort Sharing legislation. In all other source categories, emissions are a mix of ETS and Effort Sharing emissions. The share of ETS-emissions in these source categories is country specific and changes over time. In some source categories, like in 1.A.4, the share of ETS emissions is usually very small and might be set to zero in case no other information is available.

Examples of splitting emissions reductions into ETS and Effort Sharing emissions

Policies or measures can address multiple sectors or source categories. Depending on the sectoral coverage, it is good practise to apply the following methods to split emission reductions into ETS and Effort Sharing emissions:

Example 1: If a policy or measure directly addresses **stationary installations** which are covered under the EU-ETS (for example the modernization of fossil public power plants), the whole emission reduction effect of the intervention falls under the EU-ETS. If emission reductions in typical Effort Sharing sectors (e.g. transport or building sector) are realized by electrification in these sectors (e.g. electric vehicles, heat pumps), this might lead to additional emissions in public fossil power plants – in case the additional demand is not covered by increased supply from renewable energy sources.

Example 2: If a policy or measure addresses **agricultural practices**, emission reductions fall completely under the Effort Sharing legislation.

Example 3: If emission reductions are allocated to source categories, in which **both ETS and ESD/ESR** emissions occur, it will only be possible in specific cases to clearly separate emission reductions into ETS and ESD/ESR emissions: If the intervention does not exclusively address or exclude installations which are covered under the EU ETS in these source categories, it is necessary to work with an estimated split between ETS and ESD/ESR. For this, average shares of ETS emissions need to be applied. The following steps can be followed to split emissions for a group of measures as well as effects of single policies or measures:

(1) Determine national ETS emission shares per source category

The national ETS shares per source category can be estimated based on the following data sources:

- To calculate national ETS shares in the above shown source categories, it is good practice to make use of the reporting under the MMR. Information on ETS emissions on source category level can be calculated from annual reporting in the tabular format provided in the Implementing Regulation No 749/2014 of 30 June 2014 - Annex V. If this information is available, these emissions can be compared to total GHG emissions in the relevant category.

- If Annex V tables are not available, ETS shares can be calculated from latest GHG projections which have been submitted under the MMR (EU, 2013). These GHG projections are reported in the tabular format provided in Annex XII of the Implementing Regulation No 749/2014 of 30 June 2014. Following this format, GHG projections are reported on source category level, separated into ETS and ESD/ESR emissions.

(2) Calculation of ex-post ETS emission reductions by source category

The reduction of ETS emissions is calculated by multiplying the determined emission reduction for a given year with the ETS share for each relevant source category.

(3) Calculation of ex-post ETS/ESR emission reductions per source category

ETS/ESR emission reductions are calculated as the difference between total emission reduction per source category and those emission reductions that were allocated to the EU ETS sector for that category, as estimated in the previous step.

If possible, the development of ETS shares should be considered on a source category level along the complete time series.

Tools

There are no specific tools for splitting the impact of ETS and Effort Sharing.

Data sources

Shares of ETS emissions by source category can be based on:

- National, annual reporting of ETS emissions by source categories (Implementing Regulation No 749/2014 of 30 June 2014, Annex V);
- If the above is not available, results from GHG projections as reported under the MMR, in the tabular format provided in Annex XII in Implementing Regulation 2014.

If national Annex V tables are used, the total GHG emissions per source category shall be taken from GHG inventories as reported to UNFCCC in the same reporting year.

Related topics

[Uncertainty]

Case studies of approaches to address the issue

No Member States' case study describing how to split the impacts between ETS and Effort Sharing is included. Although Member States' reports are available including the split of GHG ex-post effects of their policies and measures, the methodological explanation in the reporting is very limited. Moreover, the assumptions about the split of effects is country specific especially with regard to the energy sector, depending, among others, on the share of renewable energies.

Want to know more?

- (EU, 2003) Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC (OJ L 275, 25.10.2003, pp. 32-46).
- (EU, 2009) Decision No 406/2009/EC of the European Parliament and of the Council of 23 April 2009 on the effort of Member States to reduce their greenhouse gas emissions to meet the Community's greenhouse gas emission reduction commitments up to 2020 (OJ L 140, 5.6.2009, pp. 136-148).

- (EU, 2013) Regulation (EU) No 525/2013 of the European Parliament and of the Council of 21 May 2013 on a mechanism for monitoring and reporting greenhouse gas emissions and for reporting other information at national and Union level relevant to climate change and repealing Decision No 280/2004/EC.
- (EU, 2018) Regulation (EU) 2018/842 of the European Parliament and of the Council of 30 May 2018 on binding annual greenhouse gas emission reductions by Member States from 2021 to 2030 contributing to climate action to meet commitments under the Paris Agreement and amending Regulation (EU) No 525/2013 (OJ L 156, 19.6.2018, pp. 26-42).

6 Case studies

In this chapter, the guidelines for most of the methodologies or challenges are illustrated by case studies of ex-post evaluations of Member States' policies in Effort Sharing sectors. In this way, the guidance explains how Member States deal with **evaluations in practice.** It should be remarked that not all case studies presented are exclusively related to Effort Sharing sectors. They may also be related to other climate & energy policies (e.g. energy efficiency, renewable energy). Moreover, while the guidance includes several thorough case studies, there is no showcase example of a study that includes all methodological steps suggested in previous chapters. Therefore, a common finding is that many case studies apply simplifications and shortcuts to the proposed methodologies, given the challenges and constraints that Member States face.

Similarly to the methodology descriptions in previous chapters, the case studies are presented according to a fixed structure, as explained in the next BOX.

Description of case study

Methodology A	**Methodology B**	**Challenge A**
	☛ List of evaluation methodologies or methodological challenges covered by the case study.	
Description of policy measure	☛ Describing in short terms the policy measure and its context covered by the ex-post evaluation study.	
Scope of the evaluation	☛ Summarizing the scope of the evaluation study, such as evaluation criteria, research questions, time period(s), sectoral scope.	
Methodology	☛ Explaining the methodologies used for the collection of evidence as well as for the (quantitative) analysis.	
Outcomes	☛ Summarizing the main ex-post evaluation results of the case study.	
Want to know more?		
	☛ References to the original evaluation study or other sources giving more explanation or background on the case study.	

Case #1: Application of Surveys in Agriculture Policy Evaluations (United Kingdom and EU)

Surveys **(Systematic) literature review** **Interviews**

Counterfactual analysis

Description of policy measure

This case study focuses on two examples of surveys being used in evaluations of agriculture policies. In both cases, surveys were used to understand farm practices with the aim of finding the associated impact on greenhouse gas emissions.

The first case study comes from a review of the Greenhouse Gas Action Plan (GHGAP) in England, during the period 2012 through to the end of 2016. The GHGAP is a voluntary industry led programme, and the principal mechanism for delivering reductions in emissions from agriculture in England (DEFRA, 2017).

The second case study comes from a project for the European Commission to support the evaluation of the Common Agricultural Policy (CAP) to understand the impact of certain measures on reducing GHG emissions, agriculture's vulnerability to climate change and its ability to provide adaptation and mitigation services to society (EC, 2018).

Scope of the evaluations

A review of the GHGAP was conducted in 2016 to assess the effectiveness of the plan during the period 2012 through to the end of 2016. The review evaluated both the *impact* of the plan on GHG emissions in the sector, but also the strengths and weakness of the *implementation process*. One of the key sources of evidence for the evaluation was the annual Farm Practices Survey (FPS) (DEFRA, 2015). The FPS was initially developed as part of the 2012 review of progress in reducing GHG emissions from English agriculture. It collects information on a diverse range of topics related to the impact of farming practices on the environment (DEFRA, 2019).

DEFRA's GHGAP progress report focuses on seven key areas of activity identified in the GHGAP. These are: management, skills, advice and guidance; crop nutrient/crop health management; soil and land management; livestock breeding – genetic improvement potential; animal husbandry and improved health and welfare; energy efficiency and renewable energy; and cross sector initiatives. This structure is reflected in the survey as well, which has sections of questions dedicated to these key areas of activity. The review aimed to:

- Establish the level of progress being made towards industries' commitment to a reduction target of at least 3 million tonnes of carbon dioxide equivalent per year (3 $MtCO_2e$) by the end of the third carbon budget period (2022) (Effectiveness - impact);
- Highlight the strengths and weaknesses of the approach and recommend changes that could improve performance. Highlight the links between the GHGAP and Government's long term thinking for Food, Farming and the Environment (Effectiveness – process);
- Look ahead to the challenges that the GHGAP will need to address during the next phase of activity alongside other support from Government.

The evaluation of the CAP sought to understand the impact certain measures of the CAP have had on reducing GHG emissions, agriculture's vulnerability to climate change and its ability to provide adaptation and mitigation services to society (EC, 2018). The study focused on the period 2013-2018 when the current form of CAP has been in force. New surveys were used as part of the supporting study to gather data from a small sample of farmers from ten Member States about their experience of climate pressures and relevant CAP instruments and their uptake of different types of measures and innovation support activities.

Methodology

Surveys are an effective way to gather new data and to validate or enhance other data sources. In both the CAP and GHGAP review, surveys were used to gather data on the uptake of measures by farmers to reduce GHG emissions. The need for farm-level data on specific measures is important when assessing the impact of policies in the agriculture sector in particular, as higher level farming indicators (such as output) do not adequately describe **how** any improvements in productivity have been achieved. This is because farms can be very heterogeneous in their characteristics. Therefore, greater granularity and detail is needed to better understand progress in the sector, including the measures taken, and to estimate the impact on emissions.

Surveys were suited to this task, as in both cases there were a large numbers of implementing entities (farms) from which data was needed to estimate emission impacts.

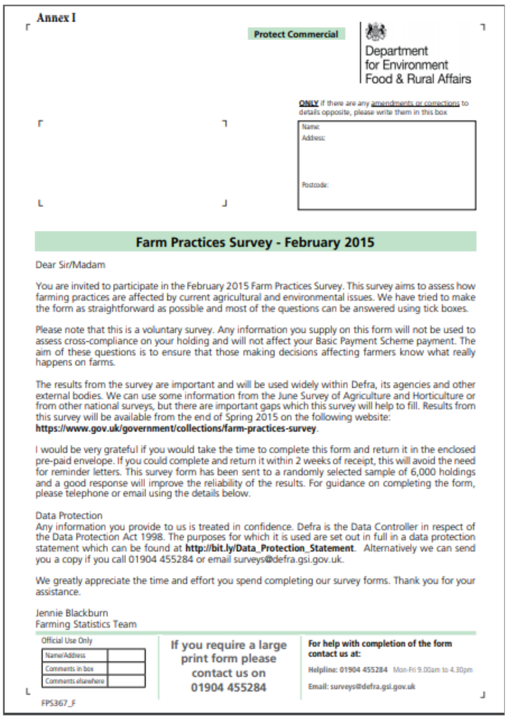
Survey distribution

In both case study examples, there were a large number of relevant stakeholders (farms) from which data was needed but it is not practical to attempt to collect data from them all. A sample of farmers was therefore targeted. Table 22 below summarises the different sampling approaches in the two evaluations. As the table shows, the FPS survey is designed to have a high confidence level (i.e. capture information that is representative of the full population) and is run in a similar way each year[17]. The representative sample of farmers is defined by the proportion of different farm types across England (cereals, other crops, dairy, pigs & poultry, grazing livestock, mixed). A 'stratified random sampling' approach is used, where the confidence level is determined for different groups (or 'strata') and a random sample is then taken for each group. This sample should then be representative of the respective group and its role in the wider population. A paper copy of the FPS survey is sent out to 6,000 English farms each year. An example of the format can be found at the end of the 2015 summary document (see Figure 20) (DEFRA, 2015). Guidance on completing the form is available via phone and email. Reminder letters are also sent via post, if the completed survey is not returned after 2 weeks.

The EC CAP survey ran via an online survey tool and was distributed to farmers, foresters and advisers via email. Each Member State's relevant Managing Authority undertook to administer the survey itself (except for HU, IE and NL). The survey was translated into each of Member States' native languages. It was necessary to offer the option to conduct the surveys over the phone, as some farmers had issues using the online tool. Unlike the FPS, the CAP surveys were not designed to be representative or to have a high confidence level. Instead, the evaluation notes that surveys were used to test that conclusions found in a literature review reflect the experiences of stakeholders. The surveys were also followed by interviews to better understand areas that need further clarification. In this way, surveys were combined with results from literature reviews and interviews to provide confidence and clarity.

[17] It is worth noting that the distribution approach to the FPS survey has been repeated in recent years. In 2019 DEFRA distributed it to 6.000 farms, and there was a similar response rate (~42%). This indicates this level of response rate is adequate for the analysis conducted.

Figure 20: FPS paper survey and accompanying letter 2015 (DEFRA, 2015).



Annex I

Protect Commercial

Department for Environment Food & Rural Affairs

ONLY if there are any amendments or corrections to details opposite, please write them in this box

Name:
Address:
Postcode:

Farm Practices Survey - February 2015

Dear Sir/Madam

You are invited to participate in the February 2015 Farm Practices Survey. This survey aims to assess how farming practices are affected by current agricultural and environmental issues. We have tried to make the form as straightforward as possible and most of the questions can be answered using tick boxes.

Please note that this is a voluntary survey. Any information you supply on this form will not be used to assess cross-compliance on your holding and will not affect your Basic Payment Scheme payment. The aim of these questions is to ensure that those making decisions affecting farmers know what really happens on farms.

The results from the survey are important and will be used widely within Defra, its agencies and other external bodies. We can use some information from the June Survey of Agriculture and Horticulture or from other national surveys, but there are important gaps which this survey will help to fill. Results from this survey will be available from the end of Spring 2015 on the following website: **https://www.gov.uk/government/collections/farm-practices-survey**.

I would be very grateful if you would take the time to complete this form and return it in the enclosed pre-paid envelope. If you could complete and return it within 2 weeks of receipt, this will avoid the need for reminder letters. This survey form has been sent to a randomly selected sample of 6,000 holdings and a good response will improve the reliability of the results. For guidance on completing the form, please telephone or email using the details below.

Data Protection
Any information you provide to us is treated in confidence. Defra is the Data Controller in respect of the Data Protection Act 1998. The purposes for which it is used are set out in full in a data protection statement which can be found at **http://bit.ly/Data_Protection_Statement**. Alternatively we can send you a copy if you call 01904 455284 or email surveys@defra.gsi.gov.uk.

We greatly appreciate the time and effort you spend completing our survey forms. Thank you for your assistance.

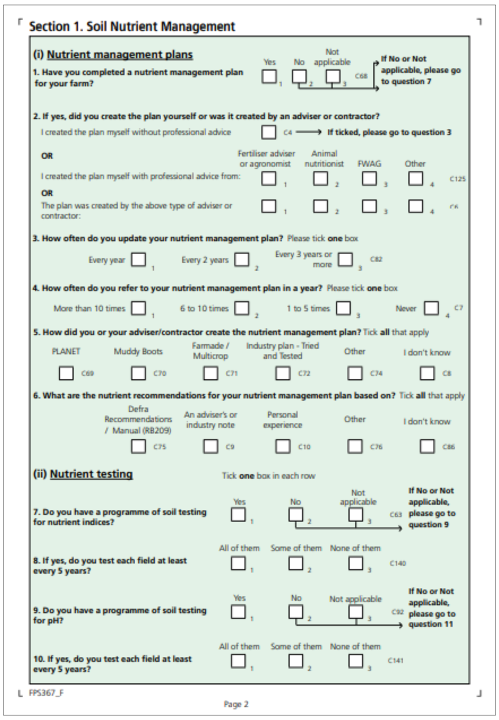
Jennie Blackburn
Farming Statistics Team

Official Use Only	
Name/Address	
Comments in box	
Comments elsewhere	

If you require a large print form please contact us on 01904 455284

For help with completion of the form contact us at:
Helpline: 01904 455284 Mon-Fri 9.00am to 4.30pm
Email: surveys@defra.gsi.gov.uk

FPS367_F



Section 1. Soil Nutrient Management

(i) **Nutrient management plans**

1. Have you completed a nutrient management plan for your farm? Yes ☐1 No ☐2 Not applicable ☐3 C68 — **If No or Not applicable, please go to question 7**

2. If yes, did you create the plan yourself or was it created by an adviser or contractor?
I created the plan myself without professional advice ☐ C4 → **If ticked, please go to question 3**
OR
I created the plan myself with professional advice from: Fertiliser adviser or agronomist ☐1 Animal nutritionist ☐2 FWAG ☐3 Other ☐4 C125
OR
The plan was created by the above type of adviser or contractor: ☐1 ☐2 ☐3 ☐4 C6

3. How often do you update your nutrient management plan? Please tick **one** box
Every year ☐1 Every 2 years ☐2 Every 3 years or more ☐3 C82

4. How often do you refer to your nutrient management plan in a year? Please tick **one** box
More than 10 times ☐1 6 to 10 times ☐2 1 to 5 times ☐3 Never ☐4 C7

5. How did you or your adviser/contractor create the nutrient management plan? Tick **all** that apply
PLANET ☐ C69 Muddy Boots ☐ C70 Farmade / Multicrop ☐ C71 Industry plan - Tried and Tested ☐ C72 Other ☐ C74 I don't know ☐ C8

6. What are the nutrient recommendations for your nutrient management plan based on? Tick **all** that apply
Defra Recommendations / Manual (RB209) ☐ C75 An adviser's or industry note ☐ C9 Personal experience ☐ C10 Other ☐ C76 I don't know ☐ C86

(ii) **Nutrient testing** Tick **one** box in each row

7. Do you have a programme of soil testing for nutrient indices? Yes ☐1 No ☐2 Not applicable ☐3 C63 — **If No or Not applicable, please go to question 9**

8. If yes, do you test each field at least every 5 years? All of them ☐1 Some of them ☐2 None of them ☐3 C140

9. Do you have a programme of soil testing for pH? Yes ☐1 No ☐2 Not applicable ☐3 C92 — **If No or Not applicable, please go to question 11**

10. If yes, do you test each field at least every 5 years? All of them ☐1 Some of them ☐2 None of them ☐3 C141

FPS367_F

Page 2

Table 22. Overview of sample approaches of both evaluation surveys.

	Stakeholders/ holdings targeted	Response rate	Total eligible holdings	Eligibility criteria	Sampling
CAP	Per country, aimed to collect information from: 50+ farms and 10+ farm advisors.	Low: on average per MS 10 farms and 8 advisors.	All holdings were eligible	NA	Targeted at 10 key member states. Non-representative sample.
FPS - GHGAP	6.000	40%	~60.000	>50 cattle, >100 sheep >100 pigs, >1.000 poultry or 20 hectares of arable crops/orchards.	Stratified random surveys

Other data collection methods

Surveys were used alongside other methods in both the CAP and the GHGAP studies. In the review of the GHGAP, interviews with the GHG Steering Group of representative industry bodies and with Farm advisors were conducted. It also used information from several key texts and previous work, including a review that took place in 2012 and 'The 7th edition of Agricultural Statistics and Climate Change' (which includes national level agriculture statistics and GHG information).

The review of CAP collected data from a detailed literature review, researched case studies and interviews with key stakeholders including farm advisers, representatives of the farming, forestry and wider rural sectors, government officials, climate researchers and NGOs. Both reports note how a combination of methods was used in order to strengthen the analysis, with the results of the surveys being **triangulated** with the other data sources.

Quantification of emission reductions
The data from the FPS survey was used to work out the rates at which mitigation measures have been taken up by farmers. This provided a measure of progress towards achieving the industry's ambition to reduce agricultural production emissions. The environmental impacts of actions were based on those reported in 'The 7th edition of Agricultural Statistics and Climate Change'. The EC CAP evaluation findings regarding the environmental effects were based on peer-reviewed literature on the effects of farming practices associated with the CAP greening measures on biodiversity, water, soils and climate.

Scaling up the survey results
The FPS is a stratified random survey which is targeted at a representative sample of farmers. However, there is no guarantee that that survey will be completed and returned by a respective sample. To produce national estimates, DEFRA therefore analysed results according to a standard methodology for stratified random surveys. With this method, the data from the different farm types are 'weighted according to the inverse sampling fraction' (DEFRA, 2015): the sampling fraction is the ratio of sample size to the population size, and the responses from different farm types are scaled according to the proportion of the farm types that exist at the national level. The EC CAP evaluation used a similar method for scaling up of the survey results. Subsequently, emission reductions were quantified by multiplying data from the survey on the uptake of CAP measures with scientific data on the emission impacts of those measures.

Defining the baseline scenario
Determining the reduction caused by a mitigation measure requires comparing the mitigation action scenario with a baseline 'business as usual scenario'.

As DEFRA's FPS survey had been conducted in 2012, it was possible to compare rates of uptakes of mitigation methods covered in the more recent survey.

The EC CAP evaluation found that there was not always a clear link between the use of a measure and the actual impact it has. While there is a large body of research establishing the potential of certain land management practices to contribute to climate mitigation, the evaluation sought to understand how the CAP had been used to support these mitigation actions. The mechanism to understand how effective CAP has been at doing so is less well established. For example, if a farmer receives advice on land management practices through CAP, it is hard to quantify the degree to which this measure encourages the farmer to changes their practices. The report highlights four main challenges in doing this:

- Lack of a baseline;
- The level of detail in some of the uptake data;
- The wide range of potential biogenic emissions depending on specific circumstances;
- Difficulties aggregating GHG emissions reported at project level.

The evaluation therefore developed a baseline. This was done using the GHG – Air Quality Interaction and Synergies (GAINS) model, which contains information about the expected evolution of GHG emissions and activity variables, and CAPRI, which is an agricultural sector-specific model, that contains an EU-wide unified data set on relevant statistics and activity variables. In combination, these were used to estimate and contextualise emissions mitigation from CAP. For all CAP measures with an expected quantifiable effect on GHG emissions, the uptake was multiplied by mitigation action factors taken from scientific literature. The product of these two variables constitutes a net emissions change. This was then applied to GHG emissions of their respective years using GAINS.

Assigning impact of a measure

Assigning the impact of a measure on emissions can be complicated. This is recognised in both the EC and DEFRA case studies. DEFRA's review involved contacting advisors from the UK Farm Advisory Service, and the report notes that all respondents indicated the difficulty of measuring impact. Judgment of the extent to which a certain measure, rather than other legislation, is causing the impacts under analysis is a further complication in assigning impacts. The evaluation of CAP highlights difficulties in establishing the impact of individual measures since multiple measures (both CAP and non-CAP) frequently apply to the same land. Due to these difficulties, impact was not explicitly assigned in both of these studies.

Outcomes

(1) GHGAP Evaluation

The GHGAP concluded that by early 2015, a 1.1 Mt CO_2 equivalent reduction in GHG had been achieved. DEFRA compared this to the 'maximum technical potential' which is the amount that could be saved if all mitigation potential was enacted. They estimated that around 31% of the estimated maximum technical potential was achieved by the end of 2015.

The report also found that the GHGAP had helped to drive the uptake of mitigation methods that have delivered the emission reductions. To achieve the target of 3MtCO2e by 2022, the report suggests that the GHGAP should push further uptake of mitigation methods already proving effective. It is thought it could also be used as a vehicle for identifying and driving uptake of new mitigation methods that will increase the potential for emissions reduction. The pace at which mitigation methods are taken up needs to increase, in line with the ambition for reductions in emissions.

The next full review of the GHGAP will be in 2020.

(2) CAP Evaluation

The evaluation found that Pillar II CAP measures implemented were responsible for a 1.1% emission reduction (~6.4 Mt CO_2eq/year) compared to the baseline of agriculture emissions. This reflects particularly low uptake for certain CAP measures (e.g. between 1-3% of planned expenditure for investments in physical assets had been committed at the time the evaluation was conducted in 2016). If all uptake targets set by Member States were met by 2020, the simulation in GAINS indicated that the Pillar II measures would reduce emissions by 1.5%.

In policy analysis, they found that significant opportunities have been missed to design the CAP in a way that would contribute more coherently and with greater relevance to climate objectives. In particular, this related to the choices made by MS in implementing the CAP 2014-2020 and by farmers. The study also found there was a lack of Government officials, farmer representatives or advisers with significant climate expertise or who understood how the CAP could be used to further climate objectives. The authors concluded that this lack of expertise would be a limiting factor in CAP's potential for climate action.

The report concluded that it was too early to see evidence of actual environmental impacts, as the study was conducted two years after implementation. It will however provide a detailed source of information for future reviews.

Want to know more?

- (DEFRA, 2015). FPS February 2015 - greenhouse gas mitigation - statistics notice, UK Government National Statistics and DEFRA, 2015, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/431938/fps-ghg2015-statsnotice-03june15.pdf

- (DEFRA, 2016) The Greenhouse Gas Action Plan for Agriculture Review, Department for Environment, Food & Rural Affairs DEFRA, 2016, , https://www.gov.uk/government/publications/greenhouse-gas-action-plan-ghgap-2016-review
- (DEFRA, 2017) The Greenhouse Gas Action Plan for Agriculture - Review 2016, Department for Environment Food & Rural Affairs DEFRA, 2017 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/599129/ghgap-2016-review.pdf
- (DEFRA, 2019) FPS - February 2019 - greenhouse gas mitigation - statistics notice, UK Government National Statistics and DEFRA, 2019, https://www.gov.uk/government/statistics/farm-practices-survey-february-2019-greenhouse-gas-mitigation-practices
- (EC, 2018) Evaluation study of the impact of the CAP on climate change and greenhouse gas emissions, European Commission (Directorate-General for Agriculture and Rural Development) and Alliance Environment (Mottershead, D.; Maréchal, A.; Allen, B.; Keenleyside,C; Lórànt, A.; Bowyer, C.; Brèche, O.: Martin, I.; Daydé, C.; Bresson, C.; Panarin, M.; Martineau, H; Wiltshire, J.; Menadue, H.; Vedrenne, M; Coulon, A), 2018, https://ec.europa.eu/agriculture/sites/agriculture/files/evaluation/market-and-income-reports/2019/cap-and-climate-evaluation-report_en.pdf

Case #2: The impact of Swedish SO_2 policy instruments on SO_2 emissions 1990-2012 (Åström et al., 2017)

Decomposition analysis **(Systematic) literature review**

Description of policy measure

The evaluation is focused on Swedish SO_2 policy instruments and their impact on SO_2 emissions from 1990 until 2012. Sweden introduced SO_2 policy instruments relatively early, which promoted cost-effective emission control measures (Lindmark and Bergquist, 2008) and reduced emissions substantially from 105 ktonne in 1990 to 30 ktonne in 2012 (Swedish Environmental Protection Agency, 2014a). Within the same time frame, the Swedish Gross Domestic Product increased by € 86 billion, which would imply a decoupling of SO_2 emissions from economic growth. However, Sweden is still having problems to reach their SO_2-related environmental policy targets and is expected to see too high values for SO_2 emissions until at least 2030 (Amann et al. 2014, Fölster et al. 2014). In order to develop effective instruments to achieve further reductions, the implemented instruments and their impact need to be studied. The decomposition analysis used here was done for SO_2 policy instruments of the Energy & Transport (ET) and Industrial Processes (IP) sectors. For the second part of the evaluation, the impact of individual SO_2 policy instruments was analyzed.

Scope of the evaluation

The objective of the study was to investigate the amount of decoupling of SO_2 emissions from economic growth that was due to SO_2 policy instruments in Sweden for the years 1990-2012. For this, a decomposition analysis was used. Further, the causality between emission reductions and the impact of individual policy instruments was determined by literature review and complementary mass balance analysis of oil imports[18].

Chronological correlations between changes in the emission requirements of the individual instruments and changes in implied SO_2 emission factors were determined to estimate the impacts of individual SO_2 policy instruments. For the study, official emission inventory data and publicly available data were used. This case-study highlights how an ex-post impact assessment on efficiency (emission reduction) of an intervention can be accomplished by combining: literature review, correlation analysis and decomposition analysis.

Methodology

Impact of SO_2 policy instruments on SO_2 emission decoupling

In order to evaluate the total impact of SO_2 instruments on the SO_2 emission decoupling from economic growth, a decomposition analysis using historical data was applied. The decomposition was based on (Rafaj et al., 2014) with the alteration that the analysis was disaggregated into separate calculations for the sectors Energy & Transport (ET) as well as Industrial Processes (IP). The different emission drivers were economic activity, fuel use, industrial productivity, emission factors. Equation (1) below reflects the ET sector, while (2) reflects the IP sector.

[18] The mass balance included a comparison of oil import statistics - starting from 1995 - with region-specific crude oil sulphur content. No significant impact could be found, as no support for the notion that the sulphur content in imported coal would have changed over time.

$$e_{sc,ET,t} = \sum_{ss,f} \left(GDP_t * \left(\frac{a}{GDP}\right)_{ss,f,t} * ief_{ss,f,t} \right) \quad (1)$$

$$e_{sc,IP,t} = \sum_{ss,p} \left(GDP_t * \left(\frac{a}{GDP}\right)_{ss,p,t} * ief_{ss,p,t} \right) \quad (2)$$

With:

e = emission levels;	sc = given scenario;
s = sector;	t = year;
ss = subsectors;	f = fuel activity;
p = product activity;	ief = implied emission factors;
GDP = Gross Domestic Product;	a = activity level.

The data and their sources used for the decomposition analysis are listed in the Table 23 below.

Table 23. Data used for decomposition analysis.

Data category	Data source	Data information
GDP	The World Bank, 2015	
Energy, transport & industry statistics	Swedish Environmental Protection Agency, 2014 (a,b,c)	Official reported Swedish emission inventories & industry's environmental accounts
Heat & electricity production	Statistics Sweden, 2015	Official Swedish statistics

Source: (Åström et al., 2017)

The data used allowed further disaggregation of the activity levels (a) into three factors for the **ET sector**, so that equation (1) was rewritten as follows,

$$e_{sc,ET,t} = \sum_{ss,f} \left(GDP_{t_1} * \frac{D_{ss,t_2} * Fs_{ss,t_3} * Fm_{ss,f,t_4}}{GDP_{t_2}} * ief_{ss,f,t_5} \right) \quad (3)$$

With:

D = total final energy demand;

F_s = ratio between total amount of fuel used and total final energy demand and;

FM = share of each fuel category in fuel mix.

To verify the alterations made in the decomposition analysis, two sensitivity analyses were executed.

For the **IP sector** the equation (2) was also modified into equation (4). The implied emission factors (ief) were expanded to capture the impact of increased productivity (eff).

$$e_{sc,IP,t} = \sum_{ss,p} \left(GDP_{t_6} * \left(\frac{a}{GDP}\right)_{ss,p,t_7} * (1 - eff_{ss})^{(t_8 - 1990)} * ief_{ss,p,t_9} \right) \quad (4)$$

Impacts of individual SO_2 policy instruments

The second part of the analysis dealt with the impact on emissions of individual SO_2 policy instruments. Therefore, chronological correlations between changes in the emission requirements of the instruments and changes in implied SO_2 emission factors were analysed. The sulphur contents allowed or emission limit values (ELV) were quantified as instrument-specific ELV pathways. Then, for the same activity and/or subsector the ELV pathways were compared to the development of the implied emission factor (IEF) pathways. If correlations between the ELV and IEF pathways occurred, factual and counterfactual SO_2 emissions (i.e. with and without SO_2 policy) for 2012 for the individual instruments associated with the IEF pathway were calculated. The difference of these two values was considered as the instruments lower bound impact on SO_2 emission decoupling.

The data on the Swedish Sulphur Law and other regulations to generate the ELV pathway were taken from different sources: the Swedish Environmental Protection Agency 1997, Regeringskansliet 2015, Rättsnätet 2015 and SPBI 2015. From (Svensson, 2003; Gillberg, 2015) information about environmental permit processes and decisions made in concession boards or environmental courts were extracted.

Outcome

Impact of SO_2 policy instruments on SO_2 emission decoupling

The decomposition analysis indicated that 43 % of the SO_2 decoupling from 1990 by 2012 was based on structural changes, 32 % were based on emission factor changes and 25 % based on fuel use changes for the ET sector. The decomposition for the IP sector showed similar results with 45 % of the decoupling due to structural changes, 31 % due to emission factor changes and 24 % due to increased productivity. In total, at least 26 % (approx. 36 ktonne) of the total decoupling of SO_2 emissions from economic growth were caused by Swedish SO_2 policy instruments.

Impacts of individual SO_2 policy instruments

The ELV and IEF pathways correlated well over time for three individual SO_2 policy instruments. The highest impact on 2012 emissions of the three instruments came from the environmental permit decision from 1996. This decision mandated the use of scrubbers in cement production by October 1998. The impact of the scrubber installation in one IP sector (cement plants) accounts for around 13 % of the IP sector decoupling and around 4 % of the total decoupling. The limitation of sulphur content in marine oil, implemented in 2007 and 2010, resulted in a decoupling of around 1 % by 2012 when compared to 2005. The third instrument, the sulphur tax, accounted for around 0.4 % of the total impact on decoupling by 2012.

The total decoupling by 2012 was mainly composed of the impacts from ET sector emission factor changes (approx. 30 ktonne). Over 29 ktonne could not be associated with any individual instrument. The impacts of the instruments on increased productivity, fuel use changes and IP sector emission factor changes were not quantified. It is stated, that further research and more statistical data would be needed to do so.

Want to know more?

- (Åström et al., 2017) The impact of Swedish SO2 policy instruments on SO2 emissions 1990–2012, Åström S., Yaramenka K., Mawdsleya I., Danielsson H., Grennfelt P., Gernerb A., Ekvalla T., Ahlgren E.O, Environmental Science & Policy Volume 77, November 2017, pp 32-39, https://doi.org/10.1016/j.envsci.2017.07.014

Relevant references provided in the study

- (Amann, M., et al., 2014) The Final Policy Scenarios of the EU Clean Air Policy Package, TSAP Report #11, http://ec.europa.eu/environment/air/pdf/TSAP.pdf

- (Gillberg, B., 2015) Åren med Miljöcentrum och Miljö och Framtid. Retrieved 2015-03-15, from www.miljocentrum.se
- (Fölster, J., et al., 2014) För var dag blir det bättre men bra lär det aldrig bli—Försurning i sjöar och vattendrag 2014, http://pub.epsilon.slu.se/11943/1/__storage-sk.slu.se_home%24_rabi0001_Desktop_folster_j_et_al_150225.pdf
- (Lindmark, M., Bergquist, A.K., 2008) Expansion for pollution reduction? Environmental adaptation of a Swedish and a Canadian metal smelter, 1960–2005. Bus. Hist. 50 (4), 530–546, http://dx.doi.org/10.1080/00076790802106877
- (Rättsnätet, 2015) Lagboken. Retrieved 2015-06-15, 2015, from www.notisum.se
- (Rafaj, P., et al., 2014) Changes in European greenhouse gas and air pollutant emissions 1960–2010: decomposition of determining factors. Clim. Change 124 (3), 477–504.
- (Regeringskansliet, 2015) Regeringskansliets rättdatabaser. Retrieved 2015-10-30, 2015, from http://rkrattsbaser.gov.se/sfsr
- (SPBI, 2015) Statistik. Retrieved 2015-04-01, from http://spbi.se/statistik/.
- (Statistics Sweden, 2015) Statistikdatabasen. Statistics Sweden, http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__EN__EN0202/?rxid=1576171b-7a58-4682-ac03-69303de12262
- (Svensson, H., 2003) Tillståndsprövningen i Sverige—Vem bör sköta den? Department of Law, Gothenburg University. Bachelor, Gothenburg, https://gupea.ub.gu.se/bitstream/2077/2218/1/200380.pdf
- (Swedish Environmental Protection Agency, 2014a) Informative Inventory Report 2014—Submitted Under the Convention on Long-Range Transboundary Air Pollution, http://www.naturvardsverket.se/upload/sa-mar-miljon/klimat-och-luft/luft/luftfororeningar/iir-sweden-2014.pdf
- (Swedish Environmental Protection Agency, 2014b) Informative Inventory Report Sweden 2014—Annexes. . http://www.naturvardsverket.se/upload/sa-mar-miljon/klimatoch-luft/luft/luftfororeningar/iir-sweden-2014-annex.pdf
- (Swedish Environmental Protection Agency, 2014c: National Inventory Report Sweden 2014—Greenhouse Gas Emission Inventories 1990–2012, Submitted Under the United Nations Convention on Climate Change and the Kyoto Protocol, www.naturvardsverket.se/upload/sa-mar-miljon/statistik-a-till-o/vaxthusgaser/2014/national-inventory-report-2014.pdf

Case #3: Applying ex-post index decomposition analysis to final energy consumption for evaluating European energy efficiency policies and targets (Reuter et al., 2019)

Decomposition analysis

Description of the policy measure

The decomposition analysis conducted in this study focusses on disentangling the extent to which different drivers of final energy consumption (FEC) contributed to FEC changes. The focus was on the period from 2000 to 2015 and the EU28 and its Member States. The corresponding European policies and measures that could influence these drivers were:

- The European Energy Efficiency Directive 2012/27/ EU (EED);
- The Energy Performance of Buildings Directive 2010/31/EU (EPBD);
- The Ecodesign Directive 2009/125/EC;
- The Energy Labelling Directive 2010/30/EU, which was replaced recently with Regulation (EU) 2017/1369;
- The European Emissions Trading Scheme (EU-ETS);
- The Effort Sharing Decision 406/2009/EC;
- The Industrial Emissions Directive 2010/75/EU;
- The Regulation (EC) No 443/2009 regarding CO2 emissions of new passenger cars and a similar regulation for LDVs;
- The Car Labelling Directive 1999/94/EC.

Furthermore, the study states that many national policies played a role for improving energy efficiency as well. The index decomposition analysis (IDA) conducted in this study aims to show the effects of energy efficiency policies in the European and national context.

Scope of the evaluation

At an aggregated level, the five sectors (industry, households, transport, services and agriculture) were examined with respect to changes in FEC. At a more detailed level, the influence of residential end-uses, transport modes and industrial subsectors were investigated. An index decomposition analysis (IDA) based on the Logarithmic Mean Divisia Index method (LMDI) was applied as the method for this study. Most of the data used for this analysis is publicly accessible.

Germany and Poland were investigated also separately in a detailed analysis. Germany was chosen as representative for Member States which have reached a saturation level concerning energy consumption. Poland was chosen to represent Member States with expected increase in energy consumption in the future based on significant economic growth.

The time period 2000-2015 was selected due to good statistical coverage for all Member States and relevance for efficiency related policies. Additionally, the time period from 2007 to 2015 was considered separately to analyze the development that directly followed an economic crisis. During these time periods, the above mentioned nine policies were adopted by the EU, alongside numerous national energy efficiency policies. Some policies directly affect the final energy consumption, while others only indirectly affect the consumption (e.g. EU-ETS).

Methodology

The widely adopted Logarithmic mean Divisia Index Method (LMDI) was used as the basis for the index decomposition analysis (IDA). LMDI is recommended by its ease of use due to the theoretical foundation, adaptability and result interpretation (see also Paragraph 4.5 Decomposition analysis). The approach follows this general equation:

$$\Delta V_{\text{tot}} = V^T - V^0 = \Delta V_{x_1} + \Delta V_{x_2} + \cdots + \Delta V_{x_n} \qquad (1)$$

With:
V = aggregate (e.g. final energy consumption (FEC)) or sub-aggregate (e.g. industry sector);
x = factors;
0 = start of time period;
T = end of time period;
tot = total.

Final energy consumption (FEC) is defined as the energy supplied to the final consumer and is not including non-energy uses (IEA, 2004). According to equation (1), the decomposition of changes in FEC can be formulated as shown in equation (2):

$$\begin{aligned} \text{FEC}^{\text{T}} - \text{FEC}^{0} = \Delta\text{FEC}_{\text{tot}} = \Delta\text{FEC}_{\text{ACT}} &+ \Delta\text{FEC}_{\text{STR}} \\ &+ \Delta\text{FEC}_{\text{MOD}} + \Delta F\text{EC}_{\text{COM}} \\ &+ \Delta\text{FEC}_{\text{HDD}} + \Delta\text{FEC}_{\text{EFF}} \end{aligned} \qquad (2)$$

With:
ACT = activity-based changes;
STR = structural changes;
MOD =shifting transportation modes;
COM = variation in comfort and social factors (e.g. living space, inhabitant per dwelling);
HDD = heating degree day (i.e. changes in average temperature);
EFF = improvements of energy efficiency.

These decomposing drivers from equation (2) were aggregated from different factors of each sector. The aggregation process is schematically shown in Figure 21.

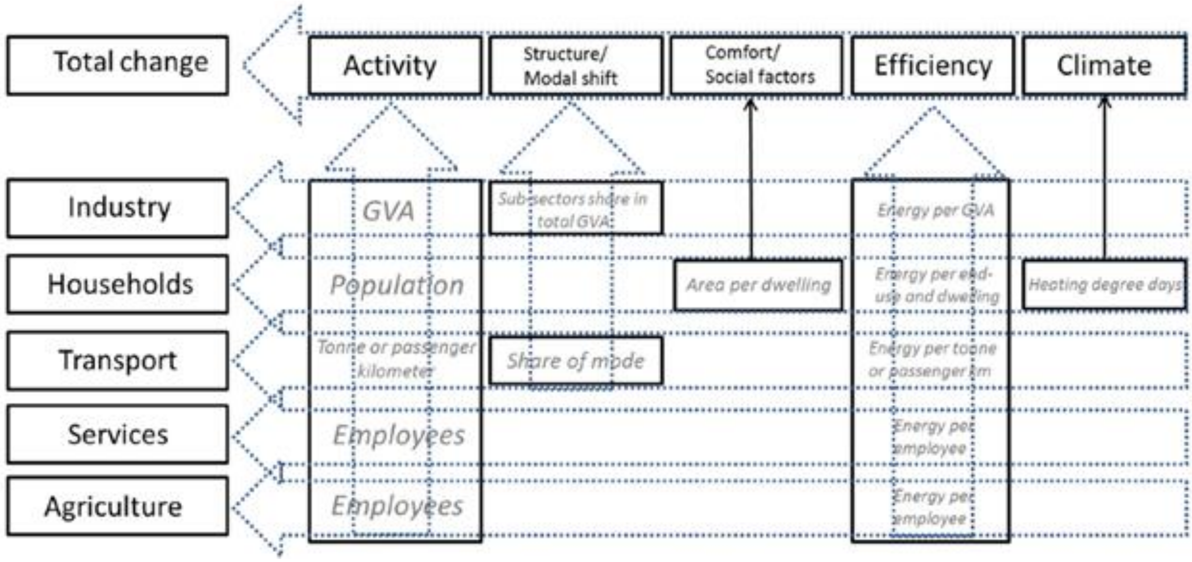


Figure 21. Aggregation of factors for the five sectors (Reuter et al., 2019).

The industry sector FEC was decomposed into three factors: ACT, STR and EFF. The results were visualised as shown below, another way in which decomposition results can be visualised (i.e. different to the way shown in Paragraph 4.5 of the section on Decomposition analysis).

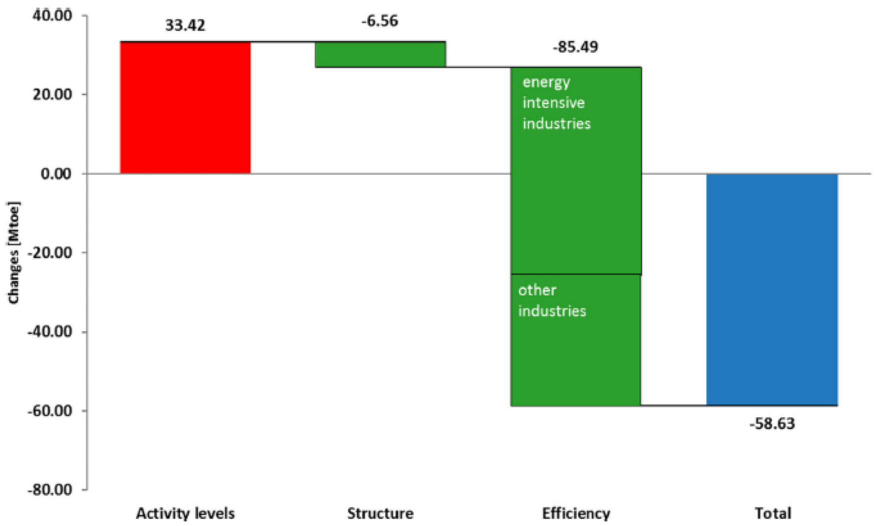


Figure 22. Changes in final energy consumption of industry (EU28, 2000 to 2015) (Energy intensive industries include subsectors of primary metals, paper and pulp, chemical and petrochemical and textile and leather) (Reuter et al., 2019).

Depending on the end-use, all of the five factors were considered for the household sector. Therefore, a separate equation was used for each end-use (space-heating, hot water, cooking, electric appliances and lighting). The FEC of the transport sector was decomposed into the factors ACT, STR/MOD and EFF and calculated separately for passenger and freight transport. The decomposition of FEC into ACT and EFF was applied for the services sector as well as the agriculture sector.

A combination of three data sources was used for the analysis. In Table 24 further details about the data sources are depicted.

Table 24. Information about the data used in this study (Reuter et al., 2019).

Category	Details	Source
Energy demand	• Simplified energy balances • Supply, transformation & consumption of electricity • Supply, transformation and consumption of heat	Eurostat [1]
Detailed sector data	• Gross value added (IND, AGR) • Number of employees (Services) • Passenger-km per mode for freight traffic, energy consumption per mode (Transport) • Number of households, average population, number of dwellings, share of end-uses, heating degree days (HH)	ODYSSEE-MURE [2]
Transport data	• Data on transport	Statistical pocketbook of DG Tren [3]

[1] http://ec.europa.eu/eurostat/data/database
[2] http://www.odyssee-mure.eu
[3] http://ec.europa.eu/transport/facts-fundings/statistics/

Outcome

Results for EU-28

In the time period from 2000 to 2015, the total FEC decreased by 48.4 megatons of oil equivalent (Mtoe) in the EU-28. The main reduction was realized in the industry sector followed by the household sector. The developments in the transport and services sector counteracted the reduction with an increase of 14 Mtoe and 17 Mtoe, respectively. Activity-based changes accounted for the largest increase for most of the sectors, whereas the main decreasing effect was contributed to an improvement of energy efficiency. For the industry sector the change in energy efficiency dominated all other impacts, especially for energy intensive industries (subsectors “primary metals” and “chemical and petrochemical”). This is assumed to be the result of the Emission Trading Scheme, the Ecodesign Directive and the Labelling Directive.

Due to the economic crisis, the economic activity in the industry sector dropped drastically resulting in an activity-based decrease of 11.6 Mtoe in the period of 2007 to 2015. The efficiency gains in the household sector resulted from more efficient space heating technologies and retrofitting of building envelopes, which could be assigned to the Energy Performance of Buildings Directive from (EPBD) and its implementation. During the period of 2007 to 2015, the economic activity had a smaller impact on FEC in this sector compared to the period 2000-2015. The total FEC decrease between 2007 and 2015 in the transport sector was mainly composed of energy efficiency as well as a rising activity level in passenger transport and freight transport. The efficiency improvements indicated that European transport policies focusing on technological progress had an impact in the investigated time period. Considering the period of 2007 to 2015, the factor EFF surpassed all other factors.

The rising number of employees in the service sector resulted in ACT being the main driver of the changes in FEC with 24.5 Mtoe. The EFF factor played only a minor role. However, from 2007 to 2015 the energy efficiency improved in this sector, it is assumed due to the introduction of the Ecodesign Directive and policies regarding energy efficiency in buildings.

In the agricultural sector, the FEC decreased by 4.6 Mtoe from 2000 to 2015 based on energy efficiency gains and almost constant values for the factor ACT.

Table 25 gives an overview about the changes in FEC at the European level as well as the two national levels by each of the five decomposing factors from 2000 to 2015.

Table 25. Changes in total final energy consumption for EU28, Germany and Poland for the time period 2000 to 2015 (+ indicates increase, - indicates decrease).

Changes per Region	EU28 (in Mtoe)	Germany (in Mtoe)	Poland (in Mtoe)
Activity levels (ACT)	+125.1	+25.9	+19.4
Structure (STR)	-6.6	-4.6	-2.9
Modal shift (MOD)	+2.6	-0.2	+2.2
Comfort/ behavioural/ social factors (COM)	+35.3	+7.3	+3.4
Annual climate variation (HDD)	+5.7	+1.8	+0.8
Efficiency (EFF)	-210.5	-38.1	-15.9
Total	-48.4	-7.9	+7

Source: (Reuter et al., 2019).

Results for Germany and Poland

In Germany, a total decrease of about 8 Mtoe in FEC was composed through the following parts: (1) the main driver of FEC changes was energy efficiency gains for the period of 2000 to 2015. These energy efficiency gains combined with structural changes and modal shift resulted in a decrease of 42.9 Mtoe; (2) Rising activity levels, comfort factors and annual climate variation caused an increase of 35 Mtoe.

A similar result is shown for the period of 2007 to 2015. Efficiency gains were the strongest for the household sector followed by transport, whereas the industry sector showed relatively little efficiency improvements.

In Poland, FEC increased by 7 Mtoe. from 2000 to 2015, which are explained by following drivers: (1) energy efficiency gains and structural changes accounting for FEC decreases of 18.8 Mtoe; (2) The other factors contributed to an increase of 25.8 Mtoe. The decreases and increases combined thus resulted in an overall FEC increase of 7 Mtoe.

The results shown in this study could imply possible policy recommendations that would positively impact the final energy demand reductions in the EU in the future. Particularly, the transport sector is emphasized as a target for future energy efficiency policies in order to counteract rising economic activity levels in newer Member States.

Want to know more?

- (Reuter et al., 2019) Applying ex post index decomposition analysis to final energy consumption for evaluating European energy efficiency policies and targets, Reuter M., Patel M.K., Eichhammer W., Energy efficiency 12 (2019), No.5, pp.1329-1357, https://link.springer.com/content/pdf/10.1007%2Fs12053-018-09772-w.pdf

Relevant reference provided in the study

- (IEA, 2004) Energy Statistics Manual. Paris: OECD Publishing, IEA, 2004, https://doi.org/10.1787/9789264033986-en

Case #4: Evaluation of Dutch long-term agreement on energy efficiency (Meerjarenafspraak – MJA3) 2008 – 2020 (Ecorys, 2013)

Cost effectiveness analysis	**Surveys**	**(Systematic) literature review**
Interviews	**Monitoring performance data**	

Description of policy measure

The Dutch long-term covenant on energy efficiency (known as MJA3) concerns agreements between government and companies about more effective and efficient use of energy. Reducing CO_2 emissions is not considered as the main goal. The objective of MJA3 is that the affiliated companies - by the end of 2012 there were around 1,160 companies – achieve 30 % energy efficiency improvement in the period 2005-2020. The companies are part of 33 different participating sectors, which can be subdivided into 4 clusters:

- Industry (19 sectors);
- Food industry (9 sectors);
- Services (4 sectors);
- Transport (only 1 sector, namely rail sector).

The companies participating in the MJA3 typically represent around 70 - 80 % of energy use of their sector, so the majority of energy consumption in the relevant sectors is therefore covered by the covenant. The Dutch Agency is responsible for implementation of the covenant, including facilitation of the involved parties. The agency offers expertise and support at all stages of the covenant process: preparation, analysis, planning and implementation.

In addition to MJA3 and its precursors, there are also other policies that more or less directly intervene in the improvement of energy efficiency. The most important other Dutch policy instruments in this area are:

(1) Energy Investment allowance (EIA): the EIA offers companies a deduction from fiscal profit or income in order to reduce the costs of energy efficiency assets.

(2) Environmental Management Act (Wm) which obligates companies to implement profitable investments for energy savings (i.e. having a payback period of about 5 years or less).

Scope of the evaluation

The purpose of the evaluation was to answer the following three main questions:

(1) To what extent has the MJA3 contributed to additional energy savings, on top of autonomous energy savings (ex-post/effectiveness)?

(2) What are the expected energy efficiency improvements in the next four years, as a result of the MJA3 agreements (ex-ante effectiveness)?

(3) What are the compliance costs for government and business in comparison to the benefits or effects (ex-post/ex-ante cost effectiveness)?

Methodology

During the evaluation it appeared that relatively **little data of good quality** were available, so to answer the evaluation questions various data collection methods had to be used, as summarized in the table below.

Table 26. Data collection methodologies applied to assess the (cost) effectiveness of MJA3.

Type of data collection method	Description of the collection data
Bottom-up monitoring & verification reports submitted to the Dutch Agency	MJA3 participants are required to submit monitoring data annually by 1st April at the latest to the Dutch Agency. These company reports provide insights into the implementation progress of the energy efficiency plans (EEP). The monitoring data show whether the companies will achieve their energy objectives as set in their EEP. The achieved saving results are aggregated per sector and reported annually by the Dutch Agency on the basis of the monitoring reports from the participating companies.
Literature review	An extensive literature study of documented research into the effectiveness of covenants as a policy instrument for energy efficiency and into the autonomous trends in energy efficiency (i.e. setting the benchmark for comparison of the MJA3 results).
Interviews	Field research through a large number of interviews (around 20) with participating companies, industry associations, policy makers and other stakeholders, giving more insights on the quality and effectiveness of the MJA3 implementation process.
Online survey	Field research through an extensive, online survey among MJA3 participating companies (response rate of about 30% was reached). This way, a better understanding could be gained into the energy efficiency improvement achieved and underlying efficiency measures, the autonomous trends as well as the burden or costs for companies to comply with the covenant.

The study sets a benchmark for comparison of the MJA3 results so the **additionality of the covenant** can be estimated on top of autonomous savings (i.e. savings that would be realized without MJA3 due to financial and sustainability objectives of the companies and/or other policies and measures). Hereto, the covenant savings are compared to European averages from similar industrial sectors. This analysis is based on a literature study as well as a comparison with EU databases (e.g. energy intensity data from PRIMES model, energy savings dataset Odyssee-MURE), but only resulted in a very rough indication of additionality. Therefore, the assessment is complemented with results from the survey, to get a better - although still uncertain - understanding of implemented savings in absence of MJA3.

In the **cost effectiveness assessment**, the following types of compliance costs are considered:

(1) Implementation costs (or administrative costs) are the costs that the government must incur to ensure compliance and/or implement the MJA3 (e.g. support and facilitation by the Dutch Agency). These types of costs were monitored by the Dutch Agency;

(2) Administrative burdens are the costs for the companies or business to meet information obligations arising from the MJA3. It's about collecting, editing, registering, saving and making available the necessary information (e.g. annual monitoring report from the companies, energy efficiency plans EEP submitted by the participants). The online survey allowed a rough estimate to be made of these types of costs by collecting the number of required hours of 3 employee categories (secretary, staff and management) to comply with the obligations.

(3) Substantive compliance costs are the additional costs of companies to comply with the MJA3 with regard to the behaviour of persons and conditions (buildings, production processes or products/services) in companies with a view to safeguard public goals (e.g. investments or implementation of cost-effective measures to improve energy efficiency within the companies). Here as well, the online survey allowed to make a rough estimate of these type of costs by collecting the number of required hours of 3 employee categories (secretary, staff and management) to comply with the obligations.

The principle of additionality counts for the compliance costs, and - again - mainly relies on the survey results (i.e. 50% of actions through MJA3 would have been realized in absence of the covenant), so half of the total, substantive compliance costs can be linked to MJA3. Concerning administrative costs, almost all surveyed costs relate to the MJA3 given its unique reporting obligations.

The additional compliance costs per year are compared to the yearly average, additional energy savings, resulting in an average cost-effectiveness. To interpret the results, the cost effectiveness is compared to the effectiveness of other policy instruments (literature study) such as the previous MJA2, the EU ETS and the Energy Investment allowance (EIA).

Outcome

The intended improvement in energy efficiency has been achieved, i.e. the *target of around 2 % efficiency improvement per year is on average achieved*. Over the period 2005-2011 (up to 2008 MJA2, after 2008 MJA3) a total energy saving of about 60 PJ was realized, if use of renewable energy is included. Without renewable energy, the savings amounted about 29 PJ in the same period corresponding to savings of 13 % over the entire 2005-2011 period, or 2.1 % per year. The latter improvements were mainly related to improved, industrial process efficiencies. The governmental support received by the industrial participants are perceived as of good quality and effective.

But, despite the uncertainties related to the various data sources, the Dutch industry does *not seem to score better than the European average*. In comparison to the Dutch industry as a whole, the MJA3 sectors seem to have an improved energy efficiency than the non-MJA sectors. Comparisons with European databases suggest that the MJA3 objective (2 % energy saving per year) is only of limited ambition, and therefore, the Dutch industry is certainly not, or not anymore, among European leaders on energy efficiency. On the other hand, some MJA3 sectors do outperform the European average. This outcome is confirmed by the participants from the surveys and interviews, as they indicate that around 50 % of the achieved energy savings through *MJA3 would have been realized in absence of the covenant* as well.

From the MJA3, companies have to deal with various obligations, which certainly results in a regulatory burden (i.e. administration costs and substantive compliance costs). *Compared to the reference alternative or benchmark*, namely the situation without MJA3 covenant but with the Environmental Management Act (Wm), the *regulatory burden varied for the MJA3 for all companies* together between € 6 and 11 million per year over the period 2008-2012. For the participating sector organizations the total regulatory burden is roughly € 1.0 million per year.

The compliances costs can be compared with the realized energy savings from different policy instruments (cost-effectiveness analysis). Comparing the additional compliance costs of MJA3 (on average € 23.8 million per year, including total regulatory burden) with the additional yearly savings (30.6 PJ over a 15-year period) leads to an *average cost-effectiveness of approximately € 0.78 per saved GJ over the period 2008−2012*. When we compare this with the cost effectiveness of other instruments, the effectiveness of MJA3 appears to be more favorable compared to the previous MJA2

(€ 1.10 - € 1.70 per GJ), is approximately the same as the ETS (€ 0.70 - € 0.90 per GJ in 2008) and less favorable than EIA (€ 0.40 per GJ).

Want to know more?

- (Ecorys, 2013) Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3) - Eindrapport van de door Ecorys uitgevoerde Evaluatie Meerjarenafspraak Energie Efficiëntie 2008-2020 (MJA3), Ecorys, 2013, https://www.rijksoverheid.nl/documenten/rapporten/2013/04/10/evaluatie-meerjarenafspraak-energie-efficientie-2008-2020-mja3

Case #5: Dutch Energy Investment allowance (EIA) 2012-2017 (CE Delft, 2018)

Cost effectiveness analysis **Counterfactual analysis** **Free riders**

Surveys **(Systematic) literature review** **Interviews**

Monitoring performance data **Indicator analysis**

Description of policy measure

The Energy Investment Allowance (hereinafter: EIA) is a fiscal measure that aims to achieve energy savings for companies. The savings (joules per year) are made by market acceleration through investments in the introduction of innovative assets which are more energy efficient than conventional assets. The operating assets that are eligible for this measure are included on the "Energy List"[19]. The scheme contributes to the energy saving objectives for 2020. With the EIA, the central government aims to promote energy savings and reduce CO_2 emissions.

A total of 78 sectors are involved. Most applications have been made in the retail sector; breeding and keeping animals (heating systems for stables); food preparation; other business services and agriculture; and, hunting and services for agriculture and hunting.

The Ministries of Economic Affairs and Climate (Economische Zaken en Klimaat - EZK) and Finance are responsible for the EIA scheme. The Dutch Enterprise Agency (Rijksdienst voor Ondernemend Nederland -RVO) and the Tax Authorities are charged with the implementation of the EIA.

Scope of the evaluation

In general, Dutch economic incentives with a legal basis must be evaluated every five years. This evaluation covers the period 2012-2017. The EIA evaluation, carried out in 2018, focuses on whether the EIA has achieved the set objectives in an effective and efficient manner. Therefore, the evaluation questions can be divided into two categories:

(1) Effectiveness: To what extent the EIA motivates energy-efficient investments and ultimately leads to extra energy savings?

(2) Efficiency/cost effectiveness: This evaluation compares the costs for government and business with (net) investments and energy savings.

Methodology

For the evaluation, multiple data sources were used. The following table summarizes the data sources used to assess the (cost) effectiveness.

[19] The Energy List is updated annually. The market acceleration of innovative company's assets is encouraged by the annual update of the Energy List. The yearly adjustment is possible in three ways: (i) a new, innovative business asset can be added; (ii) the requirements for a business asset can be adjusted, which leads to a substantive change of the text in the Energy List; (iii) if a business asset has become common, it can be removed from the Energy List.

Table 27. Data collection methodologies applied to assess the (cost) effectiveness of EIA.

Type of data collection method	Description
Literature review	The literature review was used to answer the questions that relate to the qualitative and quantitative part of the analysis. This concerns literature on enablers and barriers among investors and providers of energy-efficient technologies.
Interviews	Interviews of a small group of installers and suppliers of the innovative assets to understand the case studies or selected innovative techniques more in depth than from EIA applications (control group, benchmark).
Monitoring based on portfolio analysis	Bottom-up monitoring based on portfolio analysis of EIA applications (2012-2017 period): this data has been made available by RVO (e.g. annual energy savings, financial data of investments). In this period, a total of 87,245 applications were made under the EIA.
Surveys (e-mail)	The purpose of the surveys among a large group of investors and suppliers was to obtain information about the effects of the scheme on investments in energy-efficient technologies, both at the demand and the supply side (incl. free rider effect). In addition, questions were asked about the extent of the administrative burden, coherence with other instruments and the role of intermediaries. Two surveys were sent out to the target group of the scheme: (1) survey aimed at investors (11 % response rate); (2) survey aimed at suppliers of company resources (50 % response rate).

(1) Assessment of effectiveness

To assess effectiveness, it is important to determine the savings that will be realized with the EIA support and that would not have been achieved without the scheme. The additional energy savings achieved by the EIA indicate the extent to which the scheme is effective. In the assessment of effectiveness, a distinction is made between (1) first-order effects (additional energy savings due to investments in operating assets on the Energy List), and (2) second-order effects (the development and innovation of operating assets):

(1) For the first-order effects, the evaluators calculated the gross energy savings achieved. This **indicator analysis** concerned the energy savings that occur by commissioning the energy-efficient technology that is supported with the EIA. Hereto, the **portfolio analysis** is applied to estimate the yearly, gross energy saving effect: For the 20 most requested technologies, RVO annually calculates a savings figure of primary energy per euro invested per year (Nm3/euro per year). These techniques represent more than 60 % of all applications. The expected annual gross energy savings of the 20 most requested technologies (plus generic technologies) is calculated by multiplying the savings figure with the awarded subsidy. Afterwards, the part of the energy savings achieved that is attributable to **free-riders** is deducted from the gross savings. These are the **net energy savings**.

(2) In the case of the second order effects, market acceleration of new technologies is accounted for (i.e. introducing market potential for these technologies). This is the ‘innovation effect’ of the scheme, and is defined as the fiscal benefit needed (per kWh energy savings) in order to make the technology profitable from the perspective of the market (investor’s perspective). The evaluators

referred to this as the 'minimum required tax benefit'. The minimum required tax benefit is calculated on the basis of payback time analysis.

Since the situation without an EIA is unknown, the evaluators established a control group to properly estimate the additional effect on energy savings (**defining the baseline**). The basic or counterfactual scenario was mapped based on:

- **Free-riders:** This covers the scenario in which some companies that participated in the scheme would have made the same decision at the same time without the use of the instrument. With the help of a survey, an attempt was made to gain insights into the behaviour of the participants of the scheme, and what they would have done if the EIA scheme had not been implemented, by asking directly what the participants would have done without the scheme. Unfortunately, the survey did not lead to representative results considering the size of the sample. A second method uses financial data to gain more insights into the critical payback time with and without regulation. The difference between these payback times provides insight into possible free-rider behaviour. Hence, an investor who is already implementing a profitable measure even without the EIA can be identified as a free-rider. Since the number of observations for this method is limited (41 usable data points), this payback time method is primarily intended as an illustration.
- **Case studies based on interviews:** The control group was also mapped through five case studies of selected innovative techniques. For each case study, two to three in-depth interviews were conducted with Dutch installers and suppliers. In particular, the installers have provided an independent estimate of the proportion of customers and investments investing in the selected techniques without using the EIA.

(2) Assessment of cost effectiveness

Similar to the Dutch case study #4, the evaluation made a distinction between implementation costs and administrative burdens. Implementation costs are the costs the government must incur to realize compliance and/or implementation of legislation and regulations. Administrative burdens are the costs for applicants to comply with the information obligations arising from laws and regulations from the government (i.e. collection, processing, registration, storage, provision of information and application costs):

- Implementation costs for government: In the 2012-2016 period, RVO estimated the annual implementation costs to be between 4.6 million euros and 5.3 million euros. The Tax Office estimated their implementation costs at 0.4 million euros per year to support the scheme administration;
- Administrative and substantive compliance costs for applicants: this includes, among other things, the time investment for the application and the costs of hiring intermediaries. Based on a few survey questions, an estimate was made of the level of the administrative burden. The respondents were asked what their average time commitment and their average costs were for the application (a distinction was made between management and administrative staff).

The costs for the government include the budget loss (budget foreseen for the EIA) and the implementation costs of RVO and the Tax Authorities. The evaluation compared these costs with the CO_2 reduction achieved through the EIA. The gross efficiency is not adjusted for free riders.

As for the total national costs, these are equal to the direct costs of the government, and the costs of business. Costs for business are equal to the benefits for free riders, the substantive compliance costs of groups 1 and 3, and the administrative burden. The evaluation did not consider how large the costs were for groups 1 and 3, but estimated them to be low.

Outcome

(1) Assessment of effectiveness

The effectiveness of the EIA was limited in practice by the action of free-riders. *Free-ride behaviour* in this case specifically plays a role for energy saving technologies that already have a payback period of their own. The final estimate of suspected free riders is slightly below 50 % with a bandwidth between 30 % and 69 %.

In terms of the *portfolio analysis* (**gross** savings per year), the evaluation calculated a total of more than 38 PJ energy saved annually with assets supported by the EIA. This leads to an avoided annual CO_2 emission of more than 2 million tons.

For the **net** energy savings per year, the report mentions that the net energy savings is equal to the gross energy savings achieved, minus the part of the energy savings realised that is attributable to free riders. In the calculation of the net effects, a bandwidth of 30 to 69 % free riders is assume, based on the number of applications. As a result, the net savings amount 12 to 27 PJ.

The EIA also contributes to the reduction of greenhouse gases. Over the entire evaluation period 2012-2017, the EIA contributed to an annual CO_2 reduction of 0.7-1.5 Mtonne in the Netherlands. In the Dutch National Energy Outlook, the adopted and planned policies are intended to reduce emissions by 23% by 2020 compared to 1990 (i.e. 50 Mtonne of CO2 reduction by 2020).

(2) Assessment of cost effectiveness

From a *governmental perspective*, implementation costs amount to more than 5 million euros annually, or more than 300 euros per notification. The budget loss is around € 100 million per year. The annual energy savings are between 5,000 and 11,000 TJ (5-11 PJ), and between 0.3 and 0.6 Mtonne CO_2 is saved. This concerns the gross savings, without correction for free riders. The evaluation assumed a bandwidth in the freeride percentage of 30-69 %. As a result, every tonne of CO_2 reduction costs the government an average of 21-46 euros on average.

From a *national or social perspective*, the evaluation estimated the national cost effectiveness at 15-17 euros per tonne CO_2 reduction for the 2012-2016 period.

In conclusion, the cost-effectiveness of the EIA is relatively high, from both perspectives. This also applies when taking into account the "relatively high" proportion of free riders.

Want to know more?

- (CE Delft, 2018) Beleidsevaluatie Energie Investeringsaftrek (EIA) 2012-2017, CE Delft, 2018, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018Z11354&did=2018D33904

Case #6: Economic evaluation of French écopastille (eco-tax bonus-malus & super-bonus new vehicles) 2008-2012 (CGDD, 2013)

Cost benefit analysis	**Counterfactual analysis**	**Regression analysis**
Rebound effects	**Indicator analysis**	**Monitoring performance data**

Description of policy measure

France set up in 2008 the bonus-malus and super bonus. The combination of these two schemes is an economic incentive for the acquisition and production of lower-emission vehicles. This has significantly changed the structure of passenger car sales in France, compared to what had been observed previously.

The **bonus-malus scheme** (BM) aims to stimulate the use of more fuel-efficient techniques in the automotive field by delivering a price signal that acts both on supply and on demand. It steers consumer choice towards more fuel-efficient vehicles, on the one hand; and it encourages manufacturers to make vehicles that meet this demand and innovate in this way, on the other hand. More specifically, the scheme must be able to accelerate the reduction of CO_2 emissions of new passenger cars by applying a tariff scale depending on CO_2 emissions. In 2008, a subsidy was granted to new cars emitting less than 120 gCO_2/km, while new vehicles emitting more than 160 gCO_2/km were taxed on the occasion of their first registration. The scale has gradually hardened over the years, for instance, in 2013 the subsidy is awarded below 105 gCO2/km, whereas the penalty applies to vehicles emitting more than 135 gCO2/km.

The **super bonus** aims to accelerate the rate of renewal of the vehicle fleet and thereby reduce its average emissions. In 2008, a premium of EUR 300 is granted, subject to the acquisition of a new vehicle emitting less than130 gCO2/km, for the disposal of a vehicle older than 15 years. As part of the recovery plan, the super bonus was replaced in 2009 and 2010 by a scrappage premium. Since 2011, the super bonus system is in place again. However, it was modified in 2012, where an amount of EUR 200 is granted, in addition to the ecological bonus when a vehicle older than 15 years is disposed.

Scope of the evaluation

The evaluation presents the main facts observed in France concerning the evolution of private new vehicle registrations between 2008 and 2012, some comparisons with previous years and with what was observed in other European countries. In the **cost benefit analysis** of the *écopastille* from 2008 to 2012, different pillars to examine its effects were touched upon: economic, environmental, and a socio-economic assessment.

Methodology

The evaluation performed a socio-economic cost benefit assessment, including the following **types of costs:**

(1) Loss of consumer surplus: this is linked to a restriction of consumer's choice in comparison to their previous buying habits. In general, consumers do not value their costs over the total period of owning their vehicle, but only value these costs during the first years, namely from 3 to 5 years (in this French evaluation study, a period of 4 years is assumed).

(2) The opportunity cost of public funds (coût d'opportunité des fonds publics, COFP): the deficit of the economic incentives and the shortfall of revenues for the government linked to the fall in the internal

consumption tax of energy products (TICPE – Taxe Intérieure de Consommation sur les Produits Energétique, also formerly known as TIPP) associated with lower fuel consumption.

(3) The effects in terms of local pollution: the écopastille scheme could favour diesel vehicles, which are more CO_2 efficient. However, local pollution emitted by a diesel vehicle is more harmful compared to a gasoline vehicle.

As for **the types of benefits**, the evaluation included:

(4) Fuel savings: the report valued the reduction in fuel consumption over the lifetime of new passenger vehicles, removing the part of these savings (over 4 years) already “internalized” by consumers and therefore integrated into the ‘loss of consumer surplus’.

(5) The reduction of CO_2 emissions: a carbon value equal to 32 EUR/tCO_2 until 2010 was proposed, growing at a rate of 5.8% per year after that until 2030. This value makes it possible to monetise the gains of the bonus-malus scheme in terms of CO_2 emissions over the lifetime of the new vehicles.

So, in total 3 pillars are monetized: (i) economic. (ii) environmental, and (iii) socio economic. **Rebound effects** are also taken into account as a surplus of traffic is expected inducing social and environmental costs, which are not completely covered by the levies on road traffic (namely, tolls and TICPE).

In order to assess the impact of the eco-label scheme, a **reference scenario** was defined for the years 2008-2012 to isolate the effect of the scheme from the cyclical effects (counterfactual scenario). The assumptions to construct this reference situation are based on a comparison of the French situation with the EU on the one hand, and econometric studies on the other hand, for instance:

- Total CO_2 emissions are based on an econometric study (least squares regression) assessing the evolution of emissions as a function of the fuel price (price elasticity -0.35);
- Car registrations and the fleet are also estimated by econometric studies (least squares regression) based on the average price of private vehicles (price elasticities are -1.7 and -0.1513 respectively).

The table below presents the data sources and type of indicators applied for the cost benefits assessment.

Table 28. Data sources applied to assess cost benefits analysis of the French écopastille.

Indicator	Details	Source
Total registrations of private vehicles (per type)	Region : France and EU15 Period : 2000 – 2012	CCFA - Comité des constructeurs français d'automobiles, calculs CGDD - Commissariat général au développement durable
Emissions of new, private vehicles (gCO_2/km)	Region : France and EU15 Period: 2000-2012	ADEME - Agence de l'Environnement et de la Maîtrise de l'Énergie, CGDD
Consumption of new, private vehicles (L/100km)	Region: France Period: 1976 -2012	CGDD
Average fuel prices	Region : France Period: 1976 -2012	SOeS - Service de l'observation et des statistiques
Total size of the private vehicle fleet	Region : France Period: 1984 – 2011	CGDD
Price of new vehicles	Region: France Period: 1984 – 2011	CGDD
Rate of dieselization of new private vehicles	Region: France and EU17 Period: 2000-2012	CCFA
Registration of second hand, private vehicles	Region: France Period: 2000-2012	SOeS

Outcome

According to the report the evaluation is largely positive over the five years 2008-2012, if the rebound effect is not taken into account. However, when considering the rebound effect, the cost benefit assessment becomes negative in 2008 and 2009. Budgetary imbalance is also induced by losses in terms of TICPE which weigh negatively on the balance sheet in 2011 and 2012. Most of the benefits from the scheme consisted of gains associated with lower fuel consumption, however, these are not taken into account by the owners of vehicles (for vehicles older than 4 years). The main results from the socio-economic analysis can be seen in Table 29.

Table 29. Socio-economic balance of French écopastille (Source CGDD, 2013).

En millions d'euros			2008	2009	2010	2011	2012	Total
Coûts	Perte d'utilité des automobilistes		-28	-73	-30	-13	-7	-152
	COFP	Budget BM	-88	-188	-84	-35	0	-396
		TICPE	-82	-197	-134	-96	-120	-629
	Pollution locale		-61	0	0	0	0	-61
	Sous total		*-260*	*-458*	*-248*	*-144*	*-127*	*-1237*
Avantages	Consommation de carburant		276	461	407	372	480	1998
	Emissions de CO2		68	164	112	81	102	527
	COFP	Budget BM	0	0	0	0	14	14
	Sous total		*344*	*626*	*519*	*454*	*596*	*2539*
Bilan hors effet rebond			**85**	**168**	**271**	**310**	**469**	**1302**
Effet rebond	Circulations supplémentaires		-144	-349	-246	-186	-242	-1168
	Ventes supplémentaires		-0,5	-2,4	-0,7	-0,2	0,1	-3,7
Bilan avec effet rebond			**-60**	**-184**	**24**	**123**	**227**	**130**
Report modal			131	316	223	168	217	1055
Bilan avec effet rebond et report modal			**70**	**132**	**247**	**291**	**444**	**1185**

Abbreviations :
COFP = coût d'opportunité des fonds publics ; BM = bonus-malus ; TICPE = taxe intérieure de consommation sur les produits énergétiques

Moreover, the report mentions that the rebound effect weighs very heavily on the overall balance sheet because it leads to marginal costs which are not properly internalized by the existing tax system. The evaluation remarks that for the scheme to be fully effective, taxation on road traffic should be increased (i.e. local congestion charges) to reduce the marginal cost to society. The positive effect of the modal shift almost offsets the rebound effect, and when included in the socio-economic calculation, the evaluation finds a positive balance every year.

Want to know more?

- (CGDD, 2013) Évaluation économique du dispositif d'écopastille sur la période 2008-2012, Commissariat Général au Développement Durable (CGDD), 2013, http://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0078/Temis-0078465/20744.pdf

Case #7: Free riding on tax credits CIDD for home insulation in France (Nauleau, 2014)

Free rider effect	Surveys

Description of policy measure

The buildings sector is considered to have a high energy savings potential and the promotion of energy efficiency investments in this area is hence a key component of climate policy. The French income tax credit, the Sustainable Development Tax Credit (CIDD), is an instrument introduced to encourage households to invest in energy efficient renovations. The CIDD was launched in 2005 and ran until 2014 when it was replaced by the Energy Transition Tax Credit (CITE) (Ecoyfs & Adelphi, 2018).

All households, regardless of level of income, can apply for the CIDD, which offers income tax credits for the purchase of energy efficient equipment and materials. The tax credit subsidy is set at a maximum of € 8,000 for a one-person dwelling and double for a two-person dwelling, with additional allowances per child. The CIDD comes in the form of an income tax reduction or a direct payment for those households that do not pay income tax. Tax credits rates range from 15-50 % of investment cost depending on pre-determined categories and energy performance criteria.

More than half of households undertaking retrofitting investments have used the CIDD since 2005. Surveys also suggest that the CIDD is widely known and considered by households to be the most decisive investment since 2006. Over 1 million households have benefited from the CIDD every year since 2006. Due to its popularity the CIDD has led to large annual public expenditures ranging between € 958 million (in 2005) and € 2.8 billion (in 2008).

Scope of the evaluation

An econometric assessment was carried out to determine the efficiency of the CIDD on investment decisions for household retrofits. This aimed to determine the overall effect of the CIDD and considered free rider effects as one aspect of this evaluation. The evaluation deals with three interrelated aspects of the CIDD:

- The effect of the CIDD on the probability of retrofitting;
- How the effect of the CIDD evolved over the period 2005-2011;
- Assessing the proportion of free-riders and what individual and housing characteristics influence free riding.

This case study will not cover all data, methods, and outcomes of this evaluation study, but focuses specifically on the free rider effects.

The analysis covered the time frame of 2002-2011, including a period of three years before the CIDD implementation. The scope of the evaluation was limited as it focused on insulation measures: data on other retrofit measures related to the CIDD were inappropriate for the evaluations undertaken. The econometric model is also restricted to homeowners because the survey data provided from tenants about retrofits (conducted by owners) can lead to potential measurement bias. Statistics show that the CIDD has only impacted homeowners.

Methodology

The evaluation of the CIDD covers two types of free rider analyses. Data collected from the household survey provides descriptive information about those who are declared free riders. An additional, more comprehensive econometric analysis using this data, provides further information about the share of free riders and what factors influence free riding.

Data from the annual **"Energy Management" (EM) survey**, which is supervised by the French Agency for Environment and Energy Management (ADEME) and conducted by the French market research institute TNS-Sofres, was used for this evaluation. This survey provides detailed information on retrofitting decision processes, retrofit options, household/dwelling characteristics and subsidies received. Over the period 2002-2011, 23,879 households were surveyed with an annual observation of around 6,000 – 8,500 participants. Every year households were asked about their residential energy consumption and if relevant, what energy efficiency improvements were made. An initial questionnaire collected information on:

- Socio-economic variables (income, family size, profession, etc.);
- Dwelling information (type of building, heating energy source, building date, etc.);
- Occupant's situation (occupation status, move-in date, etc.).

Those who completed retrofitting measures (7-12 % annually) were asked to fill out a second questionnaire on:

- Retrofitting categories;
- Investment costs;
- Means of payment;
- Economic/non-economic incentives;
- Other qualitative information (motivation, personal context, satisfaction).

Table 30. Summary of household survey used for the free rider analysis.

Data Collection Survey	Description
Name	Energy Management survey
Scope	Years: 2002-2011 No. of households: 23,879 Average duration of observation: 2.4 years per household Annual observations: 6,000-8,500 households
Recruitment	Carried out by TNS-Sofres, an independent market research company Ensuring representation of all socio-economic profiles
Methodology	Two-tier questionnaire: • First questionnaire: establishes basic household data & determines whether or not retrofitting measures were undertaken within the last year; • Second questionnaire: for those households where retrofitting has taken place
Type of data collected	Household data (e.g. socio-economic variables, housing type, occupation status etc.) Retrofitting data (e.g. retrofitting categories, investment cost, means of payment, etc.)

To estimate the effect of the CIDD on investment decisions the variables indicated in Table 31 were defined.

Table 31. Variables to estimate effect of CIDD on investment decisions.

Type of variable	Variable definition	
Dependent variable	Retrofitting investment decision	
Explanatory variable	Socio-demographic	Annual income of the household
		Socio-professional category
		Family size
		Age of the head of household
	Status of occupation	
	Move-in date	
	Individual preferences	Environmental concerns
		Economic concerns
	Home characteristics	Building completion date
		Building type
		Dwelling size
		Heating energy sources
	Climatic and spatial characteristics	Heating degree days
		Location category
	Former retrofitting	

The descriptive data received from the EM survey can already be used to determine free rider effects in relation to the CIDD. This relates to those households that declare free-riding, i.e. the share of households claiming that CIDD did not affect their decision making.

Additionally, an **econometric assessment** of the CIDD was carried out using the EM survey data. In the econometric model the free rider share is defined as:

$$FRS = \frac{Invest_{Subsidized} - \hat{\Delta}}{Invest_{Subsidized}}$$

Where:
FRS = Free Rider Share;
$Invest_{Subsidized}$ = total number of renovations for which an incentive was requested;
$Invest_{Subsidized} - \Delta$ = the number of investments for which an incentive was claimed, but the household would have made the renovation anyway.

In order to investigate what factors influence the free rider effect, the Random Effect (RE) logit model, that is used to estimate the effect of CIDD on the probability of investing[20], was re-estimated to include interaction variables between the CIDD and individual and housing characteristics. The variables

[20] Details about the specification of the RE logit model can be found on page 85 of the study, and is calculated using equations (2) and (3) (Nauleau, 2014).

considered were: annual income of the dwelling, building completion date, socio-professional category, move-in date, category of city and building type.

Outcomes

According to the survey results, the *share of households* declaring themselves to be “free riders” ranges from 40 % to 70 % of CIDD beneficiaries. Such a declaration is more likely to come from households with a higher income, older (or inactive) householders, and those who have occupied their dwellings for a longer period of time. The econometric model shows that annual rates of free riding decrease from 85 % in 2008 to 61 % in 2010 and increases again to 70 % in 2011. Comparing rates of declared free riders and estimates from the econometric model display the same decreasing trend until 2010, followed by an increase in 2011. The calculated estimates from the econometric model, however, find a lower magnitude of free rider effects than those declared.

Regarding the investigation of *what factors influence the free rider effect,* the RE logit model demonstrates some clear patterns. The socio-professional category, for example, is considered to have a significant effect. The share of free riders is higher for those in the category Business (65 % of CIDD beneficiaries) than for Professionals and Employees (35 %) or the Inactive (50 %) categories. The FRS also increases with income. In the highest income bracket the FRS is estimated at 65 %, while in the lowest income bracket it is at 44 %. It is also higher for those households who have occupied their dwelling from more than 3 years (64 %) than those who have recently moved into their house (27 %). In some instances where the overall CIDD effect varies significantly between categories, such as the location category or building type, the FRS estimates are homogenous. Overall, the free rider effect is considered to be significant, ranging from 40 % - 85 % from 2006-2011, and needs to be taken into account when designing and implementing retrofitting policies.

Want to know more?

- (Nauleau, 2014) Free-riding on tax credits for home insulation in France: An econometric assessment using panel data, Nauleau M., Energy Economics vol. 46 pp. 78-92, 2014, https://doi.org/10.1016/j.eneco.2014.08.011
- (Ecofys & Adelphi, 2018) The Energy Transition Tax Credit (CITE) in France - Fact sheet, Ecofys & Adelphi for BMU, September 2018, https://www.euki.de/wp-content/uploads/2018/09/fact-sheet-energy-transition-tax-credit-fr.pdf

Case #8: Evaluating public policy instruments in the Greek building sector (Spyridaki et al., 2016)

Multi-criteria analysis MCA

Description of policy measure

The policies being evaluated in this case study consist of any measures stimulating energy efficiency and renewable energy development in the Greek building sector with the aim to achieve the 2020 energy savings targets.

Scope of the evaluation

The objective of the evaluation was to assess the policies that were implemented in the Greek building sector with the aim to achieve the 2020 energy savings targets and to use these findings to inform policy recommendations for future. The specific policy measures that were evaluated as alternative policy options were:

- Feed-in tariff for small photovoltaic rooftop systems in buildings (FiT);
- Subsidies for energy efficiency intervention in buildings (Subs);
- Tax reliefs for energy efficiency interventions/renewable energy system installations in buildings (TaxR);
- Energy Performance Certificates (EPCert);
- Energy Building Codes (EBCode);
- Energy Labelling of Appliances (ELabel);
- Green Public Procurement (GPP);
- Energy Performance Contracting (EPContr);
- Public Leadership Programmes (PLP);
- Voluntary Agreements/Co-operative measures (VA).

Methodology

A first step of this study was to set up a research framework consisting of the main research questions, the assessment criteria to be used in the MCA, the policy options to be considered and the stakeholders to be involved. Subsequently, the methodology for assessment was established including rules for scoring and weighting of criteria. This was followed by a data acquisition phase in which the stakeholders were closely involved. Lastly, each policy option was scored and weighted according to the methodology set and a final ranking of policy options was presented. An overview of this methodology is presented in the Figure 23.

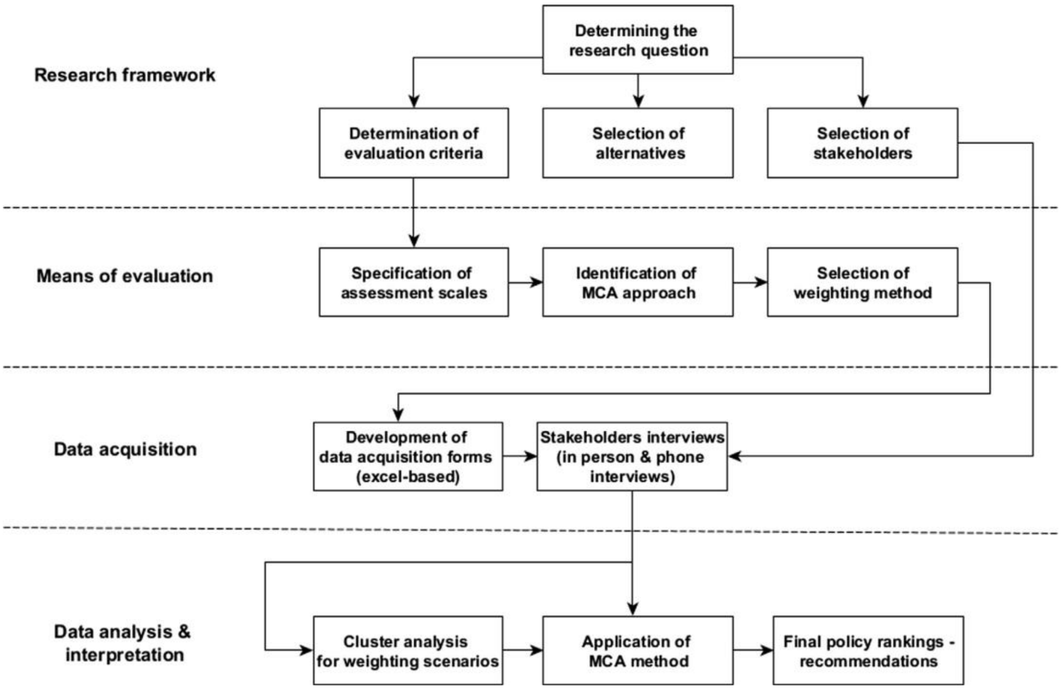


Figure 23. Evaluation methodology used for a MCA on public policies regarding energy efficiency and renewable energy in the Greek building sector.

The assessment criteria used were defined in terms of what they were measuring as well as measurement scales as scores from 0-5. An example of a set of assessment criteria used in this study is presented in the Figure 24.

Criteria	Measurement scales
Incentive to invest/comply (Mot) Strength of the yielded incentives to invest or comply due to policy intervention.	**0:** No incentive at all. **1:** Very low incentive. **2:** Low incentive. **3:** Neither high, nor low incentive. **4:** High incentive. **5:** Very high incentive
Familiarity (Fam) Public awareness associated with the PI through information/advertisements/ official websites.	**0:** No familiarity. **1:** Very little familiarity. **2:** Little familiarity. **3:** Neither high, nor little familiarity. **4:** High familiarity. **5:** Very high familiarity
Fairness in its distribution principles (Eq) Distributional effects associated with relevant benefits and compliance costs among target groups.	**0:** Not at all fair. **1:** Very unfair. **2:** Less unfair. **3:** Marginally fair. **4:** Fair. **5:** Very fair
Adaptability to exogenous changes (Adap) Flexibility in case of exogenous market signals (required time for adjustment) and available options for participation/compliance.	**0:** Not at all flexible. **1:** Little flexibility. **2:** Marginally flexible. **3:** Flexible. **4:** Very flexible
Transaction Costs (Trans) Additional costs accruing of potential barriers (economic, information, or institutional barriers) during policy implementation.	**0:** Very High transaction costs (TCs). **1:** High transaction costs. **2:** Neither high, nor low transaction costs. **3:** Low transaction costs. **4:** Very low transaction costs. **5:** No transaction costs at all
Institutional management & coordination (Coord) Management structures existence of oversight bodies, coordination of policy targets, networks of communication and established information flows.	**0:** Not at all. **1:** Limited coordination. **2:** Neither limited nor adequate coordination. **3:** Adequate coordination. **4:** Very Adequate coordination
Compatibility with national policy strategy (Comp) Addressing relevant market barriers in a way that, synergies and/or contradictions among policies in pursuit of different policy targets and objectives are promoted.	**0:** No compatibility. **1:** Very little compatibility. **2:** Little compatibility. **3:** Compatible. **4:** Very compatible
Institutional set-up and capacity (Inst) Capacity (personnel, available technologies and previous experience of associated regulators) of regulatory authorities to administer and support the implementation of the instrument.	**0:** No capacity at all. **1:** Very low capacity. **2:** Low capacity. **3:** Neither high, nor low capacity. **4:** High capacity. **5:** Very high capacity
Monitoring & control (MnC) Sanctions, inspections and monitoring processes to identify barriers during the execution of the mechanism ensuring compliance are considered.	**0:** No monitoring & control at all. **1:** Very limited monitoring & control. **2:** Little monitoring & control. **3:** Marginally adequate monitoring & control. **4:** Adequate monitoring & control
Financial viability (Fin) The ability of the mechanism to be implemented with low overall costs (operational costs and total expenditure imposed on society).	**0:** No financial viability. **1:** Very low financial viability. **2:** Low financial viability. **3:** Neither high, nor low financial viability. **4:** High financial viability. **5:** Very high financial viability

Figure 24. Assessment criteria used by (Spyridaki et al., 2016) and corresponding measurement scales.

For each of the assessment criteria established, a weighting was assigned based on stakeholder views. Each stakeholder was asked to comment on the relative importance of a criteria in the assessment process. The results of this exercise and the agreed weighting of criteria is presented in the figure below.

Stakeholder groups Criteria	Policy Makers	Energy agencies	RES market actors	ESCOs	Academia	NGOs	Total
Mot	13	16.1	15.8	11.4	15.4	27.3	14.9
Fam	5.9	5.1	9.9	6.8	0.8	9.1	5.1
Eq	15.2	13.0	7.7	13.6	8.3	22.7	12.5
Adap	6.4	8.3	9.9	4.9	7.8	9.1	7.6
Trans	5.6	4.9	3.0	5.8	12.9	4.5	6.7
Coord	7.7	9.4	7.5	12.9	7.9	0	8.2
Comp	7.9	10.2	11.5	8.0	9.9	13.6	9.5
Inst	9	7.6	8.9	16.3	9.6	0	8.9
MnC	14.3	11.8	13.8	12.6	10.1	0	12.2
Fin	15.2	13.5	12.1	7.7	17.2	13.6	14.4

Figure 25. Stakeholder group weights and total average weighting applied to assessment criteria (in %).

Outcome

Instead of setting one preferred assessment methodology, the researcher created three decision-making MCAs to show the influence of the weighting exercise on the final results. The three strategies are:

- Strategy A: in this MCA framework, assessment criteria related to the practical feasibility of *implementation of measures* and the minimisation of barriers for implementation are prioritised. Therefore, higher weighting is applied to elements such as alignment with national policies and sufficient institutional capacity available.
- Strategy B: In this MCA framework, assessment criteria related to *cost-effectiveness of policies* are priorities. Therefore, higher weighting is applied to elements such as the financial viability of options and the incentives these provide for investment.
- Strategy C: In this MCA framework, assessment criteria related to *market competitiveness* are prioritised. Therefore, higher weighting is applied to elements such as return on investment and transaction costs.

Subsequently, each of the policy alternatives was scored against the assessment criteria and the different MCA frameworks were used to aggregate the scores into a final ranking of policy options. The results of this exercise are provided in Figure 26.

The final rankings seem to suggest that tax reliefs is a preferred option for all stakeholders, as it is ranked high according to all three MCA frameworks considered. Likewise, energy efficiency subsidies and voluntary agreements are also ranked high in all three MCA frameworks. On the other hand, green public procurements and energy performance certificates are ranked the lowest in all MCA frameworks.

In addition, the results show that there are no significant differences between the rankings of the different MCA frameworks. This seems to suggest that the results in terms of ranking of policy alternatives are not sensitive to different stakeholder perspectives as considered here.

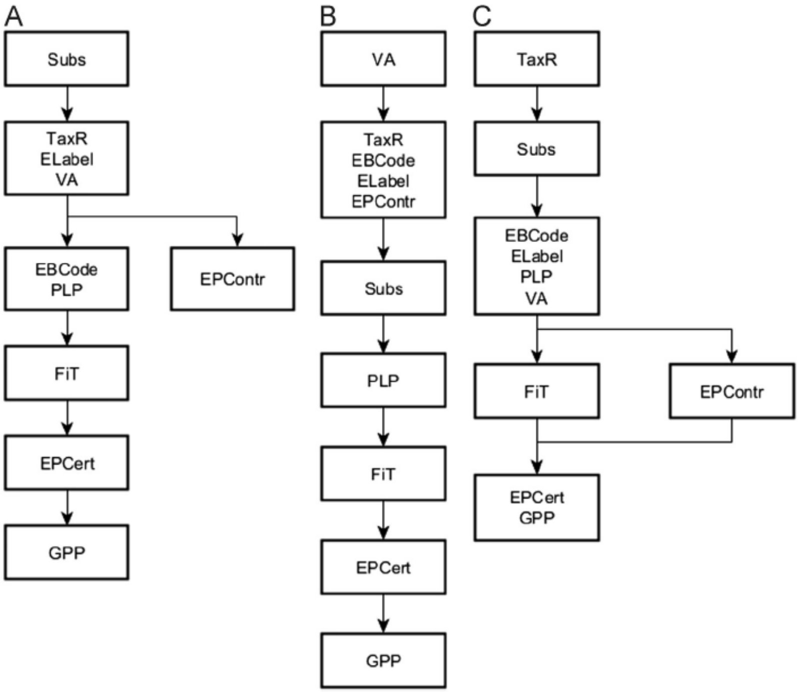


Figure 26. Ranking of policy alternatives for three MCA frameworks (A,B,C).

Want to know more?

- (Spyridaki et al., 2016) Evaluating public policy instruments in the Greek building sector, Spyridaki N., Banaka S. and Flamos A., Energy Policy 88:528-543, 2016, https://www.sciencedirect.com/science/article/pii/S030142151530183X

The Gemini Building
Fermi Avenue
Harwell
Didcot
Oxfordshire
OX11 0QR
United Kingdom
t: +44 (0)1235 753000
e: enquiry@ricardo.com

ee.ricardo.com