ENERGY EVALUATION EUROPE
2020 Conference
Here East, Olympic Park, London
Monday 29 June to Wednesday 1 July 2020
Formerly IEPPEC
Accelerating the energy transition for all: Evaluation's role in effective policy making

One size fits all? Emerging challenges to harmonising energy efficiency evaluations in a changing policy arena

Fabian Voswinkel, Fraunhofer Institute for Systems and Innovation Research ISI, Germany
fabian.voswinkel@isi.fraunhofer.de

ABSTRACT

One of the strategic pillars of the German "Energiewende" is to push for more energy efficiency. As a part of that strategy, the government set up a large Energy Efficiency Fund (EEF) that includes heterogeneous policies ranging from large-scale technology funding programmes for industry to information initiatives for households.

This paper shares the experience with a unified evaluation methodology that was developed for the fund. It presents the lessons learned from applying the methodology in three years of evaluation. The experience has culminated into a made-over Methodology Guideline for future prospective evaluators. The Guideline will be presented in this paper, taking account of how experience has been included in the development of the Guideline and what further challenges came up in the process.

A key goal of the EEF methodology was to make the diverse policies comparable and to address various requirements of different stakeholders at the same time. The latter include the need to increase transparency on public spending and on the government's efforts for climate change mitigation. The Court of Auditors monitors outputs per Euro spent and asks for a clear-cut method for calculating net savings. Further reporting requirements demand specific savings metrics, while policy makers strive to understand strengths and weaknesses of specific policies.

The main challenges for the harmonised methodology were the wide array of programme specifics, a changing political framework with tightening EU regulations in the Energy Efficiency Directive (EED), and different needs of different target groups; a trade-off between user friendliness and legal exactness. The Methodology Guideline therefore aims to overcome those challenges, i.e. being detailed enough to cover all the most important aspects of evaluation while also being broad-cut enough to leave room for flexibility. For example, indicators are suggested rather than fixed, and emissions factors suggest a regularly updated publication instead of a table of values. At the same time, the document keeps up with clear-cut methods for many aspects like savings calculation metrics. The result is increased transparency and comparability. In short, the paper discusses the degree of desired harmonisation.

Introduction

One of the strategic pillars of the German "Energiewende" is giving a high priority to energy efficiency across all sectors following the EU principle "Energy Efficiency First" (Rosenow, Cowart, Bayer, & Fabbri, 2017). Government actions to push for energy efficiency span from retrofitting incentives for buildings to subsidies for electric vehicles and consultation as well as financial support for energy efficiency investments in industry. One key element of the support of energy efficiency is the Energy Efficiency Fund (EEF, further referred to as "the Fund"). It is a wide-range funding scheme by the Federal Ministry of Economic Affairs and Energy (BMWi) and was first started in 2011. In 2018 it had a budget of over 0.5 billion Euros, making it the largest funding instrument for energy efficiency apart from the transport and buildings sector in Germany, contributing about 10% to German notified energy savings for Article 7 of the EU Directive 2012/27/EU on energy efficiency and the amending directive 2018/2002 (further called Energy Efficiency Directive or EED) (European Union, 2012), (European Union, 2018). Of note is that 59% of notified energy savings were stemming from the energy and carbon tax (Federal Republic of Germany, 2020). A challenge to the evaluation of the Fund is its heterogeneity. Before a restructuring in the year 2019, it consisted of 18 different policies ranging from individualised information and consulting campaigns to broad-range financial support schemes. The largest savings in the realm of EED notifications stem from the financial support measures. In particular, the support programme for improved use of waste heat has seen strong growth in the last years with yearly new savings of 7.1 PJ in 2018 (Federal Republic of Germany, 2020).

The ambition of the fund is to form an integral part of achieving the emissions reductions and energy savings targets laid out by the German government in the Energy Concept (BMWi and BMU, 2010). The evaluation has therefore several goals: to monitor the degree of target achievement and energy and emissions savings for international and national reporting requirements, to justify the spending of public money, to disclose information on government action to the public and to gain information on possible weaknesses and potential for improvements. For each of the goals, it is important that the results have been generated in a methodologically sound and comparable way. For this end, an ambitious common methodology was developed for all policies as part of the evaluation process before starting the first yearly evaluation round of the EEF (Fraunhofer ISI et. al, 2019). They included requirements from the EED 2012 (European Union, 2012) and corresponding Guidance notes (European Commission, 2013). The methodology was set apart from former evaluation methodologies in them being equal over a larger policy set. Energy efficiency evaluation has otherwise largely been performed in individual projects setting up method sets for each of these measures posing the challenges for unified interpretation of results (Ringel, 2017) , (Voswinkel, 2019).

Experience with the common methodology has proven mainly satisfactory in the EEF context, however it also presented crucial pitfalls that would not allow the methodology to be applied to a wider range of policies than the confined borders of the Fund and did not leave room for adjustments over time. In a one-year process in close cooperation between the evaluation consortium of the EEF and ministry experts from the evaluations and legal division for more formal aspects and the energy efficiency division for more technical aspects, a Methodology Guideline (further also referred to as "the Guideline") (Fraunhofer ISI et al., 2020) was developed. Since 2020, this has been applied as a mandatory base document for all energy efficiency policy evaluations commissioned by the BMWi, which are mainly performed by external contractors.

This paper aims at presenting the evolution of harmonised energy efficiency policy evaluation in Germany. It describes how experience in practice has given rise to a more distinguished ambition level and more refined methodological details in evaluation systems and how these experiences can benefit other harmonisation initiatives. First, the paper introduces the evaluation system of the Energy Efficiency Fund. The main part of the paper presents the concept and contents of the Methodology Guideline (Fraunhofer ISI et al., 2020) and puts the experienced pitfalls into perspective with the way the Guideline handles them.

The overall principles for evaluation of the Energy Efficiency Fund

The Energy Efficiency Fund is a publicly funded instrument, which is therefore bound to certain obligations laid down in the German constitution. The Federal Court of Auditors is obliged to evaluate government spending. The obligations span three main areas of evaluation: target achievement, impact assessment, and cost-effectiveness (Art. 114 II 2 GG) (Federal Republic of Germany, 1949). Commonly, the Court of Auditors directs evaluation to a responsible ministry, which can then in turn commission external experts to perform the task (Dittrich, 2017). Thereby, evaluation should be comprehensible, accepted, relevant, representative and measurable (Schlomann, et al., 2017). Building upon the 10 principles for good energy efficiency monitoring by the Expert Commission on the Energy of the Future Monitoring Process (2015)[1], a number of methodological considerations were done. Three key areas of evaluation were identified.

1. Objectives: Objectives have to be transparent and clear in advance. The defined objectives of the EEF are (among others) a contribution to the development of a highly energy-efficient economy, the contribution to the achievement of climate protection targets, and the exploitation of existing economic energy saving potentials.
2. Indicators: Indicators are the tool for making the achievements of objectives measurable.
3. Methods for data collection and analysis: considers details on how to generate results for the defined indicators. (Schlomann, et al., 2017)

To satisfy the principles and implement the three key areas, the evaluation system set up a nine-step method starting from the highest objectives and reaching down to methodological details as outlined in Voswinkel (2019), Schlomann et al. (2017) and Eichhammer et al. (2008).

1. Identification of general characteristics of the policy (e.g. available funding, administrative framework, funding party, target group, administrative issues and supported technologies of the policy)
2. Identification of framework conditions (e.g. underlying assumptions like energy prices, primary energy factors, GHG-emissions factors)
3. Review of policy targets. The policy targets are the basis for the definition of indicators (e.g. reduction of GHG emissions by 1 Million tonnes by 2020)
4. Definition of an indicator set based on policy targets (e.g. reduction of primary energy consumption, reduction of GHG emissions)
5. Data collection for analysis of defined indicators keeping in mind that data types should be comparable among policies as far as possible
6. Data analysis for gross values of indicators using well defined methods for comparability
7. Adjustments for baseline and effects like the free-rider or spill-over effect generating net values of indicators using equal methodologies.
8. Calculation of future projections. This step can also be useful for overarching goals like reduction in energy intensity of industry
9. Summation and comparison of different policies in an overarching evaluation project

These steps were used a as a basis for the development of the common EEF evaluation methodology. For each of these steps, methodological details, including reporting requirements for Article 7 EED (European Union, 2012), were defined and laid out in the common methodology for the evaluation of the fund. Many of

---

[1] Expert Commission on the Energy of the Future Monitoring Process (2015):
[The monitoring]

- identifies the most appropriate policies and programs for the relevant fields of action,
- adopts a suitable system of indicators,
- is based on a sufficiently reliable and up-to-date data ,
- has a suitable methodology to assess the effectiveness of instruments and measures, particularly taking into account endogenous and exogenous factors,
- distinguishes between direct and indirect effects,
- takes account of distributional effects,
- examines whether the effect of instruments is sustainable,
- reviews the efficiency of instruments and measures,
- can itself be implemented efficiently,
- is transparent and neutral

these methodological aspects have been adopted into to the more generalised Methodology Guideline developed consecutively. Other aspects were left out, changed or added based on the experience with the application of the common methodology and lessons learned. The following section presents the concept of the generalised Methodology Guideline for energy efficiency evaluations of the BMWi and its structure quoting selected defined methods and shares experience and challenges with the harmonisation approach.

The Methodology Guideline

The Methodology Guideline (Fraunhofer ISI et al., 2020) was developed in an effort to make evaluations conducted by different external experts more comparable and enable a meaningful summation of results over different policies in the realm of national reporting in the National Action Plan for Energy Efficiency (NAPE) framework and EU reporting under Article 7 of the EED. Since 2020, it has been adopted as a mandatory reference document for energy efficiency policy evaluations commissioned by the Federal Ministry of Economic Affairs and Energy (BMWi). A harmonised methodology is of particular importance in the German context because in its energy efficiency strategy (BMWi and BMU, 2010), the government has opted against an Energy Efficiency Obligation Scheme (EEOS) and in favour of a large array of alternative measures including financial support and information programmes for energy efficiency, in line with Article 7 of the EED (European Union, 2018), making comparability an important issue.

The concept of the Guideline is to cover a wide range of issues that can be handled in different manners. That includes, among others, wording definitions, the logics of impacts, definition of indicators, cross-cutting definitions, baseline and net impact calculations as well as methods for aggregation and forecasts.

The target groups of the Guideline are both evaluators--to include this information in their own evaluation approaches--and ministry or administrative agency employees, who take the information into account when designing a policy, setting up the data collection system, and interpreting results. It has proven to be a challenge to reconcile the different target groups and their individual requirements from a guideline. While evaluators' main interest is clear guidance on how to proceed in calculations and results presentation, political actors need to know what the method guidance means for policy and what it implies in design and data collection decisions. Finally, the evaluation commissioners (i.e., ministry representatives) need information on the implication of the methodological choices for interpretation and reporting. Figure 1 presents the interests of the two main target groups of the Methodology Guideline. It shows how both groups have different priorities and use partly diverging vocabulary leading to trade-offs in the suitability of the Guideline for both groups.

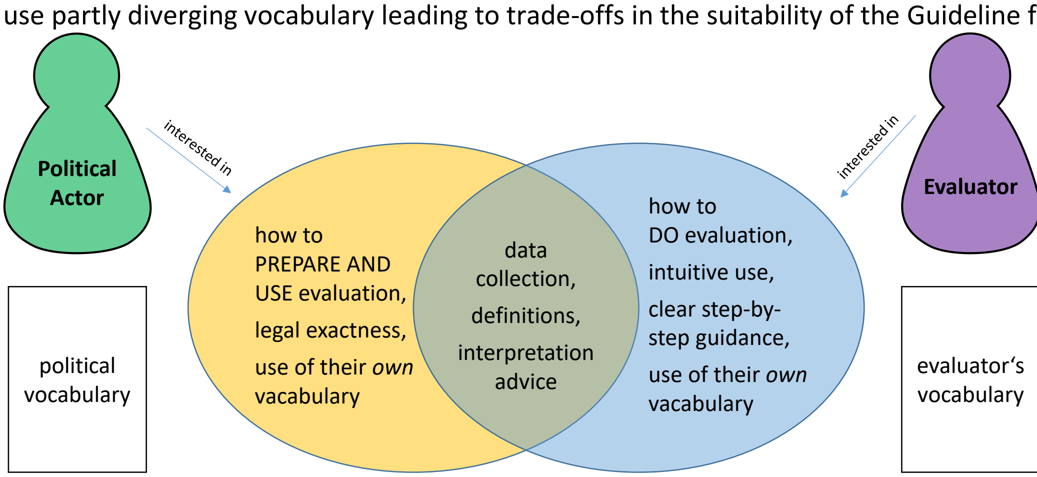

Figure 1: Illustration of target groups and their interest in the Methodology Guideline. Source: own representation

A stakeholder dialogue has led to the solution to include a two-way approach. While the main part of the Guideline is primarily destined at evaluators and gives the most detailed account of methods, grey boxes at particular positions condense the information that is most crucial for policymakers and complement it with additional interpretation advice.

The Methodology Guideline is structured according to the order in which evaluators will mostly need to plan their evaluation processes. After a brief introduction describing the goals of the Guideline, it starts with one of the most crucial aspects, the definition of terms.

Definitions

In previous evaluations, it often became clear that different target groups as well as different individual evaluators use differing terminologies. That can be seen as one of the prime barriers to sound comparisons and summations of results. This issue applies both in the German language context as well as in the international context. The Horizon2020 project EPATEE has worked on a framework of energy efficiency evaluation methods on a European and worldwide level and has encountered the same issues (Broc, et al., 2019). A good example of differences in definitions is the word "measure". It can be understood on different levels of an evaluation process and is defined differently among different sources. A measure can mean a policy measure, synonymous to "instrument" or specifically to "programme" in the context of a policy mix. On the other hand, it can be found to be used on an individual level describing the individual action, e.g. the installation of a highly efficient electric motor. Finally, on a political level it can even describe a set of policies with a common goal. Similar problems arise with the terms "programme" and "policy". In the German context for the Methodology Guideline, the word measure ("Maßnahme" in German) is used differently by different entities. Evaluators commonly use "Maßnahme" to describe the individual energy efficiency action. An energy efficiency policy would rather be called "Programm". On the other hand, §7 of the German Federal Budget Code (BHO) (Federal Republic of Germany, 1969) defines an energy efficiency policy as "finanzwirksame Maßnahme" (financially impacting *measure*). Hence, the legally correct term in line with the language used by the ministry and public administration would be to use "Maßnahme" for the policy rather than the energy efficiency action. Similarly, the EED defines the term "policy measure" in a similar sense as the German BHO (European Union, 2012). Finding an intuitive term for the energy efficiency action in German in turn proved rather difficult, leaving some sentences more complicated to understand. The challenge that has surfaced in this example as well as in several other occasions, is the trade-off between intuitiveness for use as a handbook and legal correctness. Such trade-offs can hence pose as a conflict of interest between the two main target groups (evaluators and political actors) (see Figure 1 above). While evaluators are less concerned with legal correctness but prefer an easily understandable guideline, policymakers are concerned with legal terms, especially because the Guideline is meant to be a mandatory guidance document and will be distributed in official tenders with possible legal implications.

In the next section, the Guideline outlines legal and formal backgrounds. It includes requirements laid out in German and European law, in particular concerning energy savings reporting requirements as part of the EED and Energy Performance of Buildings Directive (EPBD) (European Union, 2018) and on a national level in the NAPE. Financial support programmes are often affected by EU competition law (European Commission, 2014). All policies with an average yearly budget over 150 Million Euros are subject to an evaluation plan approved by the European Commission (European Commission, 2014).

Cross-cutting aspects

In the next section of the Guideline, cross-cutting aspects are introduced including characterisation of participants for possible clustering in the industrial sector, in private households and in the transport sector. The handling of energy prices, emissions and primary energy factors, technology lifetimes, discount factors and units of measurement is introduced. Different from the methodology of the EEF evaluation, the underlying factors are not stated as absolute values because many of them vary over time. For the CO_2-emissions and primary energy factors of electricity, an accepted public source provided by the German Environment Agency (Umweltbundesamt, UBA) is quoted and linked (Umweltbundesamt, 2020). It is noted that the most recent value is to be taken as the basis. For fuels, these values do not change to a large extent over time, and thus absolute values are included for these. This approach enables the Methodology Guideline to be used for a longer period of time without the need for updating.

According to steps 1 to 3 of the evaluation system outlined above, general characteristics and framework conditions are explained in a systemic way presented as a logical framework of intervention (i.e. the way, policy measures influence energy efficiency investments and behaviour, as well as the energy and economic system). It establishes the connection between political framework conditions like the budget code and political strategies with the concrete policy and its target. The analysis of target achievement then follows logics from input (financial means) through output (e.g. number of participants), outcome (e.g. number of installed efficient devices) to impact (e.g. reductions of GHG emissions). A distinction has to be made here between policies that directly lead to a target achievement (e.g. financial support programmes) and such policies that indirectly lead to a target achievement (e.g. information campaigns). Figure 2 illustrates the logics of impact for the two mentioned examples. It should be noted that the reduced energy consumption is a direct impact of financial measures while the direct impact of informational measures is the behavioural change that can lead to reduced energy consumption in the next step (indicated by the exclamation mark). The exact methods of how to quantify energy savings from informational measures is not part of the Guideline because such influences are too particular for each measure to generalise in a guideline document. However, general suggestions are made.

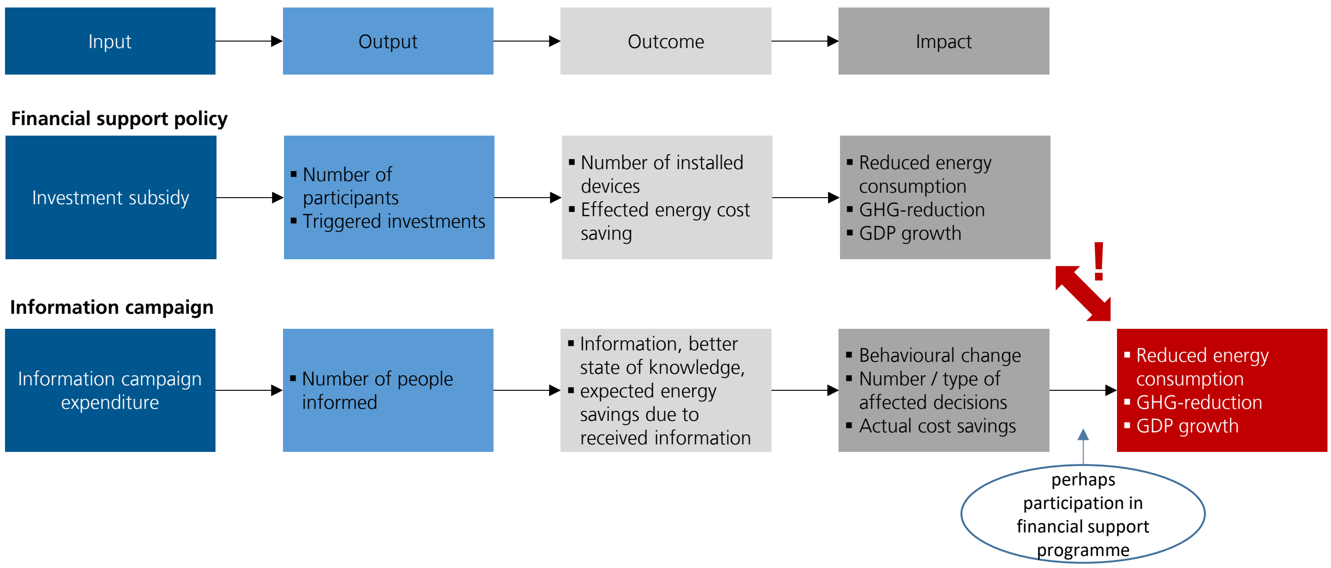

Figure 2: Logics of impact for financial support programmes and information campaigns. Source: Methodology Guideline (Fraunhofer ISI et al., 2020)

Target system

In the next section, the Guideline describes the target system as the connecting piece between the logic of impact and the definition of indicators. Indicators have to be defined for the analysis of a given target. Such targets are, on different levels, politically defined or deduced from political strategies and decisions (Step 3 of the evaluation approach). On the highest level, there are the strategic political targets in the energy field. On the level of evaluation projects, the goal of the evaluation is defined, including which reporting requirements are meant to be satisfied. Finally, what are the specific targets of the policy for example in terms of energy savings? For a meaningful target achievement analysis, it is important that the targets are defined according to the S.M.A.R.T. criteria (Specific, Measurable, Accepted, Realistic, Time-based) (Schlomann, et al., 2017). Further, under certain circumstances the individual energy efficiency action can be included as a lower layer of the target system. That is particularly useful for flagship projects in the analysis of qualitative indicators such as innovation potential.

Indicators

Step 4 of the 9-step evaluation approach deals with the definition of indicators. The German Federal Budget Code (Federal Republic of Germany, 1969) foresees three areas of evaluation: the target achievement, the impact assessment, and the assessment of cost-effectiveness. Additionally, it is of vital interest for policymakers to obtain information on programme administration performance. Indicators are therefore defined to generate information on the prior defined targets on different levels and to satisfy the foreseen evaluation areas. Indicators can in general be either quantitative or qualitative in nature. The Methodology Guideline suggests a long list of possible indicators than can be used to answer most evaluation goals in the realm of energy efficiency policy. It does not claim completeness nor defines a set of mandatory indicators. Policies are too varied in nature to be able to generate an equal indicator set for all. However, it does stress certain indicators that are recommended to use in most evaluations. For a better overview, indicators are grouped in seven "criteria", or indicator categories starting with the mainly quantitative indicators of energy savings (in MWh, PJ or ktoe) and GHG emissions reductions (in tonnes of CO_2-eq.) and the cost-effectiveness (which is defined as 'funding efficiency' in the German budget code, e.g. in MWh of savings per € of public funding). The Guideline furthermore suggests indicators for further economic effects (e.g. reduction of energy cost, triggered investments or macroeconomic effects) and for acceptance and diffusion (e.g. number of participants, target group coverage or regional distribution). Finally, primarily qualitative indicators are suggested for flagship model character (e.g. visibility), consolidation potential and the administrative performance (e.g. administrative costs, lead time for processing per application).

For quantitative indicators, the Guideline defines four different metrics and indicates for which reporting requirement or other evaluation goal they are used. Based on (1) the first-year new savings (in energy units per year), (2) the cumulated annual savings (in energy units per year) can be calculated for actions implemented in several years. Furthermore, (3) the periodically cumulated savings (in total energy units) present the total savings over a certain period. Finally, (4) the lifetime savings are the savings that are expected to be achieved until the end of each action's lifetime. The recommendation states for lifetimes to use the values from the Guidance notes to the EED 2018 recast (European Commission, 2019) unless the particular evaluation contains higher quality data on lifetime. Figure 3 shows the savings metrics in graphical form. The EED employs periodically cumulated savings, however it uses fixed reporting periods independent from the year of implementation of the policy. The last EED reporting period went from 2014 to 2020, while the next one will last from 2021 to 2030 (European Union, 2018).

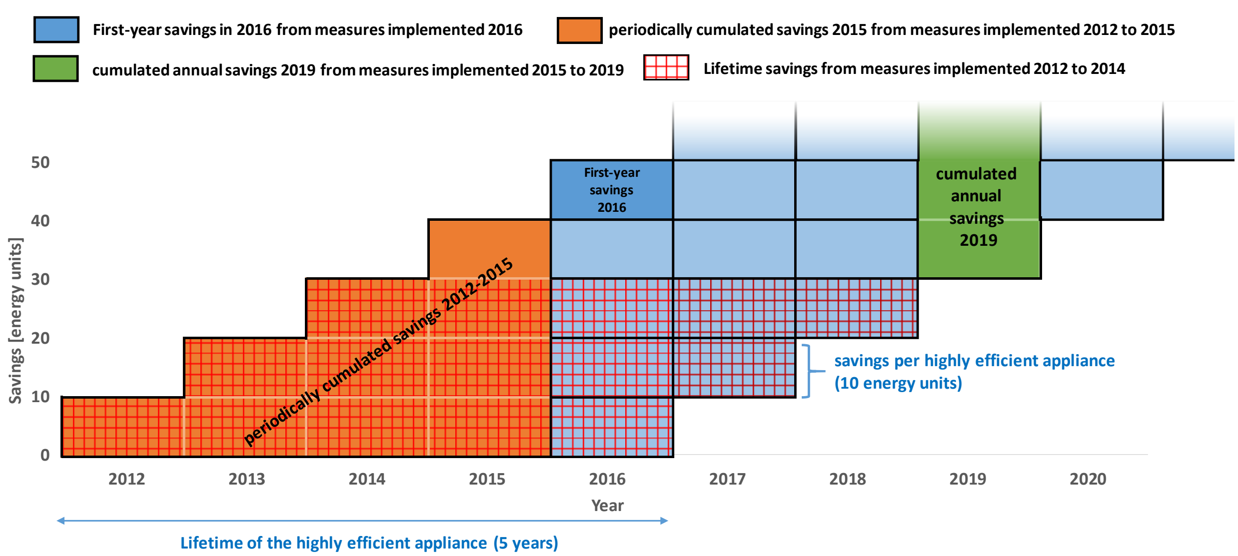

Figure 3: Accounting metrics illustrated for a generic appliance with lifetime of 5 years and yearly savings of 10 energy units. Source: Voswinkel (2018)

Further methodological details, particularly gross-to-net adjustments

In the following sections, the Guideline gives recommendations on several other methodological issues relating to steps 5 to 7 concerning e.g. univariate and multivariate analyses, sample sizes or methods of data collection. It then goes on to the analysis from gross to net values of energy savings and emissions reductions. The Guideline distinguishes gross and net in two steps. First is the adjustment for a baseline of energy consumption or emissions (what would have happened without the policy intervention?). It is required by Article 7 of the EED that other policy measures like Ecodesign minimum energy performance standards (MEPS) are not counted towards the evaluated policy. Often, the baseline correction is already done in the collected data using reference investments on the energy efficiency action level. However, sometimes it is not. Therefore, the terms "before-after gross" describing the pure difference between the state of energy consumption before and after the energy efficiency action, and "baseline-gross" for the savings after correcting for how the energy consumption would have changed anyway, are introduced for formalisation and transparency. Furthermore, effect adjustments for effects such as the free-rider effect or the spill-over effect are conducted, leading to the net values. Finally, interaction effects between different policies can generate another level of net savings that is relevant only in the analysis of a set of policies, which creates the need to avoid double-counting of savings from two policies addressing the same type of actions by the same target group using an interaction matrix (see also Table 2 further below). Figure 4 illustrates the steps of gross-net calculations.

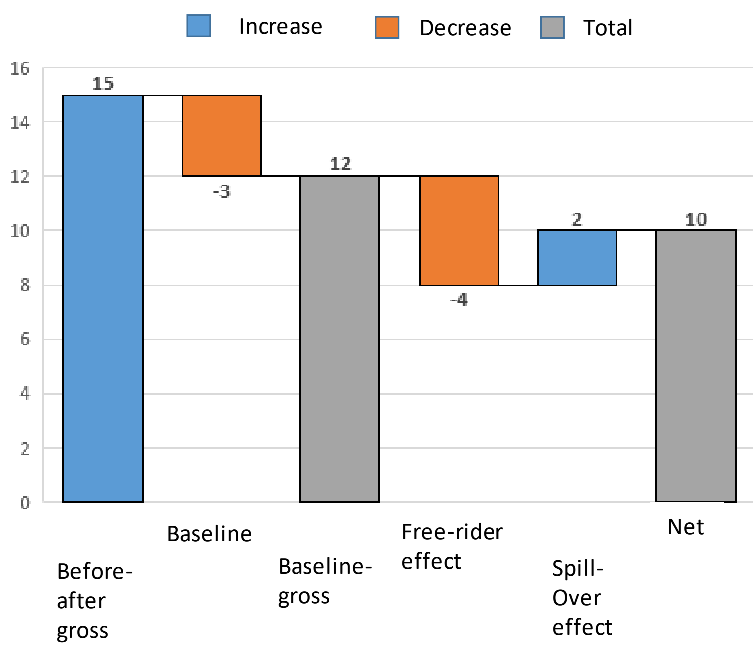

Figure 4: Graphical representation of gross-net adjustments. Source: Methodology Guideline (Fraunhofer ISI et al., 2020)

Different methods of calculations for effect adjustments are presented based on Violette & Rathbun (2017). Advantages and disadvantages are included for practical evaluation. An exemplary calculation presents effects possible to be included in a gross-to-net calculation and the way they are connected with each other (i.e. additive, subtractive or multiplicative). The example calculation is shown in Table 1. Due to the politically sensitive nature of effects adjustments and the uncertainties inevitably connected with them, interpretation advice is included in a rather detailed highlighted grey box.

Effect	Description	Example savings
Before-After Gross	Direct comparison energy consumption of example system before and after energy efficiency action	**1250 MWh**
- Baseline	Savings relative to reference investment or MEPS.	-250 MWh
Baseline-Gross		**1000 MWh**
- Free-rider effects and early implementation effects	Effects due to free-riding or the earlier than planned adoption of actions (equal to a time-lagged free-rider effect)	-38 % = -380 MWh
+ Spill-over-Effect	Effects due to Spill-over onto third parties or different functional units.	+13 % = +130 MWh
+ Time-lagged effects	Effects due to savings taking effect later and outside the evaluation period.	Not calculated here
Sub-sum additive effects		**750 MWh**
*** -** Structural effects	Effects due to changing of central structural variables (e.g. varying weather conditions during the project lifetime)	*(100 % - 5 %) = *95 %
*** -** Interaction Effects	Effects due to interactions of different measures	*(100 % - 10 %) = *90 %
*** -** Rebound-Effects	Effects due to increased use as a result of lower unit-cost	*(100 % - 8 %) = *92 %
= Net impact	Impact after effects adjustment	**589,95 MWh**

Table 1: Example calculation of effect adjustments for an illustrative generic energy efficiency action. Source: Methodology Guideline (Fraunhofer ISI et al., 2020)

Finally, the Guideline presents methods for savings projections relating to step 8 and presents an example for an interaction matrix for the interaction effects in evaluations of a policy bundle to avoid double-counting of measures targeting the same target group or type of measure relating to step 9 as shown in Table 2.

	Indicator value before interaction	Measure 2: Energy savings meter	Measure 3: Energy Management Systems	Measure 4: Waste heat flagship project	Measure 5: SME energy efficiency support	Measure 6: Cross cutting technologies	**Interaction Factor**	**Contribution of the measure to the policy mix**
Measure 1: Waste heat programme	**50**							**36**
Interaction Factor $\propto_{i,j}$		0 %	10 %	20 %	0 %	0 %		
Factor $(1-\propto_{i,j})$		1	0,9	0,8	1	1	**0,72**	

Table 2: Interaction matrix on the example of a waste heat reduction support programme. Source: Methodology Guideline (Fraunhofer ISI et al., 2020)

Lessons learned and approach in the Methodology Guideline

This section is based on the pitfalls ("catches") from Voswinkel (2019) that include experience and lessons learned from the practical application of the common methodology for evaluation of the Energy Efficiency Fund described above. The numbering of the Pitfalls does not necessarily coincide with the number of catches from Voswinkel (2019). This paper presents the way these pitfalls have been dealt with in the newly developed Methodology Guideline.

Pitfall 1: Flexibility in cross-cutting aspects

Step 1 of the evaluation process concerns general characteristics of the policy like available funding, or the administrative framework. These characteristics lie at the core of evaluation systems and have to be consistently accounted for. However, since these characteristics can change as part of a policy cycle, a methodology specifically set up can render unfit for the changed policy requirements. In the Energy Efficiency Fund, as well as in most other energy efficiency programmes in the EU, energy savings have to be reported as part of Article 7 of the Energy Efficiency Directive (EED). In the evaluation methodology, the four savings metrics described in the previous section (Figure 3) were calculated for each indicator to guarantee maximum usability for different reporting schemes. The time frames were chosen according to the evaluation cycles (i.e. from 2011, the implementation of the Fund or the implementation of the individual policy, whichever is later, until 2015, 2016 and 2017. Conversely, the EED foresees periodically cumulated energy savings for the fixed time frames 2014 to 2020 required by Directive 2012/27/EU (European Union, 2012) and 2021 to 2030 as required by Directive 2018/884/EU (European Union, 2018). Hence, they do not coincide with EEF evaluation cycles. The German national NAPE reporting is based on cumulated annual savings. The Methodology Guideline takes account of that by introducing all savings metrics and putting them into perspective with the most relevant reporting requirements. It lists which additional detailed factors, such as predefined savings periods, lifetime assumptions, attribution of savings to timeframes (also see Pitfall 4), can vary or change in these requirements prompting evaluators to choose their calculation and reporting methodologies accordingly (Fraunhofer ISI et al., 2020).

Pitfall 2: Learning effects

Pitfall 2 concerns learning effects on the sides of all stakeholders of an evaluation. Because there is no such thing as a perfect evaluation, certain assumptions and simplifications have to be made to make evaluation viable. While an evaluation system and methodology has to be defined in the beginning of a project and then applied, oftentimes, over a number of years, the learning effects on all sides will make it necessary to make fine adjustments. In a policy cycle, priorities on certain aspects of evaluation may change, calling for a flexible design of the methodology. In the EEF, the calculation of adjustment effects like the free-rider effect have moved more strongly into the focus during the project duration. While a Methodology Guideline cannot foresee all methodological details that come into focus in the future, it can present methodologies in sufficient depth to form a starting point for such changes in priorities. The challenge herein was to find a middle ground between methodological detail and flexibility to react to conditions affecting evaluation as presented in Pitfall 1. In the Methodology Guideline, a way was chosen to formulate most of the aspects as recommendations and present practical issues to take into account respectively. This approach gives policymakers the possibility to require certain aspects to be put into practice obligatorily. On the other hand, it leaves the flexibility to pick out the set of methodological aspects that is fitting for the evaluation project in the hands of the evaluators.

Pitfall 3: Joint effort on data collection

Step 5 of the evaluation process concerns data collection. Oftentimes, evaluators receive data that is collected during the implementation of the policy by the administrative agency. A common trade-off exists between the exactness of data and the cost of data collection associated with it. While metered savings can be considered the most reliable method, if the meters are able to delineate the system subject to the analysis appropriately, they are also very costly to collect (Violette & Rathbun, 2017); (Hannigan & Cook, 2015). In the EEF, physical measurements are not performed and the underlying data are collected by the implementing agents using model calculations performed by participants or their energy consultants. This method is possibly less exact but more viable to administer and finance in a large-scale programme. The Guideline helps to consider the choice of data used and their respective uncertainties for the interpretation of results. In order to obtain the data, an evaluation system relies on their availability and on the compliance by particular methodological aspects. For example, the baseline calculation is often already performed ex-ante as part of the energy savings concept using a reference investment with market average or minimum standard technology (Broc, Adnot, Bourges, Thomas, & Vreuls, 2009). It is important that already at the stage of policy design and on the side of policymakers and policy implementers, such issues are taken into account and the choice of data to be collected in the administration process is made accordingly. The Guideline aims to help policymakers and implementers in structuring their administrative processes with evaluation issues in mind well before the evaluation project is starting. Grey boxes highlight information in all sections that are of particular interest for policymakers and implementing agents. Data collection and methodological choices are among the key elements that policymakers should consider early on in the process. The Guideline can thereby serve as a communication node between the needs of evaluators and policymakers.

Pitfall 4: Timeframe attribution

Step 6 of the evaluation process concerns data analysis. It encompasses most methodological choices detailed in the Guideline. These choices can have a crucial impact on results and should therefore be made in a well-informed manner. One important pitfall for the comparability of evaluation results is related to time. Most evaluations use years, quarters or months as units of analysis. Hence, savings have to be assigned to a certain timeframe. What appears to be trivial proves complicated in practice. Impacts of a policy often only become apparent a long time after first participating in a programme. For example, it can take several months between handing in the application and receiving the acceptance notice. Afterwards, the technical implementation phase of a systemic energy efficiency measure can take up to several years. The long time lag can be impractical for evaluation because the possibility to make adjustments for policymakers is not given. Hence, it is often practical to assign savings not to the finalisation date but to another date. For information measures, often the only option is the date of participation or application. Financial measures also offer the date of acceptance. The Guideline presents advantages and disadvantages of each timeframe and elaborates on possibilities to adjust the assignment of savings to the finalization year using expected project runtimes. The necessity for the adjustment is not always given. Hence, the Guideline merely includes information on possible methods while mentioning that this choice should be made in close cooperation with the evaluation commissioner.

Pitfall 5: Cautious interpretation

Step 9 of the evaluation system considers comparisons of indicator values between different programmes and their interpretation using the unified methodology. A unified methodology can falsely suggest that all indicators are directly comparable. However, this is often not only dependent on methodology but also on programme specifics. While with a common methodology it is possible to make a comparison, the interpretation of the comparison is often not valid. A good example is the 'funding efficiency' indicator. It performs a simple cost-benefit comparison, calculating the amount of public money spent (cost) per energy unit saved (benefit). With a common methodology, the indicator can therefore provide information on which of a set of programmes is performing better in terms of funding efficiency. However, the interpretation that the one programme is therefore better than the other is oftentimes not valid. Specifics of each programme play a crucial role in that assessment. As an example, a programme that funds small cross-cutting technologies focussing on preventing barriers to participation using a simplified application procedure may have a worse funding efficiency due to its generalised and non-specific nature, than a programme supporting individual large systemic measures that require an elaborated energy savings concept that can target individual energy savings potentials more precisely. Policy makers are subject to scrutiny for justifying the spending of public money. The funding efficiency value is often most easily at hand for criticising a policy's performance. The Methodology Guideline therefore dedicates a section specifically to argumentation advice for policy makers concerning the funding efficiency indicator. A fundamental alternative to this indicator for cost-benefit analysis could be the total resource cost test that is commonly applied in the US. It relates all resources spent (cost) and output (benefit, notably energy cost savings) generated of a project from the administrative, public funding and private side. The concept allows to include external costs of carbon emissions as well as non-energy benefits of the energy efficiency investment (Felder & Athawale, 2018). With an enhanced method, it is possible to make more accurate conclusions, however possibly requiring more data and assumptions to be included in a harmonised way.

Similarly, indicator values often are difficult to compare due to the runtime of policies. While a policy at the beginning of its cycle is not yet very well known, energy consultants are not familiar with the application procedures and therefore barriers to participation are higher. At a later stage, these barriers become weaker. This can lead to higher participation rates and an overall better performance. A comparison of two policies is therefore not always directly representative. The Methodology Guideline addresses this issue in the general section on quantitative and qualitative indicators with interpretation advice for policymakers.

Pitfall 6: Free-rider effect

While the baseline issue for comparability was mentioned in Pitfall 3, this pitfall concerns the second step of the gross-to-net calculations, the effect adjustments. There is a large number of possible methods for calculating effects such as the free-rider effect. Most notably a percentage can be determined using group comparisons (randomised controlled trials or quasi-experimental methods) or using surveys (Violette & Rathbun, 2017). While the former is the more exact measure, it is often difficult to determine an exact comparison group for the analysis. In practice, data protection laws often do not permit to get in touch with the non-participants. In addition, valid group comparisons are mostly based on physical measurements, which bear the disadvantage of a long time-lag between programme participation and measurable savings (see Pitfall 4). Therefore, a survey approach is often used to determine the free-rider effect (Northeast Energy Efficiency Partnerships, 2016) (Johnson, 2014). That approach is subject to a number of biases including self-selection bias and social desirability bias (Bundi, Varone, Gava, & Widmer, 2018). A more precise approach which employs the concept by Olsthoorn et al. (2018) is included in the Guideline. It includes a partial free-rider effect for cases where the energy efficiency action would have also happened without the policy but to a lesser extent. In this case, only the lesser action is subtracted and the difference to the stronger action is attributed to the policy. Furthermore, it includes weak and strong free-riders. The concept is based on the notion that financial support measures contain two logics of impact, the direct financial support logic and the information logic. Participants with a weak free-rider effect might have not needed the financial support to implement the measure, but the existence of the programme and the information coming with it were conducive to implement the energy efficiency action. In this case, the part of energy savings due to the not needed financial support has to be deducted from policy impacts, but a part for the information content remains in count. Finally, deferred free-riders would have implemented the action at a later time, hence only from the time of the originally planned implementation, a free-rider effect has to be subtracted. Deferred, partial and weak free-riding effects can occur combined with one another (Olsthoorn, Schleich, Gassmann, & Faure, 2018). The difficulty however exists when comparing evaluation results that use different approaches, or in the case of the survey approach, even a different set of questions. The Methodology

Guideline presents different methods for free-rider effect calculation including their advantages and disadvantages in practice. It generates consciousness for the comparability issues and the necessary additional layer of uncertainty that a net effect analysis introduces. A section for policymakers gives advice on communication of effect adjustment and net values of energy savings and emissions reductions for the presentation of their policies.

Discussion and Conclusions

Evaluation methods are constantly being extended and are accompanied by learning processes. Factors ranging from political strategies to methodological priorities or the advancement of data collection and scientific methodological knowledge give rise to a changing evaluation arena. Building upon the experience with a unified methodology for the evaluation of the German Energy Efficiency Fund (EEF), a Methodology Guideline has been developed. While the EEF methodology largely proved to be a solid basis for a sound and comparable evaluation in the context of its designated policy set, the EEF, it displayed certain weaknesses and pitfalls that made it not viable for a broader context. In order to apply a unified methodology to a larger number of policies across all sectors and to guarantee compatibility of methods with EU requirements under the EED and national requirements under the NAPE, the scope had to be opened. The methodology is geared towards being both valid for a longer period of time and open to a wider range of programme characteristics. This paper presented the experience with the development of the harmonised Methodology Guideline. It presented approaches to reconcile different target groups of the Guideline accounting for both intuitiveness and legal exactness. The general approach of the Guideline is of suggesting rather than prescribing and detailing a large array of methods with advantages and disadvantages. This way it gives both an in-depth understanding of methods and leaves room for flexibility for a manifold energy efficiency policy landscape.

Certain levels of ambition in harmonisation for higher comparability in detail had to be reduced for that end. Experience with the EEF evaluation methodology has shown that this was a necessary step in a changing policy landscape. It can be argued, that comparability of results is therefore still not entirely given this way. However, stricter harmonisation rules may lead to two unwanted side-effects. Results may display inconsistencies due to incompatibilities of the prescribed method with a characteristic of the policy. Further, the harmonised methodology can take over the part of the ”predefined state” forcing policymakers to design policies constrained by the evaluation methodology. Innovative policies to reach also the not-so-low-hanging fruit would be difficult to implement. The Methodology Guideline therefore focusses on increasing transparency in order to generate comparability. For most key aspects of evaluation, it suggests different ways of action reminding to make the used methods transparently clear in the evaluation studies. This approach should make it easy for readers of the studies to draw conclusions on the extent to which results can be directly compared and on how methodological differences should be accounted for when interpreting results. The improved transparency can in turn facilitate validation of EED notifications on the part of the European Commission or its contractors. The EEF evaluation project had a runtime of four years to make experiences with its common methodology. With harmonisation becoming increasingly relevant with strengthening ambitions in energy savings and climate targets and increased supranational regulation of policies, the lessons learned and the way they have been handled in the Guideline can serve as valuable insight for upcoming harmonisation approaches in other fields or other countries. Further research should investigate on strengths and weaknesses of the Methodology Guideline both in its practical application, its theoretical foundation and scientific exactness. The Methodology Guideline is publicly available (Fraunhofer ISI et al., 2020) and has since 2020 been applied as a mandatory base document for all energy efficiency policy evaluations commissioned by the Federal Ministry of Economic Affairs and Energy (BMWi) opening the field to a wide range of experiences in real world evaluation studies.

References

BMWi and BMU. (2010). Energy Concept for an Environmentally Sound, Reliable and Affordable Energy Supply. Retrieved 01 12, 2018, from Energy Concept for an Environmentally Sound, Reliable and Affordable Energy Supply: http://www.osce.org/eea/101047

Broc, J.-S., Adnot, J., Bourges, B., Thomas, S., & Vreuls, H. (2009). The development process for harmonised bottom-up evaluation methods of energy savings. (W. I. EMEEES. Edited by IEE, Ed.)

Broc, J.-S., Breitschopf, B., Schlomann, B., Voswinkel, F., Maric, L., & Thenius, G. (2019). Sharing experience to spread evaluation practices: evaluation does help to improve policies! *Proceedings of the ECEEE Summer Study 2019*. Retrieved from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/3-policy-and-governance/sharing-experience-to-spread-evaluation-practices-evaluation-does-help-to-improve-policies/

Bundi, P., Varone, F., Gava, R., & Widmer, T. (2018). Self-Selection and Misreporting in Legislative Surveys. *PSRM (Political Science Research and Methods), 4*, pp. 771-789.

Dittrich, D. N. (2017). *Bundeshaushaltsordnung (BHO) Kommentar.* München, Heidelberg: rehm.

Eichhammer, W., Boonekamp, P., Labanca, N., Schlomann, B., & Thomas, S. (2008). Distinction of energy efficiency improvement measures by type of appropriate evaluation method. *Evaluation and Monitoring for the EU Directive on Energy End-Use Efficiency and Energy Services*.

European Commission. (2013). Commission Staff Working Document. Guidance Note on Directive 2012/27/EU on energy efficiency.

European Commission. (2014). Allgmeine Gruppenfreistellungsverordnung (AGVO). Retrieved 11 12, 2020, from https://www.gesetze-im-internet.de/euv_651_2014/AGVO.pdf

European Commission. (2014). Guidelines on State aid for environmental protection and energy 2014-2020. *Official Journal of the European Union*. Retrieved 11 13, 2020, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52014XC0628(01)

European Commission. (2019). Commission Recommendation 2019/1658 on transposing the energy savings obligations under the Energy Efficiency Directive. Retrieved 11 13, 2020, from http://data.europa.eu/eli/reco/2019/1658/oj

European Union. (2012). Energy Efficiency Directive. *Official Journal of the European Union*.

European Union. (2018). Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 on the energy performance of buildings. Retrieved 11 12, 2020, from http://data.europa.eu/eli/dir/2018/844/oj

European Union. (2018). Energy Efficiency Directive. Retrieved from http://data.europa.eu/eli/reg/2018/1999/oj

Expert Commission on the Energy of the Future Monitoring Process. (2015). Statement on the Fourth Monitoring Report of the Federal Government for 2014. Retrieved 11 13, 2020, from http://www.bmwi.de/Redaktion/EN/Downloads/stellungnahme-zum-vierten-monitoring-bericht-der-bundesregierung-fuer-das-berichtsjahr-2014-zusammenfassung-englisch.pdf?__blob=publicationFile&v=4

Federal Republic of Germany. (1949). Grundgesetz für die Bundesrepublik Deutschland. Retrieved 11 13, 2020, from http://www.gesetze-im-internet.de/gg/GG.pdf

Federal Republic of Germany. (1969). Bundeshaushaltsordnung. Retrieved 11 13, 2020, from https://www.gesetze-im-internet.de/bho/BHO.pdf

Federal Republic of Germany. (2020). Jahresbericht 2020 gemäß Artikel 24 Absatz 1 EED. *EED Annual Reports*.

Felder, F. A., & Athawale, R. (2018). "PACT-a-Mole": the case against using the Program Administrator Test for energy efficiency programs. *Energy Efficiency*, pp. 1-11.

Fraunhofer ISI et al. (2020). Methodikleitfaden für Evaluationen von Energieeffizienzmaßnahmen des BMWi. *commissioned by the German Federal Ministry of Economic Affairs and Energy*. Retrieved 11 13, 2020, from https://www.bmwi.de/Redaktion/DE/Downloads/M-O/methodik-leitfaden-fuer-evaluationen-von-energieeffizienzmassnamen.pdf

Fraunhofer ISI et. al. (2019). Endbericht der Evaluation des Energieeffizienzfonds. *BMWi*. Retrieved from https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/evaluierung-und-weiterentwicklung-des-energieeffizienzfonds.html

Hannigan, E., & Cook, J. (2015). Matching and VIA: Quasi-Experimental Methods in a World of Imperfect Data. *Proceedings of IEPEC 2015*. Retrieved 08 16, 2018, from https://www.iepec.org/wp-content/uploads/2018/02/2015paper_hannigan_cook-1.pdf

Johnson, K. (2014). Decision-Framework for Determining Net Savings Approach - Supplemental Document #2 to Principles and Guidance. *Northeast Energy Efficiency Partnerships*.

Northeast Energy Efficiency Partnerships. (2016). Gross Savings and Net Savings: Principles and Guidance. Retrieved 08 07, 2018, from http://www.neep.org/sites/default/files/FINAL%20GS%20and%20NS%20Principles%20and%20Guidance%20Document_2016May17.pdf

Olsthoorn, M., Schleich, J., Gassmann, X., & Faure, C. (2018). Free riding and rebates for residential energy efficiency upgrades: A multi-country contingent valuation experiment. *Energy Economics*.

Ringel, M. (2017). Energy efficiency policy governance in a multi-level administration structure — evidence from Germany. *Energy Efficiency*, pp. 753-776.

Rosenow, J., Cowart, R., Bayer, E., & Fabbri, M. (2017). Assessig the European Union's energy efficiency policy: WIll the winter package deliver on "Efficiency First". *Energy Research & Social Science*, pp. 72-79.

Schlomann, B., Hirzel, S., Nabitz, L., Heinrich, S., Jessing, D., Paar, A., . . . Antoni, O. (2017). From Targets to Impacts: Eight Steps for Evaluating Energy Efficiency Policies. *Proceedings of the 2017 IEPEC Conference*.

Umweltbundesamt. (2020). Spezifische Emissionsfaktoren für den deutschen Strommix. Retrieved 11 12, 2020, from https://www.umweltbundesamt.de/themen/luft/emissionen-von-luftschadstoffen/spezifische-emissionsfaktoren-fuer-den-deutschen

Violette, D. M., & Rathbun, P. (2017). Chapter 21: Estimating Net Savings - Common Practices. In NREL, *The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures.* Golden, Colorado.

Voswinkel, F. (2018). Accounting of energy savings in policy evaluation. How to get a least 8 different (correct!) results from the same data. (I. E. Conference, Ed.) *Proceedings of the IEPPEC 2018 Vienna Conference*. Retrieved 01 20, 2019, from http://www.ieppec.org/wp-content/uploads/2018/05/Voswinkel_paper_vienna.pdf

Voswinkel, F. (2019). Gotta catch 'em all - Catches to evaluating heterogeneous energy efficiency programmes. *Proceedings to the ECEEE Summer Study 2019*. Retrieved from https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2019/4-monitoring-and-evaluation-for-greater-impact/gotta-catch-em-all-catches-to-evaluating-heterogeneous-energy-efficiency-programmes/