Gotta catch ‘em all – catches to evaluating heterogeneous energy efficiency programmes

Fabian Voswinkel
Fraunhofer Institute for Systems and Innovation research
Breslauer Strasse 48
76131 Karlsruhe
Germany

Keywords

evaluation, evaluation methods, energy efficiency policy, energy policy, impact evaluation, policy mix, comparing evaluation

Abstract

Think about policies! When do you consider policies successful? Ideally, successful polices strive to influence something in an efficient and effective manner to reach a defined goal. However, when it comes to asking about the efficiency and effectiveness of policies, one challenge is well known: How to verify whether these conditions are met? The answer is good evaluation practice.

The German Energy Efficiency Fund, a special budget of the Federal Ministry of Economic Affairs and Energy (BMWi), finances more than twenty national programmes to support energy efficiency in private businesses, households and the public sector. The heterogeneity of the programmes ranging from energy savings check-ups for households to large-scale financial support programmes for energy efficient technologies in industry makes evaluation even more challenging. For evaluating the Fund in its entirety, a way had to be found to catch them all and present them in a harmonised way. Therefore, an evaluation system, which encompasses a definition of indicators, savings metrics, effects and additional assumptions was developed and applied.

This paper provides an introduction to the evaluation methodology. It focuses on practical catches for adequately covering the wide range of programmes. A major catch that lies within a uniform methodology for heterogeneous programmes is the interpretation of results. While using indicators to compare evaluation results across different programmes sounds appealingly simple, direct conclusions are often misleading. The success of a programme is very individual and an equal value of e.g. savings per Euro of funding does not necessarily mean that two programmes are equally successful. Detailed examples for such catches from the up-do-date evaluation of the Energy Efficiency Fund are presented in this paper and suggestions are made for avoiding premature conclusions from multi-programme evaluations.

Introduction

The Energy Efficiency Fund (EEF) is a broad-range funding scheme by the German Federal Ministry of Economic Affairs and Energy (BMWi). It has first launched in 2011 and has since then grown to over half a billion Euros in 2018 (Voswinkel, 2018). Its composition changed slightly over the course of the evaluation project. In the last evaluated year, 2017, the EEF consisted of a set of 18 policies of different types ranging from individual information and consulting campaigns to broad-range financial support. The list is shown in Table 1.

The political ambition of the EEF is to play a crucial role in achieving the emissions savings goals laid out by the German government after the ratification of the Paris agreement (BMWi and BMU, 2010). To monitor the outcomes of the EEF, receive information about improvements and to satisfy reporting obligations towards the Federal Court of Auditors and the European Energy Efficiency Directive (EED) (European Union, 2012) (European Commission, 2016), the ministry has commissioned an extensive evaluation over the course of three evaluation periods between 2015 and 2017. In preparation for the evaluation work, a methodology was developed specifically for the evaluation of the EEF. The detailed methodology document henceforth served as a reference for the evaluation of the different programmes. Over the course of the three evaluation periods, several aspects of the methodology were updated and retroactively applied to the evaluation. Because it was the first attempt to define a methodology of such detail for a large-scale policy mix in Germany, the elaboration and putting into practice of the methodology was accompanied by an intense learning process. This paper aims to share some of the experiences with a unified methodology in energy efficiency policy evaluation. While the methodology has by and large proven well suited for a methodologically sound and comparable evaluation, the evaluators encountered a series of catches that shall be presented in this paper. The catches should not be considered shortcomings of the methodology, as shortcomings can be overcome by continuous improvements. The catches mainly concern the limitations of methodological harmonisation in practice. The question that this paper wants to take part in answering is, how far can harmonisation go and how high should ambitions for comparability and aggregation of results go?

The paper first introduces the crucial aspects of the evaluation methodology developed for the EEF. In the main part, it goes on to present nine catches to applying the unified methodology in practice. Each of the catches is presented first theoretically, underlined by an example from the evaluation of the EEF and is finally briefly discussed. The paper closes with a short conclusion.

The evaluation methodology

The EEF is a public funding instrument subject to certain obligations based on the obligation laid down in the German constitution, which requires the Federal Court of Auditors to evaluate government spending. Hence, evaluation of the cost-effectiveness of the EEF and implicitly of target achievement is mandatory (Art. 114 II 2 GG Federal Republic of Germany). It is common practice for the Federal Court of Auditors to direct evaluation to the responsible governmental body which then further directs it to independent external experts in the required field (Dittrich, 2017). General principles state that evaluation should be "comprehensible, accepted, relevant, representative and, as far as possible, measurable" (Schlomann, et al., 2017). To be able to satisfy these needs, it is necessary to apply a unified methodology that enables policy makers to compare results for the manifold and heterogeneous programmes part of the EEF. This section gives an introduction to the methodology that was developed for the evaluation of the EEF.

The evaluation of each programme follows a series of steps based on the eight-step approach presented below. It was first elaborated in Schlomann, et al. (2017) and (Eichhammer, Boonekamp, Labanca, Schlomann, & Thomas, 2008). For the comparability of a policy mix, the approach has been extended with a ninth step concerning comparability and combination. The catches in the main part of this paper follow the steps of the evaluation system in their argumentation.

IMPACT EVALUATION SYSTEM

In **Step 1**, the general characteristics of the policy measure are identified. That includes the available funding, the administrative framework, the funding party, the target group, administrative issues and supported technologies and methods of the policy. **Step 2** concerns framework conditions including underlying assumptions such as energy prices, primary energy factors for different energy sources and greenhouse gas emission factors.

Table 1. List of policies in the Energy Efficiency Fund in 2017 with indication of policy type.

Original name	English translation	Policy type
Abwärmerichtlinie	Exhaust heat guideline	financial: broad-range
Effizienzhaus Plus mit Elektromobilität	Efficiency house plus with electric mobility	financial: individual
Nationales Effizienzlabel für Heizungsaltanlagen	National efficiency label for old heating systems	informative: broad-range
Energie- und Stromsparchecks	Energy and power savings checks	informative: individual
Energieeffizienzgenossenschaften	Energy efficiency cooperatives	informative: individual
Energieeinspar-Contracting	Energy saving contracting	informative: individual
Energieeinsparzähler	Energy savings counter	financial: individual
Energiemanagementsysteme	Energy management systems	financial: broad-range
EnEff.Gebäude.2050	Energy Efficiency buildings 2050	financial: broad-range
Leuchttürme Abwärme	Flagship project exhaust heat	informative: individual
Unterstützung der Marktüberwachung	Support for market monitoring	regulatory
Mittelstandsinitiative	Medium-enterprise initiative	informative: individual
Energieberatung für Kommunen und gemeinnützige Organisationen	Energy consultancy for municiplaties and charitable organisations	informative: broad-range
Paket BMUB	Environmental Ministry package	financial: individual
PKW-Label	Passenger car label	informative: broad-range
Produktionsprozesse	Production processes	financial: individual
Querschnittstechnologien	Cross-cutting technologies	financial: broad-range
Nationale Top-Runner-Initiative	National top-runner-initiative	informative: broad-range

Table 2. Select targets and indicators. Source: Schlomann, et al. (2017), own representation.

Policy Target	Indicator
Target achievement	
Contribution to the achievement of climate protection targets	Greenhouse gas reduction (t CO_2-eq.)
Exploitation of energy savings potentials	Reduction of final and primary energy consumption (MWh_{final} and $MWh_{primary}$)
Reduction of energy costs	Achieved energy cost saving (EUR)
Impact assessment	
Net values for quantitative indicators after adjustment for effects (e.g. free-rider, spill-over)	
Economic efficiency	
Funding efficiency (view: funding body)	GHG - funding efficiency (t CO_2-eq./EUR of funding) Energy-funding efficiency (MWh_{final}/EUR of funding)
Policy administration	
overarching	Satisfaction with policy administration by participants Satisfaction with policy administration by administrators

In **Step 3**, policy targets are reviewed. The targets form the basis for the definition of indicators in the next step. **Step 4** translates policy targets into indicators and the respective indicator targets. Target and indicator definitions are described in more detail in the following subchapter. **Step 5** collects the data necessary for the calculation of indicators. For the evaluation of a policy set, it is important to organise data collection in a similar way. Differing data availability has to be taken into account. In **step 6,** the data is analysed for gross values. In this step, methods have to be defined in a very detailed way to reduce hindrances to comparability. Considerations for accounting methods are detailed below. **Step 7** concerns the effect adjustments for effects like the free-rider effect to obtain net values as part of the impact assessment. A more in-depth view into effect adjustments for free-rider and spill-over effects is given below. **Step 8** treats future projections from the calculated data. This can be useful for overarching indicators to which the policy measure's target achievement contributes. (Schlomann, et al., 2017). Finally, **step 9** concerns the combination and summation and comparison of separate policies to an overarching evaluation project. A more detailed account is given at the end of the methodology section.

TARGETS AND INDICATORS

In order to define indicators for a policy measure or policy mix, goals and targets have to be clearly defined. Often, they are implicitly given by government strategies or explicitly stated in legal policy guidelines. For the EEF the targets have been defined by the Federal Ministry of Economic Affairs and Energy as follows (Fraunhofer ISI et. al, 2019, forthcoming):

- Contribution to the development of a highly energy-efficient economy;
- Contribution to the achievement of climate protection targets;
- Exploitation of the existing economic energy saving potentials;
- Exploitation of the existing economic electricity saving potentials;
- Decreasing the energy costs of all energy consumers (private households, companies, public institutions).

Out of these targets, indicators were formed according to four main aspects of evaluation. (Fraunhofer ISI et. al, 2019, forthcoming)

- **Target achievement (A)** monitors whether the previously defined targets are met. Gross values are calculated;
- **Impact assessment (B)** analyses whether the policy is causal for the outcomes, e.g. emissions savings. Net values are calculated;
- **Economic efficiency (C)** relates the outcomes to the monetary inputs, e.g. financial support;
- **Policy administration (D)** evaluates the satisfaction with the policy design from both the participant's side and the programme administrator's side.

Table 2 presents example indicators for the given categories.

DATA COLLECTION

In parts, evaluation can utilise detailed data from the programme administration. Such data may contain engineering estimates of energy savings, energy consumption before and after the action, the sector and size of the enterprise, employed technologies and more. However, in other cases, a full sample including that level of detail may not be available. Then, estimates can be made using primary, secondary or tertiary data from surveys or other sources. Primary data have been generated specifically for the programme, secondary data are existing data that can be utilised for programme evaluation and tertiary data are only available in aggregated or compressed form.

If a full sample is not available, either a random drawing or a selected sample based on participant characteristics is used. This way the sample can be representative.

Additional and complementary data are generated using surveys. Such data are used, among others, for the determination of adjustment effects and the indicators of category D (Policy administration).

SAVINGS ACCOUNTING

In step 6, the data analysis, for the definition of indicators in categories A, B and C, energy and emissions savings have to be calculated in a uniform way. The disclosure of the employed metric is of crucial importance for a transparent evaluation. Because different audiences of the evaluation and reporting standards require different metrics, the EEF employs four metrics in parallel, which are based on Schlomann, Rohde, & Plötz (2015) and (Voswinkel, Grahl, & Rohde, 2018). They will be presented in the following with year *i* being the year of evaluation:

- **First year savings** are savings in the year *i* from actions implemented in the same year.
- **Cumulated annual savings** are the sum of savings in year *i* from actions implemented in all years up to year *i*, whose lifetime has not yet expired.
- **Periodically cumulated savings** are the sum of savings in all evaluated years up to year *i* from actions implemented in all years up to year *i* whose lifetime has not yet expired.
- **Lifetime savings** are the sum of savings over the course of actions' individual lifetimes from actions implemented in all years up to year *i*.

A crucial issue for the lifetime savings is the determination of lifetimes. While in some programmes, detailed energy efficiency concepts give an account of estimated technical lifetimes on a micro-level for the individual technological components, for some programmes, such micro-data are not available. In this case, default values for sectors and technology categories based on previous evaluations and literature are used.

EFFECT ADJUSTMENT

In step 7, gross results are adjusted for a number of effects to determine the net impact of the policy on the outcomes. The effects are listed in Table 3.

In the EEF, the first three effects are calculated as part of each policy. The follow-on effect refers to the assignment of savings to a certain point in time as detailed in Catch 4 below. Double-counting effects are addressed in step 9 in the combined evaluation of the EEF. The method is detailed in the last part of the methodology section.

The free-rider effect describes the fraction of the total impact that would have been implemented in the absence of the policy. The effect is subtracted from the gross values (Breitschopf, Voswinkel, & Schlomann, 2018).

Olsthoorn et al. (2018) and (Violette & Rathbun, 2014) categorise free-rider effects in three groups. Strong free-riders, weak free-riders and deferred free-riders. Strong free-riders had already planned a savings action and additionally decided to benefit from the financial incentive. Weak free-riders had not originally planned a savings action but the information about a fitting opportunity from the incentive programme gave them the idea. After receiveing the information, they would have invested also without the financial incentive. Hence, they are free-riders for the financial aspect of the proramme but no free-riders for the information aspect. Deferred free-riders had planned the investment for a later time, but pulled the investment forward due to the financial benefit. Hence, for savings that occur after the time of the initially planned time of investment, deferred free-riders are counted as free-riders. For the timeframe that the policy pulled the investment forward, they are no free-riders. Deferred free-riders can aditionally be strong and weak (Voswinkel, 2018). The value for the free-rider effect is subtracted from the gross values.

The spill-over effect is the flipside. It describes how many additional savings have been generated both inside and outside the participant company that are not directly credited to the programme. (Voswinkel, 2018). Word of mouth effects and the generation of general knowledge about the programme or energy efficiency investments are part of the spill-over effect. The value for the spill-over effect is added to the gross value.

The EEF uses a survey method to determine the mentioned effects. For the free-rider effect, the base question is directly "Would you have also implemented the measure without financial incentives?" with the options "1: no", "2: yes, on the same scale", "3: yes, on the same scale, but later" "4: yes, but on a lower scale", "5: yes, but on a lower scale and later", "6: n/a". Answer 1 is associated with no free-rider effect. Answer 2 with 100 % free-rider effect. Answers 3 to 5 have to be looked at in combination with other questions. They include questions about the degree of decreased investment without the programme and with the timeframe at which the investment was originally planned, to account for deferred and partial free-riders or a combination of both. Finally, a question was asked to determine the importance of the information content of the programme. This serves to discount the free-rider effect for the weak free-riders.

The spill-over effect is determined by a set of questions concerning the priority that energy efficiency has inside the company, how it is treated in their internal and external communication and whether they are recommending the support programme in their professional environment (Voswinkel, 2018).

The determined values for the two effects are then added up to form the net value as shown in Figure 1. The percentages are a fictional example for illustration purposes.

COMBINED EVALUATION

In the evaluation of a policy mix, apart from analysing the single policies, policymakers are interested in results about the policy mix as a whole. Results from the different individual programmes therefore have to be combined. For this end, the EEF uses a selection of methods depending on programme specifics (Fraunhofer ISI et. al, 2019, forthcoming).

- **Aggregation of quantitative indicators.** For indicators concerning savings (category A), values can be summed up for cumulated annual savings and periodically cumulated savings. The summation of lifetime savings are not added up due to differences in lifetimes between programmes, which would not be visible anymore in aggregate values leading to possible unwanted interpretations. Catch 5 details related issues. Relative values like the indicators from category C (e.g. savings per money spent) can be combined by first adding up numerator and denominator and dividing them. However, due to the widely differing programme characteristics, the resulting value would have little meaning as detailed in Catch 6. Hence, in the EEF it was decided not to aggregate the indicators from category C. However, the individual values are jointly presented and compared while considering programme differences.

Table 3. Adjustment effects for net values in the EEF. Source: Schlomann, et al. (2017), own representation.

Impact / Effects		Description
	Gross value	Impact before considering effects
-	Free-rider effects	Saving that would have occurred without policy
+	Spill-over effects	Effects on third parties and other areas not directly credited to the programme
+	Follow-on effects	Effects through not yet completely realized actions
-	Double-counting	Effects through interactions of measures
=	**Net value**	Impact after adjusting for effects

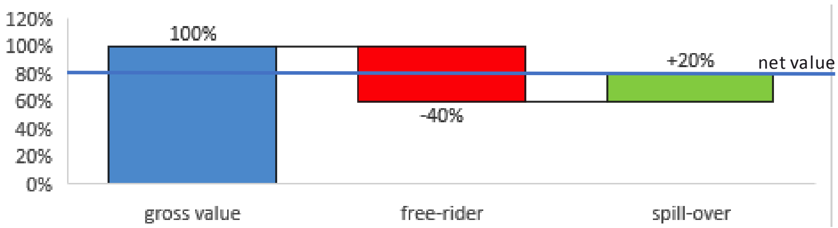

Figure 1. Net savings value calculation.

- **Argumentative account for qualitative indicators.** The aggregation of qualitative indicators is trickier. One possibility is to mirror results between programmes considering differences in programme specifics and finding common directions in results. However, due to the great heterogeneity of policies, this method is mostly difficult to implement in practice. Hence, the EEF evaluation employs this method only for the most important indicators concerning satisfaction with the programme administration. Surveys and interviews for this indicator have been performed in a unified way opening the possibility to assign a numeric value to the answers. These numeric values scaled between "not at all satisfied" to "highly satisfied" are jointly presented and compared. Answers to more specific questions and comments are compared and parallels between programmes are identified.

ACCOUNTING FOR INTERACTION

Finally, for quantitative indicators, the evaluation employs estimated coefficients for interaction effects between programmes that are part of the EEF. The estimation of these effects is based on survey results asking whether any and which of the other programmes in the EEF are known or have been used by the participant. Furthermore, reference values from the MURE database (Odyssee-Mure, 2018) and other literature as well as plausibility considerations from programme characteristics are employed.

The catches

This part of the paper presents a selection of difficulties in applying a unified methodology as described in the previous section. While certain aspects to evaluation can be methodologically harmonised, other aspects are met with practical difficulties that render a complete comparison impossible. There is a row of catches to the appealing methodology that fits all.

Most of these catches are due to fundamental differences in policy designs. They differ in properties ranging from policy type (e.g. informative or financial) to addressed sector or intensity of intervention. Table 4, however by definition incomplete, lists possible policy aspects that should be considered when designing an evaluation system.

The following part presents catches experienced during the application of the unified methodology of the EEF. They are organised according to the nine steps of the evaluation system described above. Each catch will be presented with the underlying aspects leading to it, a concrete example from the evaluation and a short discussion.

CATCH 1

Step 1 of the evaluation process concerns the characterisation of the policy measure including aspects like budget, target group, administrative issues, supported technologies, methods or regulatory framework. The catch is that these aspects can change over time or be incompatible with each other. Especially with multiple reporting requirements on individual, national or supranational levels, a unified methodology for evaluating a policy or policy set may not suit all at the same time due to incompatibilities in the different reporting standards. In other words, in a complex field like energy efficiency policy evaluation, there is a large number of right answers that may not match the other large number of right requirements.

In the EEF, as in most other European energy efficiency policies, results have to be reported as part of the European Energy Efficiency Directive (EED) Article 7. Methodological requirements are laid out in detail in Annex V of the EED (European

Table 4. Differing aspects of policies with examples.

Aspect	Example
Sector	Industry, households
Policy size	Assigned budget
Intensity of intervention	Individual consultancy, broad range support
Policy type	Financial support, informative programme
Data availability	Individual energy consumption data, number of cases, baseline data
Policy runtime	Timeframe of action
State of policy development	Pilot, newly established, established
Time lag for energy savings	Lag between the time of implementation and the first measured savings
Technical differences	Different technical lifetimes
Access to participants	Possibility to survey or interview participants
Geographic distribution	Geographic distribution of participants

Union, 2012). The methodology of the EEF, however was based on the requirements of the German National Energy Efficiency Action Plan (NAPE) (BMWi and BMU, 2010). The different savings accounting methods presented in the methodology section account for the fundamental differences in the requirements. However, a large array of details including the savings timeframe of the EED from 2014–2020 (European Commission, 2016) and the EEF evaluation periods of 2011–2015, 2011–2016 and 2011–2017 do not coincide. In the process of developing a methodology, it is therefore advisable to get informed very well about all relevant legal frameworks. This way, data collection, analysis, and other methodological aspects can take differences in these frameworks into account and possibly disclose results in different ways in parallel.

CATCH 2

This catch is both curse and blessing. The learning effect concerns all steps of evaluation from 1 to 9. It is uncontested that evaluation and methodology are highly complex and that a perfect evaluation in the strict sense of the word does not exist because certain simplifications have to be done to make evaluation viable. Learning occurs on all sides, the evaluators, the policy makers, the policy administrators and other stakeholders. Methodologies have to be developed in the beginning and applied afterwards. Over the course of an evaluation period, certain aspects of a methodology may prove to have potential for improvement. Stakeholders may face changing priorities in the required information to be taken from evaluation during the process.

In the EEF, particularly the treatment of adjustment effects like the free-rider effect has advanced strongly with learning processes. The introduction of the weak free-rider effect that takes the information aspect of a financial support programme into account was introduced at a later stage and applied retroactively. At the same time, methodological requirements for calculations were put more into focus by the ministry due to increasing experience from different evaluation studies in the field. To avoid pitfalls, a detailed methodological document has been agreed on in the beginning of the project and consequently applied. Only slight changes were made in the course of the evaluation project. This way all parties have the security from a transparent evaluation process, potentially at the cost of possible improvements. However, learning effects can improve evaluation studies over the course of time.

CATCH 3

Step 5 of the evaluation process concerns data collection. Evaluation is often dependent on data that is generated during the implementation or evaluation of the policy. To establish a methodology that avoids as many pitfalls as possible and leads to the soundest results, evaluators always have to consider the limitations in available or obtainable data.

The generation of higher quality data may come at a higher price leading to the issue of justifiability of the costs. Furthermore, certain data may not be possible to generate in a reliable way. In the context of energy efficiency policy evaluation, a crucial aspect is the calculation of energy savings. Arguably, the most reliable method is to use metered savings (Violette & Rathbun, 2017) (Hannigan & Cook, 2015). It requires physical measurements of energy consumption of the analysed technological system before and after the implementation of a savings action and furthermore control for outside variables like the weather and inside variables like production volume. However, in a large-scale programme with several thousand participants every year, performing physical measurements is very time-consuming and cost-intensive.

In the EEF, physical measurements after the implementation are not performed in large-scale programmes. Performing such would have a number of considerable drawbacks. The main one is the additional cost for the procedure. Another one is the necessary time lag between the programme participation and the evaluation time. Measurable savings results can often only be obtained years after the programme participation.

As an example, an information measure like the energy consultancy for small and medium enterprises can incentivise a company to participate in a financial support programme for waste heat reduction in the next year. The necessary technological upgrade is very complex and takes two years to implement. Hence, measurable results first become available three years after the first programme participation. Deemed savings based on engineering estimates on an individual participant level can be evaluated soon after the time of participation. First evaluation results can be available quickly and give policymakers the chance to react better if there is a need for reform. A cost-benefit consideration based on these aspects has to be performed. On the other hand, metered savings can also help to generate better estimates for effects like the free-rider effect. For this, a comparison group in a quasi-experimental design is necessary. The analysis can measure changes in energy consumption for statistical twins of the participant group and conclude what part of the savings would have happened without the policy intervention. While physical measurements are generally costly, but possible, a valid comparison group may not exist. If the groups are too different in their characteristics, results will end up unreliable.

Availability of data and consequently methods that can be employed make a comparison or even summation of results in a policy set difficult and possibly biased. Such distortions must be taken into account when interpretations and conclusions are made.

CATCH 4

Step 6 of the evaluation process concerns the data analysis including the type of employed methods for calculations. In this step, a series of methodological details can have a large impact on final results. The mentioned issues in this catch are time-related. One of the main difficulties concerns allocation of savings to timeframes.

Programme participation and implementation of savings actions usually consist of more than one step occurring at different points in time. Evaluation also has to be performed at a defined point in time and for a defined period of time. This raises the issue of time allocation. It has to be decided which step during the process of the implementation should be used as benchmark. It can be the moment of application for participation, the moment of acceptance into the programme or the moment of finalisation of the savings action. Depending on programme design, other moments can be possible. For programmes with similar characteristics, such a decision can be made, and the results can be compared. However, oftentimes programmes vary considerably and the same choice for point in time cannot be made for different ones. The definition of each of these points in time has its advantages and disadvantages.

In the EEF, usually the time of acceptance for participation was used. At that point, in most cases, the deemed savings have been calculated and verified and the actual physical implementation can be largely expected while the evaluation can occur already at an early enough stage for timely data reporting on the German and EU levels as well as for policy adjustments. On the downside, in the process of implementation, changes may occur that affect the amount of energy savings. For example, a company may only implement a part of the project or do something differently due to a technical difficulty encountered during the process. A participant could also go out of business or cancel participation altogether. The results are hence possibly not entirely correct. However, because implementation in some programmes can take up to two years and possible additional legal action in case of disagreement, the trade-off for an earlier and reliable timeframe over the exactness of results has been made. The difficulty for comparison though, remains. While in the waste heat or the cross-cutting technologies programme, the deemed savings are present at the time of acceptance, in the technology-open support programme for climate friendly production processes includes a strong focus on individual consultancy. The exact savings actions are only defined during the participation in the programme. Hence, the timeframe has to be chosen differently causing problems in direct comparability that cannot be solved by a unified methodology. The same applies to informative programmes or preparatory programmes for savings like the energy management systems support programme. In these cases, savings do not result directly from the programme. This type of programme rather prepares participants for later investments and energy savings that may or may not be financially supported by other programmes. Effects therefore may occur at any given time in the future that cannot be immediately assigned to the policy. One way is to gather data on the time of the actual savings action and assign the savings to the year of programme participation for gross values. Another possibility is to treat such savings as a delayed effect as part of the effect adjustment leading from gross values to net values. Savings in the future after programme participation do not accrue directly to the programme but, similar to spill-over effects, take part in the impact assessment of the programme, the net values.

Finally, one of the savings metrics presented in the methodology section are lifetime savings. The information on the amount of savings that a policy will have led to until the lifetime of the given savings action has passed is of great interest. However, the value depends strongly on the assumed lifetime. Lifetimes can be calculated on a micro-level per participant if data can be accessed in such detail. Oftentimes though, the level of detail is lower, so averages for technologies or for sectors have to be used. While technical lifetimes or financial depreciation can often not be defined with absolute certainty, different ways of defining lifetimes due to data availability leads to limited possibilities for comparison of lifetime savings and also savings projections to a defined year like 2020 in the EED (European Union, 2012).

CATCH 5

While catch 4 included methodological difficulties related to time, catch 5 relates to a methodological difficulty that is strongly tied to the programme administration and the definition of gross values. Programmes display differences in how savings are calculated. One of the practical issues is the definition of an appropriate baseline. The energy consumption after implementation of the savings action has to be compared to a baseline (Broc, Adnot, Bourges, Thomas, & Vreuls, 2009). Possibilities for the baseline include the before-consumption, a minimum standard technology defined in regulatory action, a market average technology or variations of these methods considering market trends. It is apparent that the choice of baseline has a strong influence on the resulting gross savings. While a methodology can define a unified way to establish a baseline, in practice it is often, depending on programme specifics, difficult to acquire the necessary data.

In the EEF, one such example is the support programme for cross-cutting technologies. One of the two support lines in this programme specifically targets small investments in single technologies with minimal administrative burden to avoid barriers to participation particularly for smaller companies. Energy savings have to be stated in the application form. However, numbers can be taken from the manufacturer's technology data sheet and the calculations are done by the applicants themselves who may not have specific knowledge of the subject matter. For simplicity, most will calculate the difference between their current consumption and the consumption after technology implementation. While the before and after calculation is a possible baseline that can be defined, it may not satisfy the demands of reporting requirements. Under the EED and EU state aid rules, financial support action can in general only be justified for savings that are additional to existing regulatory action (European Union, 2012). While the technology before the savings action may be older and therefore not meeting the newer minimum standard regulation, the savings in the before and after calculation will include savings due to regulation plus savings due to the support programme. But a change towards calculating a baseline would increase the administrative burden considerably and weaken the main purpose of the programme.

CATCH 6

Step 9 finally considers difficulties in comparing indicator values between different programmes and their interpretation using the unified methodology. Indicators like savings amounts or funding efficiency suggest a direct comparability between programmes. In a sense, that is true. However, the catch lies in the interpretation. Differences mainly concern programme specifics and targets. The conclusion, that the programme with higher funding efficiency or higher savings is more successful, is not necessarily true. Functioning energy efficiency policy requires actions that cover the whole economy. A programme with very high funding efficiency might have required intense consulting processes that would not be possible in a large-scale programme. Otherwise, it may specifically target low-hanging fruits. This way the specific programme cannot be compared to a broad-range programme aiming at implementing single technological measures for energy efficiency. The same applies to values on different indicators. Does the high funding efficiency leverage the overall low savings?

In the EEF, several examples could be found. One being the comparison between the support programme for production processes and the cross-cutting technologies. While the former aims at exploiting advanced energy efficiency potentials using individualised and intense consulting action for a select number of participants from mainly large enterprises, the latter focuses on more easily implemented technologies in a large number of small and medium enterprises. The two funding efficiency values and overall savings differ greatly. While the overall savings are larger in the cross-cutting technologies, the funding efficiency is considerably larger in the production processes programme. A clear argument for the one or against the other can therefore not be made. Both take part in a policy mix addressing a wider range of the economy.

CATCH 7

As with most other matters, performance of energy efficiency programmes is subject to level of awareness about the programme and experience in administration. Policies are created or are phased out constantly making the comparison between programmes once again more difficult. In the first year, word will not have yet spread very far, which results in lower participation. Administrative processes may take longer with a lack of experience from both sides, the administrators and the applicants. Hence, a long-running programme cannot be simply compared to a new one. This issue applies strongly to the EEF whose composition of policies slightly changed during the evaluation period. It contains long-running programmes, pilot projects and newly established programmes.

CATCH 8

While catch 5 raised the issue of the baseline applying to the comparability of gross values, catch 8 discusses the comparability of effect adjustments for net values. As described in the methodology section, there is an array of possible effects that can alter the fraction of savings that accrue to the policy. The net values are therefore dependent on the way effect adjustments are calculated. In catch 3, the group comparison approach for the calculation of free-rider effect is presented in the context of difficulties to find a necessary comparison group or receiving the means to make physical measurements. Due to these difficulties, another approach for the calculation of free-rider and other effects is the survey method as presented in the methodology section. In this, participants and possibly a non-treatment comparison group is questioned about their behaviour towards energy efficiency. This way, effects can be determined, but results may be subject to biases. A further method for the determination of free-rider effects is based on former studies or expert estimations (Johnson, 2014) (Northeast Energy Efficiency Partnerships, 2016). The difference in sophistication, data requirements and consequently quality of results between these methods pictures that comparisons of net values add another layer of contingencies.

In the EEF, most policies for which net effects were calculated used a unified survey approach. However, reaching out to participants was not possible in all cases. Hence, a certain comparability is given. The absolute values, however, should be taken with care as the survey approach is subject to biases. In harmonised methodologies, oftentimes trade-offs have to be made between quality of results and comparability. If measured energy consumption data and a comparison group is available for one programme, but not for another, a trade-off has to be made between the highest accuracy and the highest comparability.

CATCH 9

Finally, in the evaluation of a policy mix or in aggregations as part of reporting initiatives, interactions between different programmes can exist. To some extent, these interaction effects can be evaluated and calculated. Consulting or information programmes that aim to inform about participation in financial support programmes have calculable interaction effects with these support programmes. However, on a larger scale like the EU in the EED reporting or even the national level, the interactions are as complex as the economic system. Reliable data on the amount of interactions cannot be calculated. It may be possible to compare evaluation data with input-output tables for energy resources and deduct an estimate for the interactions using the deviations. Further research can take a crucial part in establishing such estimates.

Conclusion

The paper presents experiences from the practical application of the unified methodology for the evaluation of the EEF, a policy mix of largely heterogeneous policies in Germany. The methodology was developed specifically for the evaluation of the EEF and was part of an intense learning process. It has been improved in certain aspects over the course of three evaluation periods. The paper aims to give insights into the practical limitations to harmonisation. How far is harmonisation of the methodology possible and how far does it make sense? The methodology of the EEF and the nine-step system proved to work well for the purpose of evaluating the policy mix. Aspects that could be treated in a unified way were mainly defined beforehand. Improvements over the years have made the evaluation methodology better, but the complexity of evaluation is infinite. Work towards more reliable methods of data collection and analysis and effect adjustments is underway. However, this paper argues to question the level of ambition in harmonisation efforts. Harmonisation efforts can be complemented by more qualitative approaches to comparability.

Harmonisation is very important in a setting of a complex energy efficiency policy landscape with manifold actors and evaluators. Results have to be comparable as far as the practical limitations permit it. Furthermore, harmonisation is also a crucial step for transparency of methods. Only with disclosure of all necessary information about the data and employed methods, can results be interpreted the right way and put into perspective towards other policies. This paper presented a series of catches that reduce the direct comparability of results. These limitations should be considered when creating and applying a methodology. Moreover, they should especially be considered in the interpretation of results and the culmination in policy action.

The conclusion of this paper is that the ambitions of harmonisation should be questioned. Not only how far is it possible to go, but how far do we want to go? Policies are different and they need to be in a complex economic system that heavily determines the worldwide challenges of our time. More unification in evaluation might mean more unification in policy approaches. This way data could be generated in a more unified way and lead to more comparable results. However, while to a certain degree, such a step is necessary, the innovativeness of policies and their specific qualities should not be left out of sight. Some catches can be reduced by thorough evaluation planning and a programme administration that considers all necessary steps of evaluation already during the creation of a policy. But some catches persist and will not seize to exist. They are proof of the wide scope of policy action aiming at reaching everybody to achieve the overarching goal – reducing and ending climate change.

References

Baumgartner, R. (2016). Chapter 12: Survey Design and Implementation for Estimating Gross Savings Cross-Cutting Protocol. *The Uniform Methods Projects: Methods for Determining Energy-Efficiency Savings for Specific Measures*. Retrieved 08 07, 2018, from https://www.nrel.gov/docs/fy17osti/68568.pdf.

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.

Breitschopf, B., Voswinkel, F., & Schlomann, B. (2018). How relevant are free-rider effects for taget achievement? *Proceedings of the IEPPEC 2018 Conference*. Retrieved 08 07, 2018, from http://www.ieppec.org/wp-content/uploads/2018/06/Breitschopf_paper_vienna.pdf.

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.).

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. (2016). Directive of the Eurpean Parliament and of the Council amending Directive 2012/27/EU on energy efficiency. proposed EED 2016. *Official Journal of the European Union*. Retrieved 08 07, 2018, from http://eur-lex.europa.eu/resource.html?uri=cellar:efad95f3-b7f5-11e6-9e3c-01aa75ed71a1.0009.02/DOC_1&format=PDF.

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

Federal Republic of Germany; Grundgesetz; Art. 114 II 2 GG.

Fraunhofer ISI et. al. (2019, forthcoming). Endbericht der Evaluation des Energieeffizienzfonds. *BMWi*.

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*.

Labanca, N., & Bertoldi, P. (2016). Energy Savings Calculation Methods under Article 7 of the Energy Efficiency Directive. *JRC Science for Policy Report*. Retrieved 08 07, 2018, from https://e3p.jrc.ec.europa.eu/publications/energy-savings-calculation-methods-under-article-7-energy-efficiency-directive.

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.

Odyssee-Mure. (2018). Mure database (Mesures d'Utilisation Rationelle de l'Energie). Retrieved 01 20, 2019, from http://www.measures-odyssee-mure.eu/.

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*.

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*.

Schlomann, B., Rohde, C., & Plötz, P. (2015). Dimensions of energy efficiency in a political context. *Energy Efficiency*.

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., Grahl, A., & Rohde, C. (2018). Less hot air for a less hot climate - Evaluating the German waste heat reduction programme. (eceee, Ed.) *Proceedings of the eceee Industry 2018 Berlin Conference*. Retrieved 01 20, 2019, from https://www.eceee.org/library/conference_proceedings/eceee_Industrial_Summer_Study/2018/1-policies-and-programmes-to-drive-transformation/less-hot-air-for-a-less-hot-climate-evaluating-the-german-waste-heat-reduction-programme/2018/1-035-18_Voswinkel.pdf/.

Acknowledgements

This paper is based on a research project funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). In this project, the results and experience of the evaluation of the Energy Efficiency Fund has been elaborated by the consortium of Fraunhofer Institute for Systems and Innovation Research (ISI), Prognos AG, ifeu and Stiftung Umweltenergierecht. An extension project will start in 2019 aiming to generalise the harmonised methodology for energy efficiency policy evaluation in Germany. The guideline document is planned for publication.

I would like to thank the representatives at the ministry and all project partners for the support of this study and for the important discussions during the period of project work.

The considerations have furthermore partly been included in a case study about the German Energy Efficiency Fund as part of the project EPATEE, which analyses evaluation studies worldwide. Results can be found on www.epatee.eu. EPATEE has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement No. 746265.