

Heinemann, Friedrich et al.

Research Report

Enhancing objectivity and decision relevance: A better framework for evaluating cohesion policies

EconPol Policy Report, No. 50

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Heinemann, Friedrich et al. (2024) : Enhancing objectivity and decision relevance: A better framework for evaluating cohesion policies, EconPol Policy Report, No. 50, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/325303>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Enhancing Objectivity and Decision Relevance: A Better Framework for Evaluating Cohesion Policies

*Friedrich Heinemann, Zareh Asatryan, Julia Bachtrögler-Unger,
Carlo Birkholz, Francesco Corti, Maximilian von Ehrlich, Ugo Fratesi,
Clemens Fuest, Valentin Lang, Martin Weber*

Report for the German Federal Ministry of Finance (BMF)

This report was prepared as the outcome of a research cooperation between the ZEW – Leibniz Centre for European Economic Research in Mannheim (coordinator), ifo Institute - Leibniz Institute for Economic Research (ifo), the Austrian Institute of Economic Research (WIFO), the University of Mannheim, the Center for European Policy Studies (CEPS), the University of Bern, the Politecnico di Milano and the University of Heidelberg on behalf of the German Federal Ministry of Finance (BMF). The views expressed in this report are those of the authors and do not necessarily represent the views of the BMF or the minister.



EconPol Europe is CESifo's economic policy platform. With key support from the ifo Institute, it seeks to leverage CESifo's globe-spanning network of more than 2000 high-ranked economists – at least a dozen of whom have won the Nobel Prize – and ifo's decades-deep research expertise to provide well-founded advice to European policymakers and to facilitate informed decisions. Drawing on the wide range of specializations of its members, EconPol's mission is to contribute to the crafting of evidence-based, effective economic policy in the face of the rapidly evolving challenges faced by the European economies and their global partners.



EconPol POLICY REPORT
A publication of the CESifo Research Network

Publisher and distributor: CESifo GmbH
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 89 9224-0, Email office@cesifo.de
Editors: Clemens Fuest, Florian Dorn
Reproduction permitted only if source is stated and copy is sent to CESifo.

EconPol Europe: www.econpol.eu

Abstract

By international comparison as well as compared to other EU policies, the EU's Cohesion Policy (CP) evaluation system is far developed and institutionalized. This paper analyses the remaining gaps and shortcomings in the CP evaluation system against principles established by the OECD and others and provides recommendations on how to further improve it. The presence of a broad and imprecise CP objective function emerges as a key challenge for evaluations. The evaluation culture is not equally developed among all Member States and regions. In quite some cases, an unfavorable equilibrium is found which is characterized by limited evaluation capacities, poor methods, and a formalistic approach to evaluations. Program evaluations in the Member States are usually commissioned by national or regional managing authorities who have a vested interest in promoting the success of their programs. Evaluations are carried out by evaluators who are functionally independent, but often lack factual independence. There is also limited international competition in the market for evaluations commissioned by national or regional authorities. Evaluation methods applied in CP program evaluations mostly lag behind academic advancements and evaluation reports often do not transparently describe their methodological limitations. As the EU body responsible for implementing CP across all 27 Member States, the Commission may also have an overly optimistic perspective on CP. Finally, there is little evidence that evaluation findings are used for decision-making processes, funding allocation and the design of programs. The paper offers a number of recommendations how to advance the evaluation system: (1) Reorient CP reforms towards a more focused set of objectives; (2) Specify evaluation obligations more precisely in the Common Provision Regulation and set out a 'charter for evaluators'; (3) Introduce an 'evaluate first' requirement when preparing or updating programs; (4) Promote the use of counterfactual methods; (5) Explicitly link funding decisions at program and policy level to evaluation results; (6) Implement measures to stimulate a European market for CP evaluations; and (7) establish a standing European Advisory Panel on CP evaluation to foster independent third-party reviews.

Table of Contents

Executive Summary	4
1 Introduction	7
2 CP Evaluation – the Status Quo and the 2021-2027 Regulation Changes	11
2.1 History and Main Trends.....	11
2.2 Actors Carrying Out Evaluations	11
2.3 Other Stakeholders.....	12
2.4 Evaluation Requirements According to Current CPR.....	13
2.5 Changes Compared to the Last Program Period.....	13
2.6 Methods Used	14
3 Robust Evaluation Systems and Methods	15
3.1 A Growing Consensus on Evaluation Principles	15
3.2 Clarity on Policy Objectives.....	15
3.3 Evaluation Culture	16
3.4 Capacity and Expertise	16
3.5 Impartiality and Unbiasedness of Evaluators	16
3.6 Appropriate Methods.....	16
3.7 Transparency on Methodological Limitations and External Validity	17
3.8 Unbiased Aggregation of Insights	18
3.9 Use in Decision Making	18
3.10 Timing Aligned to Policy Decisions	19
4 Imperfections in the Current Evaluation CP System	19
4.1 Goal Congestion.....	19
4.2 Heterogeneous Evaluation Culture.....	21
4.3 Lack of Capacities and Expertise.....	22
4.4 Lack of Effective Evaluator Independence	23
4.5 Slow Methodological Progress and Lack of Clarity on Methodological Limitations	24
4.6 Lack of Differentiated and Unbiased Aggregation of Evaluation Results	26
4.7 Limited Use in Decision Making	27
4.8 Evaluation Timing not Synchronized with Decision Taking	28

5 Recommendations	29
5.1 Reorient CP Reforms Towards a More Focused set of Objectives	29
5.2 Specify Evaluation Obligations More Precisely in the CPR and Set out a ‘Charter for Evaluators’	30
5.3 Introduce an ‘Evaluate First’ Requirement when Preparing or Updating Programs	30
5.4 Promote the Use of Counterfactual Methods	31
5.5 Explicitly Link Funding Decisions at Program and Policy Level to Evaluation Results	32
5.6 Implement Measures to Stimulate a European Market for CP Evaluations ...	33
5.7 Establish a Standing European Advisory Panel on CP Evaluation.....	34
References	35
Appendix I: Common Provision Regulations – Chapter II Evaluations.....	39
Appendix II: Excerpt from Recommendation of the OECD Council on Public Policy Evaluation (OECD 2022)	40

Executive Summary

The EU's Cohesion Policy (CP) aims to promote economic, social and territorial cohesion in Europe. To this end, the EU's core budget devotes about one third of its financial resources to this policy. High-quality independent evaluations are key to ensuring that these resources are used effectively and efficiently. This report applies principles of good evaluation to assess the CP evaluation system. By international standards and in comparison with other EU policies, the system is well developed and institutionalized. Nevertheless, this analysis identifies shortcomings and derives recommendations on how the CP evaluation system can be further developed in the next programming period.

The report begins with a brief summary of the history, actors and main stakeholders of the CP evaluation system. It describes the main rules defined in the Common Provision Regulation (CPR) and the changes compared to the last programming period. Some of these changes were aimed at simplifying the system. These changes include the end of an obligation for ex-ante evaluations and the deletion of the explicit obligation for managing authorities to provide the necessary resources for evaluations.

Gaps in the Light of Principles for Good Evaluations

The assessment of the CP evaluation system applies the OECD's principles for the design of a good evaluation system. Although the CP is the most evaluated EU policy there are the following significant gaps:

- Meaningful evaluations require a well-defined policy objective against which performance can be measured. There has been an inflation of CP policy objectives in recent years with the transformation of CP towards a policy that increasingly seeks to support the full EU policy agenda. The resulting ambiguity of the objective function is a key challenge for evaluations.
- The evaluation culture is not equally developed among all Member States and regions. In quite a few cases, an unfavorable equilibrium is found which is characterized by limited evaluation capacities, poor methods, and a formalistic approach to evaluations. Moreover, according to current observations and assessments, the lack of evaluation capacity and expertise remains a significant problem in regions and Member States.
- The CPR requires evaluators to be functionally independent. This rule does not preclude that stakeholder interests have an impact on evaluation outcomes. The European Commission, Member States and their managing authorities all have an

interest in demonstrating the successful use of EU funds through good evaluation results. Evaluators are confronted with these client incentives. Although reputation concerns point in the direction of unbiased reports, service providers have a business interest in meeting their clients' expectations.

- The lack of impartiality is exacerbated by closed and oligopolistic national evaluation markets. Very few evaluators are involved in evaluations outside of their home country. Typically, only a few groups of evaluators tend to dominate these segmented national markets.
- Methods such as counterfactual approaches, which could verify a causal effect are rarely used. The same is true for theory-based evaluations that consider possible causal channels. While simpler methods are often justified by data and resource constraints, there remains a lack of transparency about the methodological limitations involved.
- Finally, there is little evidence that evaluation findings are used in decision-making, resource allocation and program design.

Recommendations

The report makes a number of recommendations for advancing the evaluation system:

- Reorient CP reforms towards a more focused set of objectives: Reform proposals, such as the suggestion to refocus Cohesion Policy on the convergence of the poorest regions, respond to the need for a clearer division of labor among EU policies. While a comprehensive reform along these lines may prove difficult to achieve, it is crucial for the upcoming negotiations to recognize a fundamental trade-off: the proliferation of policy objectives for CP makes its success or failure increasingly difficult to measure.
- Specify evaluation obligations more precisely in the CPR and set out a 'Charter for Evaluators': The next revision of the CPR should reintroduce an explicit statement that Member States must allocate sufficient resources to their evaluation process. Similarly, the European Commission's obligation to provide guidance on evaluation standards (methods, capacities, and procedures) should be reinstated. The European Commission should set out a 'Charter for Evaluators' indicating the minimum quality standards an evaluation should meet.
- Introduce an 'evaluate first' requirement when preparing or updating programs: Like the Commission, Member States should apply an 'evaluate first' principle when preparing or updating programs. In this regard, we consider that no longer requiring Member States to carry out ex-ante evaluations has been a too far-reaching change. A better compromise on the trade-off between simplification and

evaluation benefits is possible. We recommend the reintroduction of a focused ex-ante evaluation requirement, focusing on particularly important programs.

- Promote the use of counterfactual methods: Counterfactual policy evaluations require not only methodological skills, but also experimental or quasi-experimental setups that allow a control group to be compared to a treated group. The design of CP programs should therefore pay more attention to the preparation of (quasi-experimental) evaluation designs.
- Explicitly link funding decisions at the program and policy level to evaluation results: We propose to refer back to the ‘evaluate first’ principle outlined above. In particular, we propose that Member States should be required to support any changes that involve a significant reallocation of funds on the basis of an evaluation. Similarly, any proposal for a program would need to be accompanied by an ex-ante evaluation or impact assessment of the previous program.
- Implement measures to stimulate a European market for CP evaluations: We recommend opening the borders of the market for CP evaluation services. International team composition should be used as a criterion in tenders for national evaluation service contracts above a certain budget threshold. Quality assurance would benefit from peer review requirements where some of the reviewers must come from other Member States.
- Establish a standing European Advisory Panel on CP evaluation: We recommend the establishment of a “European Advisory Panel on CP Evaluations” (EAP-EVAL). As a fully independent and non-political body, the EAP-EVAL would advise the European Commission and the managing authorities in the Member States on their evaluation activities and provide an independent third-party review of their evaluation activities and reports.

Enhancing Objectivity and Decision Relevance: A Better Framework for Evaluating Cohesion Policy

Friedrich Heinemann¹, Zareh Asatryan, Julia Bachtrögler-Unger, Carlo Birkholz, Francesco Corti, Maximilian von Ehrlich, Ugo Fratesi, Clemens Fuest, Valentin Lang, Martin Weber²

1 Introduction

The EU's Cohesion Policy (CP) aims at promoting economic, social, and territorial cohesion. These overarching policy objectives are set out in the Treaty (Art. 174 TFEU). Over time, and from one programming period to the next, the EU legislators have substantially broadened and further specified the policy's objectives, using the CP in support of a significant part of the Union's comprehensive policy agenda (Leino-Sandberg 2024) including pandemic stabilization policies (European Court of Auditors 2023a). In the 2021-2027 period, the policy's objective function encompasses the support of a more competitive and smarter Europe, a greener and resilient, carbon-free Europe moving towards a net-zero carbon economy, a more connected Europe by enhancing mobility, a more social and inclusive Europe implementing European Pillar of Social Rights and a Europe closer to citizens (Common Provision Regulation: CPR, Art. 5).

Evaluations are a tool for assessing whether CP delivers on these broad ambitions. The explicit EU approach to an "EU budget focused on results" (Begg et al. 2023) implies that a policy should not be judged solely by its budget and the amount of financial resources allocated to it (the input), but by the results, the outputs and outcomes, as well as the impact of this policy and its budget.

Generally, the EU budget, and CP in particular, are seen as frontrunners in performance-orientation compared to budgetary systems in most OECD countries (Downes et al. 2017). In addition to the traditional financial information, performance budgeting requires the collection of relevant and reliable monitoring data on outputs and outcomes. Both are an indispensable contribution to performance budgeting, because a CP

¹ Corresponding author: Friedrich Heinemann, friedrich.heinemann@zew.de, ZEW – Leibniz-Centre for European Economic Research, L7,1, 68161 Mannheim, Germany.

² This text expresses the personal opinion of the author and not that of the European Court of Auditors.

budget focused on results needs a reliable understanding whether and to which extent the budgetary resources have contributed to reach the intended results and defined policy objectives. Evaluations have the role to provide this knowledge.

More precisely, the insights from evaluations serve different purposes (Fratesi 2024b; Pellegrin et al. 2020). They support a learning function as they help program managers to adjust and improve the program design and operation and the legislators to reconsider the policy design and to take informed decisions on budgetary allocations. Evaluations also serve an accountability function as they inform the public on whether public money has been used in a responsible way. Additionally, they fulfil an advocacy function: The knowledge generated can be used in debates and negotiations on how to fund and develop the policy in the future.

At face value, the EU's CP evaluation system appears to be rather mature. As CP entails significant spending from the EU budget, it falls under the budgetary ex-post evaluation requirement. In accordance with the EU's Financial Regulation (Article 34) and its rules of application, such evaluations are mandatory at least once during each MFF term. Detailed provisions regarding the time and nature of the evaluation are usually set out in the basic act of the respective program. The CPR (European Union 2021) assigns specific roles to the European Commission and the Member States whereby the main responsibility for evaluation is assigned to the national and regional Managing Authorities running the CP programs. It also defines concise rules and criteria according to which all CP programs have to be evaluated by functionally independent evaluators within the Member States. In particular, the CPR introduces a legal requirement to evaluate programs: as a consequence, the evaluation coverage of CP programs is extremely high, as nearly all CP programs financed from the European Regional Development Fund (ERDF), the European Social Fund (ESF), the Cohesion Fund (CF), and the Youth Employment Initiative (YEI) in the 2014-2020 period were subject to at least one evaluation (Asatryan et al. 2024). Moreover, the CP evaluation system is transparent since all EU and Member State evaluations are available publicly at the Cohesion Open Data Platform.

The Commission's 2015 Better Regulation reform and the 2016 Inter-institutional Agreement on Better Law-Making have further emphasized the role of evaluation in the EU's policy cycle: as a backward-looking tool, ex-post evaluations provide an evidence-based assessment of how existing legislation, policies and programs perform. This, in turn, is meant to help in the design of new interventions. In particular, the Commission applies an 'evaluate first' principle whereby any revision of EU legislation should build on the lessons learned from a preceding evaluation.

Overall, the EU evaluation system is more developed than similar systems in most OECD countries (European Court of Auditors 2019, S. 40). However, even a highly developed

system with a long history is not necessarily perfect. Several observations are due at this point.

First, there is a contrast between the messages from evaluations and the assessments by academics and researchers. Member State and Commission evaluations tend to paint a largely positive picture on the success of both single programs and the policy as a whole. A textual analysis of more than 2,500 Member State evaluations shows that none of them comes to very negative conclusions whereas the large majority presents good or very good assessments (Asatryan et al. 2024). Similarly, prominent Commission synthesis documents like the 8th Cohesion Report presents a list of largely positive evaluation findings across all policy priorities and complements this with messages from macro-simulations claiming a high and long-lasting overall growth effect (European Commission 2022).

In contrast, the assessments in the advanced academic literature are more nuanced. The bottom line from this literature is one of conditional and sometimes limited effectiveness (Bachtrögler et al. 2020; Di Caro and Fratesi 2022; Bachtrögler-Unger 2024; Ehrlich 2024; Fratesi 2024a, 2024b; Ehrlich and Overman 2020; Lang 2024). CP causally impacts growth and employment in a positive way, but related results often show up only conditional on institutions (e.g. administrative capacity, quality of institutions) and the availability of crucial production factors (e.g. human capital). Moreover, effects are often found to be place-specific or of a limited duration. A further contrast between policy evaluations and academic literature relates to methods used and the interpretation of results and data. Economic scholars are typically much more hesitant in attributing economic developments to causes of CP whereas EU evaluations are quick in interpreting changes that occurred during a project to a cause of the project. As a consequence of the “credibility revolution” in economic research (recently illustrated in 2021 with the Nobel Memorial Prize in Economics for David Card, Joshua Angrist and Guido Imbens for their contributions to the analysis of causal relationships), the more recent scientific papers have resorted to quasi-experimental research designs which are more convincing on causal chains.

Second, the specific perspectives of crucial players continue to be important in the evaluation process. Program managers, managing authorities, national governments, and the European institutions (European Commission, European Parliament, Council) all have different roles and interests that might impact the evaluation process and its outcome (Bachtler and Wren 2006). The Commission, as the budgetary executive, wants to demonstrate a responsible use of budgetary resources. For the Members of the European Parliament, CP is an attractive financial policy instrument with significant visibility in home constituencies. The positions of national governments in the Council are influenced by the (differing) financial motives in CP. All these interests might potentially

interfere with the impartiality of the evaluations carried out and how evaluation findings are assessed and interpreted.

Third, carrying out evaluations and publishing evaluation findings does not guarantee in itself that learning processes actually take place. Evaluations can support a budget focused on results only if evaluations have consequences for decisions on the program design and the budget allocation. For this purpose, all important European and national stakeholders would have to share an evaluation culture where evaluations are seen as an input to decision-taking rather than a mere formal requirement.

Against this background, this paper develops suggestions on how to further advance the CP's evaluation system in the next programming period. Here, we pay particular attention to strategies that may further foster the factual independence, impartiality and quality of Member States' program evaluations given current limitations of the evaluation market's supply side (studied in-depth in Asatryan et al. 2024). Moreover, we explore how policy evaluations could be inspired from the above sketched insights from the academic literature on conditional effectiveness and more advanced methods.

Two caveats are important: First, even an ideal evaluation system would not improve budgetary decisions by itself. An effective evaluation system is merely a necessary, but not a sufficient condition for a budget that is effectively focused on results (Robinson and Last 2009). The crucial conditions include, for example, a commitment of all budgetary stakeholders to consider policy outcomes and impacts instead of a mere input perspective (OECD 2019). In this context, the setting up of the Recovery and Resilience Facility (RRF) in the wake of the COVID-19 crisis has a particular relevance. Like CP, the RRF aims at funding investments to promote the economic, social and territorial cohesion; at the same time its performance-based delivery mechanism ties payments directly to the satisfactory achievement of pre-defined results, i.e. milestones and targets (European Court of Auditors 2023b). Second, CP operates in a multi-annual context, whereby allocations are fixed and allocated to specific Member States. This limits policy learning from evaluation within a programming period. Third, evaluations including necessary data provision are costly and can impose a large financial and administrative burden on managing authorities. Therefore, evaluation requirements should always be proportionate. Advanced methods that are appropriate for academic papers in leading international scientific journals cannot and should not always be applied to each single CP program evaluation. However, this does not preclude the search for a continuous and gradual improvement of evaluation methods.

The analysis presents in the next section a brief description of the current CP evaluation system. Section 3 describes principles of an ideal evaluation system that are contrasted with the status quo in Section 4. Section 5 develops a list of 5 proposals on how to improve the system in the coming programming period.

2 CP Evaluation – the Status Quo and the 2021-2027 Regulation Changes

2.1 History and Main Trends

The origin of the current CP evaluation system can be traced back to the late 1980s. Over time, the system has evolved and the evaluation approach has become more formalized and specified in the legislation (Pellegrin et al. 2020). In particular, the assignment of evaluation tasks to the European and national authorities has been clarified. Since the 2007-2013 period, the main legal basis defining the formal rules and procedures is the Common Provision Regulation (CPR). For each programming period, a new CPR is adopted by both EU legislators, the EP and Council, based on a Commission proposal which is negotiated and modified during the legislative procedure (European Union 2006, 2013, 2021). When looking at the CPRs of the last three periods, there was a move towards a more results-oriented approach, with a particular push in the 2014-2020 period, and a tendency to unify the monitoring system with respect to the collection and use of indicators (although most of them remained program-specific rather than horizontal and therefore not suitable for a cross-program assessment). Since 2015, the Better Regulation Guidelines have exerted significant influence in promoting comprehensive evaluation throughout the policy cycle and on the specification of evaluation criteria (Pellegrin et al. 2020, see also below 2.5). These guidelines also recommend evaluations based on five distinct criteria, namely effectiveness, efficiency, relevance, coherence, and EU added value.

2.2 Actors Carrying Out Evaluations

The important players in the CP evaluation system are the following:

European Commission: The European Commission has the responsibility for certain types of evaluations (see below). Moreover, it supports the evaluation activities of Member States and their authorities through methodological guidance (European Commission 2013, 2014, 2021) and by providing a forum for exchange and discussion through the Evaluation Network. The Commission also provides syntheses of Member State evaluations and makes these evaluations accessible through the Cohesion Open Data Platform.

Member States: In the Member States, managing authorities are tasked to carry out the evaluations of CP programs under their responsibility. Centralized models with national responsibility as managing authority coexist with decentralized models where regional authorities manage and evaluate the programs (for case studies see: Pellegrin et al.

2020). For each program, Member States have to set up monitoring committees. These committees approve the managing authority's evaluation plan and monitor the evaluation process. The Member State's evaluation activities can be financed from the technical assistance provided through the CP programs.

Evaluators: Finally, the evaluators, who actually conduct the evaluations commissioned by the managing authorities or the European Commission are important stakeholders as well. This group comprises public and private research institutes, universities, private consultancies, individual experts, but also internal evaluators from the civil service.

2.3 Other Stakeholders

Other stakeholders in the CP evaluation system are:

European Parliament: The Commission is accountable to the Parliament and has to report to the Parliament in annual consultations (Art. 8 CPR) and on important milestones like mid-term reviews (Art. 18 CPR), or in specific circumstances, for example, if a suspension of cohesion payments is imminent (Art. 19 CPR). The Commission also has to communicate its own CP evaluations to the Parliament (Art. 45 CPR). The Parliament is free to commission own evaluations or conduct assessments of CP through its own services.

Council: similar to the EP, the Council has no direct role in the CP evaluation, but also monitors the implementation of the policy. The Council plays, however, a particular role as the competent grouping is composed of or represents the ministries in charge of carrying out the program evaluations in the Member States.

European Court of Auditors: The European Court of Auditors in its capacity as the EU's independent external auditor, carries out performance audits, not unlike evaluations, which examine the effectiveness and efficiency of programs, operations, management systems and procedures. These audits complement its compliance audits which examine the regularity of the expenditure incurred and co-financed from the EU budget.

2.4 Evaluation Requirements According to Current CPR

The CPR regulation currently in place for the 2021-2027 period includes a chapter on evaluations with two articles (for the full text see appendix).³ They mandate the following:

- Member States or their managing authorities have to write an evaluation plan, which is approved by the monitoring committee.
- Member States or their managing authorities have to carry out two types of evaluations (Art. 44 CPR): First, they have to conduct evaluations of all programs according to at least one criterion from a list (effectiveness, efficiency, relevance, coherence and Union added value), which correspond to a large extent to the internationally accepted criteria for good government evaluations (OECD 2022). Further criteria may be examined in addition⁴. Second, Member States or managing authorities have to provide ex-post impact evaluations for each program by 30 June 2029.
- All evaluations have to be published digitally.
- The Commission is responsible to carry out mid-term evaluations by the end of 2024 and an ex-post evaluation of CP (“retrospective evaluations”) by end of 2031.
- Evaluations can be entrusted to internal or external experts who have to be functionally independent.

2.5 Changes Compared to the Last Program Period

Compared to the preceding programming period, with the revised CPR for 2021-2027, the evaluation rules have changed (Naldini 2018; Corti et al. 2024; European Commission 2021). The adaptations include:

- Member States have to set out an evaluation plan for their programs. This plan has to be produced within one year following the program adoption.
- Member States must evaluate their programs using five criteria (stemming from the Better Regulation agenda): effectiveness, efficiency, relevance, coherence with other EU priorities and policies and EU added value. During the CPR negotiations with the EP and the Council, it was agreed that evaluations must apply at least one

³ This analysis focuses on CP financed by the core budget. For an analysis how the performance orientation and evaluation differ for the Recovery and Resilience Facility see European Court of Auditors (2023b) and Corti et al. (2024).

⁴ In practice, CP evaluations have mostly covered the effectiveness and efficiency criteria (Pellegrin et al. 2020).

of these criteria, and that they can also add other criteria (such as inclusiveness, non-discrimination, etc.).

- Member States are no longer required to conduct ex-ante evaluations. This requirement has been dropped as a way to simplify the process and reduce the administrative burden (European Commission 2021).
- Member States have to conduct impact evaluations as before, but they have to be provided only by mid-2029, i.e. two years after the end of the 2021-2027 programming period (rather than before the end of the period as was previously the case).
- The object of Member States' evaluation is defined broader than before. In the past period, evaluations had assessed how a program has contributed to each "priority axis", which defines a specific policy aim like support of SME or employment. Under the new CPR, the evaluation requirement only refers to an operational program as a whole. Evaluations may even cover several programs at the same time. This change has decreased the number of obligatory evaluations substantially, but also poses methodological challenges since an operational program includes very different instruments with various aims (Naldini 2018).
- The Commission's obligation to conduct mid-term evaluations is new (Naldini 2018).
- The evaluation articles in the CPR have been shortened. Some explicit obligations that had been included in the preceding CPR edition have been deleted. These cuts include the Commission's task to provide guidance on evaluations and the Managing Authority's responsibility to provide the necessary resources for evaluations.

2.6 Methods Used

The CPR does not define methodological details. Various evaluation methodologies can be employed, including mainly basic quantitative and qualitative approaches conducted via desk research, interviews, surveys, and case studies. The use of methods is necessarily different for ex-ante and ex-post evaluations (for an overview: Fratesi 2024b, chapter 7). For ex-post impact evaluations, counterfactual methods focus on whether the intervention has had an effect that can be causally linked to the cohesion policy intervention. These methods have become the state of the art in the academic literature, but as they are more demanding, they are relatively rarely employed in CP evaluations. They make use for instance of difference-in-difference approaches or regression discontinuity designs. Theory-based impact evaluations focus on "why" and "how" questions with a particular interest into the precise causal channels that can explain causality from the intervention up to its outcome and impact ("theory of change" approach) (for a detailed discussion: Begg et al. 2023). More qualitative analysis is justified when there are only a small number of beneficiaries, or if data constraints or

proportionality considerations prevent the use of more demanding counterfactual methods. Lastly, cost benefit analysis estimates both the project’s financial profitability and its economic rate of return, thereby evaluating the cohesion policy intervention’s benefits to society as a whole (European Commission 2008, 2013a).

In the next section, we describe how a state-of-the-art evaluation system should look like, to then identify the weaknesses in the current system and the gaps between the status quo and the ideal stylized evaluation system.

3 Robust Evaluation Systems and Methods

3.1 A Growing Consensus on Evaluation Principles

A recent concise summary of principles that should guide the design of a robust evaluation system has been provided by the OECD Council on Public Policy Evaluations (OECD 2022) (see Appendix II).⁵ Fratesi (2024b, chapter 7) develops how such principles could be applied to CP. Other requirements result from the role of evaluations within performance budgeting (Robinson and Last 2009), or the emerging new methodological breakthroughs in the academic literature summarized above. We regard the following requirements as key for the assessment of the CP evaluation system:

3.2 Clarity on Policy Objectives

A meaningful evaluation of a public intervention is impossible without a well-defined policy objective: “any assessment of EU expenditure should start from a clear definition of the logic of intervention” (Begg et al. 2023, p. 47). The requirement of clarity includes transparency about priorities or weights if a policy intervention serves different objectives at the same time. The ex-ante definition of success implies also to have a notion of when to discontinue an intervention. Ideally, the extent to which a policy objective is reached can be quantified on the basis of robust and reliable performance indicators that are available when needed to inform policy decisions. In practice, it is however often difficult to specify such indicators upfront. This makes an ex-post evaluation even more important.

⁵ See also OECD (2021) and Gesellschaft für Evaluation e.V. (2023).

3.3 Evaluation Culture

A thriving evaluation culture describes a setting where the important stakeholders share a willingness to learn. This includes the readiness to challenge existing policy approaches and programs. This principle is in tension with stakeholder interests (Bachtler and Wren 2006; Naldini 2018). For example, national authorities and European institutions may have a motivation to demonstrate a successful use of CP resources. A developed evaluation culture requires the readiness to learn from open-minded and impartial evaluations. This includes the readiness to admit failures. An evaluation culture can improve over time and needs to be supported by institutions and rules that foster impartiality and learning effects.

3.4 Capacity and Expertise

High-quality evaluations need resources and expertise both at the level of the program authorities and the evaluating institutions. To some extent, this is the responsibility of the EU and national institutions that have to allocate sufficient resources to the evaluation task. With respect to the evaluating contractors, capacity and expertise also depend on how the evaluation market is organized. An open and international European evaluation market with links to academic research will be able to provide higher methodological standards as compared to narrow national markets without those links.

3.5 Impartiality and Unbiasedness of Evaluators

A high evaluation culture critically relies on credible and impartial evaluations. Credibility grows with impartiality of evaluators, and sound methods and data (see below). A mere functional independence of evaluators as required in the CPR is a necessary, but not a sufficient condition for impartiality. Evaluators have institutional interests as well. They want to provide a service that meets the expectations of the contracting authority in order to stay in the business and to win follow-up projects. An evaluator's financial dependence on a few contracting authorities, related career motives, or narrow links from a long-standing cooperation in a small national evaluation market could impair impartiality and credibility. Evaluators might then face incentives to whitewash unfavorable empirical evidence, or to blur the conclusions from and communication of disappointing results.

3.6 Appropriate Methods

The use of a variety of methods in CP evaluations is legitimate and well justified. Evaluations have various functions that range from an assessment of a program

implementation to an ex-post impact evaluation.⁶ These varying functions also translate into different methodological requirements. The choice of methods should follow the principle of proportionality: Methodological ambitions and the invested resources have to increase with the strategic and budgetary importance of a project. Equally, higher efforts are indispensable if there is a particular lack of knowledge on an intervention's possible impact, e.g. if the type of intervention is innovative. The 2021-2027 Commission guidance document (European Commission 2021) clarifies, that even for the Member States' ex-post impact evaluations a variety of methods can be used including "simpler impact evaluation techniques" (p. 14). However, methodological efforts should be higher if, for example, large budgets are involved, or if an intervention is innovative and there is a lack of prior evidence.

Although a variety of measures is legitimate, this is no excuse for sticking to outdated methods in many cases. CP impact evaluations should strive for a continuously increasing level of rigor that echoes the development of the academic literature on causal inference. The buildup of capacities and expertise discussed above can support the familiarity with these methods. Proportionality and costs are relevant limiting factors, but the costs for rigorous impact analysis can be lowered if the data requirements are already considered in the initial program design and legislation.

The methodological choices should always be well explained and ideally be documented beforehand. Pre-analysis plans (PAPs) and pre-registration have become more important in science and ideal CP evaluations would include these. This serves as a precaution against specification searching for the most desired outcome. While full PAPs may, in many cases, be too much to ask from CP evaluation, it would be important to agree on standards, methods, and data to be used before the actual evaluation starts.

Methodological standards should be supported through appropriate quality assurance. This includes the definition of quality standards, publication requirements, and external scrutiny through peer review and meta-analyses of each authority's evaluations.

3.7 Transparency on Methodological Limitations and External Validity

The credibility of the evaluation system can also be supported with a high transparency on strengths and weaknesses of methods used. In particular, any communication of evaluation results that addresses non-experts should signal the limits of knowledge and the different reliability of the various methods. For example, macro-modeling simulation studies that demonstrate the CP's potential to increase growth and employment

⁶ See Fratesi (2024b) for a survey on the arsenal of evaluation methods that can inform ex-ante and ex-post evaluations of CP.

should not be presented as if they could prove these results ex-post. Equally, point estimates should be presented in a way that conveys the statistical uncertainty that relates to confidence intervals. Most importantly, mere case studies, other qualitative approaches, correlational analyses and output counts (e.g. number of jobs/firms that have received support) should not be misleadingly reported with claims of indicating causality or “success”. In general, the realized values of performance indicators relative to target values set ex ante should not be the sole basis for the assessment. As a minimum requirement, the evaluation should describe a consistent logic of the mechanism of how the intervention may contribute to the policy objective. In the best case, this mechanism is verified using counterfactual methods. However, also the limits of causal inferences should be addressed openly. For example, sound evidence for a temporary effect from a program should not be sold to the public as if this could already demonstrate a long-lasting impact. Finally, in the last programming period, Member States’s managing authorities made widespread use of the possibility of modifying performance indicators prior to performance reviews leading to better assessments (European Court of Auditors 2021). Such moving of goalposts should also be made transparent when assessing the achievement of performance targets.

3.8 Unbiased Aggregation of Insights

Thousands of Member State evaluations create an information overflow that cannot inform the policy debate on the overall performance of CP without aggregation. Here, synthesis reports are essential by condensing the various findings, and identifying strengths and weaknesses of the policy. In this aggregation process, impartiality and unbiasedness are equally important principles as in each single program evaluation. These principles are even more important if this synthesis and its communication are dominated by actors which are perceived to have a strong institutional interest.

3.9 Use in Decision Making

Evaluations are not an end in themselves but have the function to inform decision making. Evaluations that are not considered in decisions are therefore a waste of resources. For this reason, insights from evaluations should be easily available to program managers, the budget executive, the legislator, and the general public, including the media. All these stakeholders should be able and incentivized to base their reflections and decisions on what they have learned from evaluations. A good evaluation culture should seek integration of the evaluation system into the decision-making processes. That includes formalized follow-up mechanisms that track how decision makers react to recommendations by the evaluations. In the end, evaluation results should have a significant impact on budget allocations and program design.

3.10 Timing Aligned to Policy Decisions

A specific precondition for an evaluation system that looks to influence decision making is the timely provision of insights. Insights from evaluation should therefore be available when the budgetary authorities or the executive prepare or take decisions. The existence of a developed system of ex-ante, interim, and ex-post evaluations that follow the budgetary cycle from the phase of budgetary negotiations and decisions up to implementation and, finally, completion of programs, would serve this purpose. In addition, the exact deadlines for each type of CP evaluation should be aligned to the exact time of the decisions. For example, program impact evaluations should be ideally available when the reflections on the next programming period approach the decision stage, although in practice this may be difficult.

With these principles for good evaluations, we developed a yardstick to identify weaknesses of the current CP evaluation system.

4 Imperfections in the Current Evaluation CP System

As emphasized, the CP evaluation system is highly developed and mature as it has gone through continuous adjustments over decades. In terms of its formalization and coverage it is advanced compared both to other EU policies and to evaluation systems in most EU Member States. Nevertheless, it still has a number of shortcomings, some of which are of a systematic nature and relate to the institutional interests of key stakeholders.

Based on various recent reports and papers (European Court of Auditors 2019; Naldini 2018; Pellegrin et al. 2020; Asatryan et al. 2024; Fratesi 2024b), and our own judgment, we see the following main imperfections with respect to the preceding principles.

4.1 Goal Congestion

Already for decades, CP has been confronted with the problem of the “inflation of objectives” (Heinemann et al. 2009). This trend has accelerated over more recent years with the transformation of CP towards a policy that increasingly wants to support the full EU policy agenda (Leino-Sandberg 2024). The CP’s objective function for the 2021-2027 period, as defined in Art. 5 CPR (see introduction), is so broad that it would be hard to imagine which specific policy objective could not be subsumed under these headlines. Relatedly, the ECA stresses that the policy is confronted with the problem of

overlapping EU strategies and sector-specific commitments. The resulting coexistence of multiple strategic frameworks, periods, objectives, indicators and targets creates confusion (European Court of Auditors 2019). The problem is exacerbated by a lack of transparency in how trade-offs are to be assessed. For example, a prominent trade-off of the CP is between regional disparities and aggregate (national) efficiency, as more regional convergence may come at the cost of lower growth in agglomerations.

This goal congestion, combined with a lack of precision in the weighting of competing objectives, raises several problems. The fuzziness of the policy objective function makes a coherent policy design difficult. Policy beneficiaries can practice “target shopping” by selecting from a large universe of possible objectives and targets those which best serve their own agenda and are most useful to legitimize a given budget. For evaluations, the lack of precision of the objective function poses a fundamental problem as any meaningful evaluation needs a well-defined yardstick to assess performance (Fratesi 2016). Multiple objectives thus immunize a policy against a negative performance for single objectives and serve the interests of those with an interest in keeping and enlarging the CP budgets under all circumstances. Another reason for multiple objectives is a preference for flexibility in steering the policy on current (and changing) necessity.

This issue’s relevance is confirmed by evaluators. Asatryan et al. (2024) report in their evaluator survey that more than 60 percent regard unclear policy objectives as a bottleneck for the CP evaluation system and rank this problem at position three (only surmounted by data imperfections and a lack of evaluation impact). The academic literature summarized in the introduction tends to circumvent this problem by applying a relatively narrow focus on growth and employment effects. However, policy makers can easily refute those insights as lacking policy relevance since these studies do not pay sufficient attention to the broad universe of CP policy objectives.

On a more operational level, evaluations could try to avoid these challenges for the policy as a whole with a focus on the narrower objective of each single policy intervention. However, the fuzziness at the aggregate level translates into practical difficulties for evaluation designs. This is particularly relevant as the revised CPR only prescribes Member State evaluations at the program level, and no longer at the level of priorities that relate to a more narrowly defined policy target (section 2.4). This broader scope of evaluations in the 2021-2027 programming period will exacerbate the fuzziness of objectives problem even more.

A related aspect is the increasing overlap between EU and national funding areas. Indeed, CP (and more recently the RRF) basically can fund nearly all types of measures that are traditionally funded from national budgets. In fact, over the last three programming periods, the ‘additionality principle’ for CP has been gradually abandoned

(European Court of Auditors 2023b). This means that there is no longer any requirement for CP to fund interventions that are additional to those funded by national budgets, including recurrent public expenditure. In such a system of mixed financing, attributing specific results to one funding stream is inherently difficult.

4.2 Heterogeneous Evaluation Culture

Evaluation culture has always been dissimilar across Member States. For the EU-15, Bachtler and Wren (2006) describe that the evaluation of regional policies had a stronger tradition in parts of Northern Europe. Cultural differences may also have an impact on whether evaluations are seen as a bureaucratic burden rather than a tool for continuous learning. In Member States with such a preconception, any willingness to conduct high-quality evaluations is naturally limited, and the main intention might be to just comply with the minimum evaluation requirements. For example, a comprehensive data collection is a pre-condition for evaluations using counterfactual designs. Regarding reporting of data on funded projects, several Member States also in Southern and Eastern Europe stand out in providing detailed data exceeding the formal requirements set out in the CPR since the 2007-2013 programming period, while others - including Northern and Middle European countries - report exactly or less than the minimum information required (see Bachtrögler-Unger et al., (2021), p. 4 and 7, for the 2014-2020 programming period). From interviews and a survey among CP evaluation stakeholders, Pellegrin et al. (2020) confirm that a lack of evaluation culture is still a relevant issue today: Evaluations are often regarded as a formal obligation and, as a consequence, as an exercise that generates higher costs than benefits. Begg et al. (2023) report that, with respect to EU-financed social programs, some administrations rather account for how much was spent rather than to assess targets, let alone the causal impact.

Interestingly, Pellegrin et al. (2020) find that project managers and respondents from management authorities tend to find evaluations more useful for learning purposes than high level national politicians. This may be related to the more political perspective that looks at CP funds predominantly as a European financial transfer to the own country or region. The sense of accountability towards the own electorate through credible evaluations is naturally lower for resources that come from external sources, compared to the money financed by the own country's taxpayers.⁷

⁷ The lack of domestic accountability is an important reason discussed in the development economics literature as an explanation for a lack of state capacity and the frequent ineffectiveness of foreign aid. With the external financing, domestic politicians need not develop an implicit contract with their citizens and prove the responsible use of money in exchange for levying taxes (Deaton 2015).

4.3 Lack of Capacities and Expertise

A lack of evaluation capacities and expertise remains a significant issue in regions and Member States according to current observations and judgments (Naldini 2018; Cerqua and Pellegrini 2023). Pellegrin et al. (2020) summarize insights from European Commission scrutiny that 80 percent of Member State evaluations from the 2014-2020 reviewed in 2018 had major quality deficiencies; in 2019 the Commission rated the average quality of evaluations at 2.5 out of a maximum of 4 points indicating a mediocre average quality.

The use of the more advanced methods (e.g. theory-based and counterfactual methods) requires substantial expertise, constant training, and investment into methodological knowledge, as well as the data infrastructure. The required skills comprise not only methods, but also the ability to write concise and relevant policy conclusions that can guide decision makers.

The heterogeneity in evaluation culture described above correlates and is mutually reinforcing with a lack of capacities and expertise (Pellegrin et al. 2020). Countries with a weak evaluation culture have an incentive to limit resources to the minimum that is just about sufficient to fulfill the formal requirements. Conversely, a lack of evaluation expertise explains that the understanding for the merits of evaluation stays at a low level. There can thus be a stable “bad equilibrium” of low capacities, poor expertise, and a weak evaluation culture that is self-enforcing.

The lack of methodological skills tends to be more pronounced in Member States that have a lower level of human capital, and where universities and research institutes are not yet well-integrated into the international academic communities. In those countries, the availability of experts that are able to apply modern evaluation tools is relatively limited. In principle, an open European market for evaluations could compensate for this handicap of smaller and poorer countries. This is however far from being the European reality today. A common market for CP program evaluations does not yet exist, also due to linguistic barriers. Member States’ CP evaluation markets are largely separated and closed with almost full absence of international collaboration. Asatryan et al. (2024) show in their analysis of evaluations from the last two programming periods that evaluators only rarely collaborate across borders. From more than 2,000 evaluation authors included in their study, only 2.5% have contributed to evaluations in more than one Member State. This points to the obvious absence of incentives to incorporate evaluators from abroad. De iure, the market is not expected to hinder international collaboration.

Another capacity constraint relates to data. The availability of reliable and timely data to measure outputs, outcomes and impact are a precondition for any quantitative

evaluation. However, the data available in the Cohesion Open Data Platform are self-reported from Member States. Self-reporting as such is an inherent necessity, but reliability may be low due to lack of external verification and auditing. Thus, there is no guarantee for their quality. Nevertheless, the obligatory provision of detailed project-level data requires Member States to engage more in documenting their activities than in the past.⁸

As described above (2.4) the 2021-2027 CPR has not only dropped explicit obligations to strengthen evaluation capacities in the Member States, but also the European Commission's obligation to give Member States guidance on how to carry out evaluations. Equally, it no longer obliges Member States explicitly to provide the necessary resources for evaluations. This streamlining of the regulation may hinder further improvements and implies the risk that some Member States with lacking evaluation culture could take this as a signal to further cut back on evaluation resources, efforts, and quality assurance (Naldini 2018; Pellegrin et al. 2020).

4.4 Lack of Effective Evaluator Independence

The CPR requires evaluators, which may be internal or external, to be functionally independent (Art. 44). This rule is helpful but does not preclude that stakeholder interests have an impact on evaluation outcomes (Bachtler and Wren 2006).

These interests are a logical consequence from above-described evaluation functions (learning, accountability, and advocacy) which stand in a certain mutual tension. The learning function requires unbiased and impartial guidance from the evaluation. Incentives are different for the accountability and advocacy functions. On accountability, the European Commission, Member States and their managing authorities all share the interest to demonstrate to the national and European public a successful use of EU funds through good evaluation results. In this respect, the central role of the managing authorities to steer evaluations is seen as particularly problematic since this authority will not be immune to national interests of CP; this constellation is likely to result in pressure on evaluators regarding the findings of evaluations (Pellegrin et al. 2020, p. 58; Naldini 2018). On advocacy, beneficiaries and public administrators of CP funds will find positive evaluation results more useful than subdued ones. Pellegrin et al. (2020) reports the example of how the four Visegrad countries have used CP evaluations in negotiations on the 2014-2020 EU budget to call for generous funding of CP. Similar incentives exist for regional authorities also in richer Member States. Even if these Member States are CP net payers, sub-national authorities will still have an interest to keep their regional programs financed through the EU budget.

⁸ A related and fundamental issue is that of timeliness of data, which is addressed in Section 4.8.

Likewise, the European Commission as the central budget authority with a political and institutional interest to command a substantive and growing budget may want to use evaluation results to argue for a stable or growing EU budget, of which CP accounts for a significant share (see below 4.6).

Evaluators from consultancies, research institutes and freelancers are confronted with these motives whenever they bid for, or conduct and submit evaluation reports commissioned by national or sub-national authorities. Although reputation concerns point into the direction of unbiased reports, these service providers have a business interest to satisfy their customers' expectations in order to stay in the market. With this trade-off between reputation concerns and business interests, effective independence would be supported if evaluators had a diversified business model with many different customers. As shown above, this is however not the case. Due to a strong national segmentation of CP evaluation markets, most evaluators provide their services to the authorities in one single Member State only, which almost always is their home country. Hence, impartiality not only suffers from a non-diversified customer structure but potentially also from a home bias. It is known from other contexts like internationally operating rating agencies that experts tend to give better assessments to their home country (Fuchs and Gehring 2017).

Moreover, these national evaluation markets have an oligopolistic structure where just a few (not more than three author clusters) tend to dominate the market, often even with a 100 percent market share (Asatryan et al. 2024). Asatryan et al. (2024) also show from their evaluator survey and an AI-based analysis of evaluation reports that the intensity of sponsor involvement in the evaluation process is correlated with a more positive tonality of the reports. The same survey shows that even 29 percent of the responding evaluators admit that a lack of impartiality is an issue for the CP evaluation system, although this survey question is clearly sensitive and evaluators should be hesitant to admit the relevance of the problem.

Overall, the functional independence of evaluators as prescribed by the CPR is therefore not enough to safeguard factual independence of evaluators. This comes at the risk that the current evaluation process may not be regarded as sufficiently independent.

4.5 Slow Methodological Progress and Lack of Clarity on Methodological Limitations

Generally speaking, a variety of methods is justified and appropriate for evaluations in CP, also in view of capacity constraints and the large resources needed for more demanding evaluation methods. However, research shows that very few evaluations make use of causal methods, thereby lagging behind what is practiced in other fields, with

development policy evaluations being the most advanced one. An analysis by Pellegrin et al. (2020, p. 49) for almost 1,400 Member State CP evaluations from the last programming period classifies 48 percent as qualitative and only 2.8 percent as counterfactual impact analysis. Also, sound theory-based evaluations that explicitly consider possible causal channels, which are a promising direction especially for ex-ante analyses and are verifiable in ex-post evaluations, are still rare. As a consequence, there is generally a lack of understanding how and through which specific mechanism a budgetary intervention contributes to a policy objective; moreover, the actual extent of the contribution in light of possible confounding variables is rarely assessed (Begg et al. 2023).

Another shortcoming is a lack of transparency on methodological limitations in the presentation of evaluation results. For example, evaluations using simple qualitative methods or indicator-based assessments should clearly point out their limitations. The careful addition of caveats that is a standard for academic papers is rarely practiced in CP evaluations. It is a common exercise both at Member States and the European level, for example, to count jobs or SMEs which have received EU funding without clarifying that this does not at all show whether the EU money causally created or preserved jobs, or fostered company performance.

The lack of methodological clarity also relates to more advanced methods. Counterfactual methods, in order to estimate a causal relationship, often have to narrow the scope of their analysis, which is not always sufficiently discussed. Another issue exists in particular for model-based simulation approaches that aim to demonstrate the growth and employment effect of CP on a macro-perspective. Often, they are presented as if they could prove the *actual impact* although they are only able to show the *policy potential* because modelling results depend crucially on model assumptions. For example, to assess the impact of infrastructure on regional growth the usual assumption is that the new infrastructure is useful and actually used by economic agents, which may not be true in case of a poor project selection. A further assumption is that all financial resources used have been allocated to the building of the new infrastructure. However, this may be not fully the case if construction delays occur or if there is corruption. If such implicit beneficial assumptions do not correspond to reality, econometric counterfactual models that are able to identify the actual impact will show different and often lower effects as compared to the model simulations (Fratesi 2024b). Furthermore, it is important to be transparent about the benchmark scenario on which the modeled policy potential is based. For example, as compared to exploring the impact of adding “fresh” money to the model framework, modeling the policy impact vis-à-vis an alternative distribution of funding (e.g., the current distribution of national public funding) could allow more relevant insights.

There is also a lack of clarity on the time dimension. A quantitative evaluation that confirms a positive effect from CP e.g. on the number of jobs should ideally indicate whether this effect is short-lived or longer lasting. If the method and the data are unable to decide this question this needs to be clarified. Fratesi (2024b) points out that politicians seeking

re-election may be keen to get immediate results. By contrast, a sustainable CP should rather focus on long-run and persistent effects. Policy design may imply trade-offs where long-run effects need more patience because of longer time lags. Evaluations should be aware of this crucial time dimension, otherwise they promote policy myopia.

Another methodological issue is the consequence of the narrow focus of Member State evaluations on their own national or regional programs. Because these evaluations do not include cross-country comparisons (and rarely cross-regional), they cannot account for the impact of uniform national factors that do not vary within one country (or region).⁹ Hence, by construction, Member State analyses are blind for the significant role of national (or regional) bottlenecks. This is a serious deficiency, since these factors have been identified to crucially explain the success or failure of CP in the academic literature (e.g. regulation, lack of human capital, low administrative capacity, deficient institutions etc., see introduction above). These methodological limits of merely national assessments interact with evaluator incentives (see 4.4) not to lose favor with national authorities through a too explicit critique of deficiencies for which the national government bears the responsibility. The lack of cross-country collaboration of evaluators also lowers the evaluators' awareness of the role of differing national factors for the success of a European policy intervention.

4.6 Lack of Differentiated and Unbiased Aggregation of Evaluation Results

The Cohesion Policy Open Data Platform currently provides access to more than 2,500 Member State evaluations from the last two programming periods. On first sight, this is an impressive number, also in view of the fact that there are around 400 CP programs in each programming period. However, the sheer number of these program-specific evaluations also makes it difficult to learn about the overall success of CP and about the reasons for different degrees of success in different circumstances. Hence, information aggregation is key for policy learning. Here, the European Commission currently holds a crucial position. This institution has the task to provide comprehensive ex-post evaluations, and it condenses their insights into reports that play a prominent role in the CP policy debate (e.g. the Cohesion Reports).¹⁰ This central role of the European Commission in this aggregation exercise is not without risks. As argued above, this

⁹ If there is regional variation, national studies are able to control for the impact of institutional constraints, however.

¹⁰ One could argue that, for example, the Cohesion Reports sometimes give an overly optimistic account of the impact of CP. For example, Chapter 9 of the Eighth Cohesion Report (European Commission 2022) summarizes evaluations results on the impact of CP. Essentially all evaluation results included in this chapter have a positive message. Where a limited impact is admitted or where an impact is not yet visible this is often justified e.g. by hints that the measures were not yet completed, or that the objective is long-term and can be expected to materialize later. Furthermore, evaluation results are listed without reference to the underlying methods or their methodological shortcomings. Even for an informed reader, it is difficult to judge which of the results originate from a more descriptive approach, and which are related to a more credible counterfactual analysis.

institution is a stakeholder with a particular institutional interest. With respect to the accountability function of evaluations, the Commission as the EU budget executive has an interest to prove a responsible use of EU money and, hence, a successful CP. As a political actor, it legitimately takes a position in support of a policy that it deems important for European cohesion, political integration and the achievement of EU policy objectives. However, this institutional interest may go against providing a balanced synthesis of evaluation insights that could best support an open and unbiased learning process on the success and failures of CP.

4.7 Limited Use in Decision Making

Evaluation findings are rarely used for decision-making processes, funding allocation and the design of programs. This is an overwhelming finding of relevant reports and studies, including by the European Court of Auditors (European Court of Auditors 2019). This not only holds for Member States but also for the European Commission. For none of the operational programs that had failed to meet their milestones the Commission ever suspended payments; instead, milestones have been adjusted to the actual level of performance to reach formal compliance (European Court of Auditors 2021). Evaluators themselves share the frustration that their insights do not reach the decision stage: 69 percent of the evaluators surveyed by Asatryan et al. (2024) regard a lack of decision impact as a problem (and much more important than issues like a lack of resources or impartiality). Similarly, the comprehensive stock taken by Pellegrin et al. (2020) paints a mixed picture according to which the European Commission tends to make use of evaluation findings for drafting proposed regulations for the next period, but that evaluation insights compete with political priorities, also those of the Commission (in line with the reasoning from the preceding section). For Member States, Pellegrin et al. (2020) only find rare examples that document the use of evaluation findings for programming decisions.

Ex-ante evaluations can help in an early stage when budget decisions are taken, but also when a program is implemented. Evaluations that use theory-based approaches can, for example, raise manager awareness for the program logic and important design features. However, contrary to the past period, the obligation for Member States to conduct ex-ante evaluations has been removed from the CPR for 2014-2020. The intention for this step was to cut back on the large evaluation burden that, due to an often merely formal compliance, was regarded to be a waste of resources. However, this end of any ex-ante evaluation obligation can be seen as too far reaching and detrimental for evidence-based policy making (European Court of Auditors 2019; Naldini 2018; Pellegrin et al. 2020).

4.8 Evaluation Timing not Synchronized with Decision Taking

Evaluation results cannot inform decisions if they are not available when decisions are taken. Evaluation requirements and deadlines should therefore be well integrated into the policy cycle from the preparation to the implementation and assessment of the policy initiative. In this respect, some of the revised evaluation obligations and deadlines for the 2021-2027 (see 2.4) may be counterproductive.

For example, Member States have to carry out their impact evaluations of 2014-2020 programs by June 2029 and the European Commission is obliged to provide retrospective evaluations for each fund by the end of 2031. Both types of ex-post evaluations come far too late to inform the Commission's preparation of the post-2027 legislative proposal (due in mid-2025) and its subsequent negotiation with the EP and the Council (European Court of Auditors 2023b). They even come too late to inform program managers when they start to implement new programs from 2028 onwards. Obviously, timing decisions involve a trade-off. Data requirements and the particular interest in longer-term effects (3.7) point in the direction of longer deadlines, whereas the need to provide information at the decision stage requires shorter deadlines. In light of this trade-off the current CPR's innovation to oblige the European Commission to provide mid-term evaluations of each fund by the end of 2024 is to be welcomed as a positive novelty (Naldini 2018; European Court of Auditors 2019) as this will provide useful insights in time for the debate on the next programming period.

However, the span of data that can be used for these mid-term evaluations is limited to a maximum of three full years (2021-2023). In reality, even less data will be available since cohesion programs had a slow start in the 2021-2027 period. Apart from the delays related to the pandemic there was a very low initial absorption of cohesion funds. The simultaneous implementation of the RRF without national co-financing needs and a shorter eligibility period, ending in 2026, as well as the two-year extension of the 2014-2020 programs through the REACT-EU initiative has contributed to this situation (Bachtrögler-Unger 2024). Therefore, the 2024 mid-term evaluation of the 2021-2027 CP programs may have little to analyze.

5 Recommendations

The preceding gap analysis has revealed that a lot needs to be done to provide CP with a truly unbiased, self-critical and decision-oriented evaluation system. Reforms should be aimed at gradually breaking away from any remaining cases of bad equilibria of low expertise, mediocre methods, biases, low credibility, and formalistic application of evaluation obligations. We think that our following seven recommendations below can facilitate a transition towards a good equilibrium characterized by high expertise, more stringent standards, impartiality, credibility, and genuine policy learning.

5.1 Reorient CP Reforms Towards a More Focused set of Objectives

Addressing the deeper-rooted obstacles to a more independent, methodologically sound, and decision-relevant evaluation system poses a complex challenge with no easy fix. One fundamental and overarching problem lies in the broad and ambiguous objective function of the policy, which could only be rectified through a substantial policy reform that strives for a clearer assignment of policy objectives to different types of EU policy instruments. It may be worth recalling the Tinbergen Rule that advocates a distinct policy tool for each policy target. This approach would provide a more structured framework, emphasizing responsibilities and significantly enhancing the foundation for effective performance assessment.

Reform proposals, such as the suggestion to refocus Cohesion Policy on the convergence of the poorest regions (Fuest 2024), align with the need for a clearer division of labor among EU policies. While a comprehensive reform along these lines may prove challenging to attain, it is crucial for upcoming negotiations to recognize a fundamental trade-off: the proliferation of policy objectives for Cohesion Policy renders its success increasingly challenging to measure.

A related aspect is the increasing overlap between EU and national funding areas. Indeed, CP basically can fund nearly all types of measures that are traditionally funded from national budgets. A clearer demarcation between EU and national policies with a focus of the EU funding on areas with a demonstrable EU added value could be envisaged. The re-introduction of an updated ‘additionality principle’ as a key conditionality for CP could help making clearer (and easier to evaluate) what the EU funding is meant to achieve.

5.2 Specify Evaluation Obligations More Precisely in the CPR and Set out a ‘Charter for Evaluators’

The next CPR revision should reintroduce an explicit statement that Member States have to provide sufficient resources to their evaluation process. Equally, the European Commission’s obligation to provide guidance for evaluation standards (methods, capacities, and procedures) should come back. We also agree with Naldini’s (2018) recommendation to make the Commission’s operational guidelines for the managing authorities more specific in terms of minimum methodological standards. We add to this that the guidance should emphasize the need for more transparency on methodological limits of the methods used for the evaluation.

Finally, the European Commission should set out a ‘charter for evaluators’ indicating the minimum quality standards an evaluation must meet. This could be used as a basis for developing relevant training programs, both for evaluators and for staff at managing authorities dealing with evaluations. Courses on causal evaluation methods have become part of standard curricula in economics programs at many EU universities, and tailor-made courses for CP evaluators could be easily developed.

The necessary evaluation capacity building could be supported through a strategy of sticks and carrots. Some funds could be set aside to support Member State evaluation capacities, and sanctions if a Member State fails to reach the minimum standards could be introduced.

5.3 Introduce an ‘Evaluate First’ Requirement when Preparing or Updating Programs

Like the Commission, Member States should apply an ‘evaluate first’ principle when preparing or updating programs. In this regard, we consider that no longer requiring Member States to carry out ex-ante evaluations has been a too far-reaching change, even if the motive to simplify the process is understandable and justified. A better compromise on the trade-off between simplification and evaluation benefits is possible. We recommend reintroducing a focused ex-ante evaluation obligation, which concentrates on particularly important programs. The selection could be decided on the basis of budget size or other more content-oriented criteria that identify innovative approaches with a particular need for thorough impact reflections and program design. An alternative would be to require managing authorities to legitimate the planned allocations and activities in each operational program based on previous evaluations and experiences. This should also involve why certain measures included in the previous period’s program are not implemented anymore and why. This could pave the way for fewer but more thorough evaluations. Whatever criteria are used to choose which programs to

evaluate ex-ante, they should not leave room for interpretation in order to avoid any selection bias.

Finally, the issue of the timing of evaluations is crucial. While the mid-term evaluation could be helpful in this aspect, the next CPR should reconsider its deadlines for retrospective impact analyses to inform policy makers and programmers in a timely manner before programming the next period's policy. At least some of these evaluations (both by Member States and the European Commission) should be conducted before the end of the programming period, and early enough before the decisions on the next period are taken.

5.4 Promote the Use of Counterfactual Methods

Counterfactual policy evaluations do not only need methodological skills, but also experimental or quasi-experimental set-ups that allow to compare a control group to a treated group. Hence, the design of CP programs should pay more attention to prepare (quasi-experimental) evaluation designs. We see three possibilities, which may differ in their feasibility.

Implementation of actual experiments: From a methodological perspective, this would be the ideal approach. This “gold standard” requires randomizing funding (its levels, types, design features etc.) across eligible projects/regions. Even if this is the preferable approach from a methodological perspective, it may be difficult to implement, and it could also raise ethical and legal problems that relate to unequal treatment. But to the extent that some effects are clearly unknowable ex ante (and, hence, the treatment would not advantage or disadvantage anybody ex ante) some experimental features might be feasible. Preparations into this direction could benefit from an exchange with experimental researchers who have applied field experiments in development economics already for a long time, and where a lot of cooperation between scholars and public agencies, and policymakers is the norm.

The following two quasi-experimental evaluation designs pose less ethical and legal problems since they exploit structural differences between observations that are a consequence arising naturally from project selection and policy implementation.

- Quasi-experimental evidence with discontinuities: Exploiting discontinuities in CP spending has brought the breakthrough in the academic literature towards causal impact identification (Becker et al. 2008, 2012, 2013, 2018; Lang et al. 2023). This could inspire evaluations much more. It should be possible to make more use of discontinuities in funding allocated according to rankings of eligible projects where

funding goes only up to a cut-off. This cut-off can then be used in a regression discontinuity design by comparing projects just above/below the threshold.¹¹

- Staggered Difference-in-difference (DID) designs: Randomize the start date of projects and study them with staggered DID models that look at an event window before and after the start. Here, it would be important to collect data before the projects start.

As a general rule, for the experimental and quasi-experimental approaches it is important to collect data on units (projects, beneficiaries, regions) that are not funded, in order to be able to quantify policy effects against credible control groups.

5.5 Explicitly Link Funding Decisions at Program and Policy Level to Evaluation Results

Ultimately, evaluation results should also be reflected in the funding allocation for and within programs. This is however a complex undertaking. The experience made during the 2007-2013 period has shown that setting aside a performance-reserve whereby part of the funding is reallocated during program allocation cannot be the way forward. For example, it is not self-evident what to do if an evaluation finds that a program is not achieving its initial objectives: is this due to a lack of funding, or should the funding be reduced? Taking funding decisions on the basis of evaluation results requires robust evaluations that examine the underlying reasons and factors contributing to a (non-) successful program implementation.

Against this backdrop, we propose referring back to the ‘evaluate first’ principle set out above. In particular, we suggest that Member States must back up any program amendments which involve a significant reallocation of funding up by an evaluation. Similarly, any proposal for a program would need to be accompanied by an ex-ante evaluation or impact evaluation of the predecessor program. The European Commission (which must approve programs and their amendments) would then act as a goalkeeper to ensure that evaluation findings are properly taken into account in the program (re)design and the related funding (re)allocations. In that respect, the European Commission should also be given the right to unilaterally decide on funding (re-)allocations based on these evaluation findings.

In addition, a fundamental change of the post-27 CP may result from a potential adaptation of the RRF delivery mechanism where payments from the EU budget are directly linked to the satisfactory achievement of milestones and targets in the Member States. But even the change to such a performance-based funding system would not do away

¹¹ An illustration from EU research policy is the European Research Council (ERC) that builds such a ranking when deciding on allocating ERC research grants.

with the need to evaluate the effectiveness, efficiency, relevance of CP programs, and their coherence with other EU priorities and policies and EU added value.

5.6 Implement Measures to Stimulate a European Market for CP Evaluations

National evaluators possess a commendable understanding of their home country, including its institutions and political system, and normally have a superior command of the language compared to experts from other EU Member States. However, potential drawbacks associated with relying solely on national experts include their limited awareness of alternative institutions in other countries, financial dependence on national authorities for follow-up studies, a potential home bias in judgment, and informal ties to the government. These factors may compromise factual independence, the ability to identify deficient national institutions and policy failures, and the openness and unbiased nature of the evaluation process.

Hence, there is a trade-off. National teams have more specific country knowledge, but are less neutral and may suffer from a narrow perspective. Presently, the trade-off for Member State evaluations leans heavily towards an almost entirely closed evaluation market (Asatryan et al. 2024), a situation that poses significant challenges. A more open market would not only support independence, but could also quickly import evaluator expertise into those Member States which suffer from a particular shortage of advanced evaluator expertise.

We recommend to start a broad initiative to open the borders of the CP evaluation service market with appropriate incentives. These incentives could entail:

- use international team composition as one criterion in the calls for tenders for national evaluation service contracts: the collaboration with partner institutes from a minimum number of other Member States could become a formal requirement for evaluation contracts that relate to programs above a certain budget threshold;
- a peer review stage into the quality assurance system with the requirement that some of the reviewers must come from other Member States. A cross-country approach to quality checks and peer review could help to overcome national biases and to identify deficiencies in national evaluations.

We propose that the requirements for internationalization could be gradually increased, since over time cross-border collaboration of national evaluators would become a normality.

5.7 Establish a Standing European Advisory Panel on CP Evaluation

Finally, we recommend to set up a “European Advisory Panel on CP evaluation” (EAP-EVAL). As argued above, the European Commission has specific institutional interests that may hinder fully neutral CP assessments. As a fully independent and non-political body, the main role of the EAP-EVAL would be to advise the European Commission and the managing authorities in the Member States in their evaluation activities and to provide an independent third-party review of their evaluation activities and reports.

The EAP-EVAL could play a supporting and advisory role in the preparation of evaluation plans, the tendering of evaluation services and the setting-up of a European evaluator data base. It could also put into practice the ‘charter for evaluators’, prepare guidance and organize trainings for evaluators and people in managing authorities dealing with evaluations. Finally, its role could be to carry out an independent third-party review of evaluation reports. This would include ‘meta-evaluations’, which assess the processes, methods and quality of evaluations from one specific institution (e.g. one managing authority). Moreover, these meta-evaluations would draw more general conclusions from the aggregation of single evaluation results. These reviews should always be made public. Members of the EAP-EVAL should have the necessary expertise and familiarity with modern evaluation methods. They could come from academia, be evaluation practitioners, or dealing with evaluations in public administrations. Their work must be firmly governed by professional standards and practices. As a member of the EAP-EVAL, however, they should not be involved in any evaluation of EU policies or programs, to preclude any risk of a conflict of interest.

Alternatively, the EAP-EVAL could also play a more active role, and in particular directly carry out impact evaluations of CP. For example, it could be designed with features similar to the Intergovernmental Panel on Climate Change (IPCC). The IPCC serves the purpose to accumulate knowledge on climate change and the effectiveness of climate policy instruments based on scientific knowledge and through a transparent procedure that wants to guarantee credibility of the results. The IPCC reports have become an authoritative source of information due to a careful procedure to safeguard a high degree of impartiality and variety. IPCC authors are selected on the basis of their expertise and with the intention to cover a diversity of socio-economic views and backgrounds, as well as geographical and gender balance. IPCC reports undergo an external expert review (IPCC 2024).

Adding such a body to the CP evaluation system could lend additional credibility to the evaluation findings reported by the Commission and the managing authorities in the Member States.

References

- Asatryan, Z., C. Birkholz, F. Heinemann (2024), “Evidence-based Policy or Beauty Contest? An AI-Based Meta-Analysis of Cohesion Policy Evaluations by Member States”, *Contribution to BMF Expert Network*.
- Bachtler, J., C. Wren (2006), “Evaluation of European Union Cohesion policy: Research questions and policy challenges”, *Regional Studies* 40 (2), 143-153.
- Bachtrögler, J., U. Fratesi, G. Perucca (2020), “The influence of the local context on the implementation and impact of EU Cohesion Policy”, *Regional Studies* 54 (1), pp. 21-34.
- Bachtrögler-Unger, J. (2024), “The Role of Administrative Capacity for an Effective Implementation of EU Cohesion Policy”, *Contribution to BMF Expert Network*.
- Bachtrögler-Unger, J., E. Arnold, M. Doussineau, P. Reschenhofer (2021), *UPDATE: Dataset of projects co-funded by the ERDF during the multi-annual financial framework 2014-2020*, Publications Office of the European Union, Luxembourg.
- Becker, S. O., P. H. Egger, M. v. Ehrlich (2008), “Going NUTS: The effect of EU Structural Funds on regional performance”, *CESIFO Working Paper* no. 2495.
- Becker, S. O., P. H. Egger, M. v. Ehrlich (2012), “Too much of a good thing? On the growth effects of the EU's regional policy”, *European Economic Review* 56 (4), 648-668.
- Becker, S. O., P. H. Egger, M. v. Ehrlich (2013), “Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects”, *American Economic Journal: Economic Policy* 5 (4), 29-77.
- Becker, S. O., P. H. Egger, M. v. Ehrlich (2018), “Effects of EU Regional Policy: 1989-2013”, *Regional science and urban economics* 69, 143-152.
- Begg, I, F. Corti, A. Liscai, T. Kiss-Gálfalvi, T. Grossi, R. Laura (2023), Social tracking methodology for the EU budget, Study requested by the BUDG Committee, European Parliament, [https://www.europarl.europa.eu/RegData/etudes/STUD/2023/742788/IPOL_STU\(2023\)742788_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2023/742788/IPOL_STU(2023)742788_EN.pdf) (accessed 3 June 2024).
- Cerqua, A., G. Pellegrini (2023), “I will survive! The impact of place-based policies when public transfers fade out”, *Regional Studies* 57 (8), 1605-1618.

References

- Corti, F., M. Pedralli, C. Pancotti (2024): “The Recovery and Resilience Facility: key innovations and the interplay with Cohesion Policy”, *Contribution to BMF Expert Network*.
- Deaton, A. (2015), Weak States, Poor Countries, Blog Social Europe, <https://www.social-europe.eu/weak-states-poor-countries> (accessed 3 June 2024).
- Di Caro, P., U. Fratesi (2022), “One policy, different effects: Estimating the region-specific impacts of EU cohesion policy”, *Journal of Regional Science* 62 (1), 307-330.
- Downes, R., D. Moretti, S. Nicol (2017), “Budgeting and performance in the European Union”, *OECD Journal on Budgeting* 17 (1), 1-60.
- Ehrlich, M.v. (2024), “The Importance of EU Cohesion Policy for Economic Growth and Convergence”, *Contribution to BMF Expert Network*.
- Ehrlich, M.v., H.G. Overman (2020), “Place-Based Policies and Spatial Disparities across European Cities”, *The Journal of Economic Perspectives* 34 (3), 128-149.
- European Commission (2013), The Programming Period 2014-2020: Guidance document on monitoring and evaluation - European Regional Development Fund and Cohesion Fund, https://ec.europa.eu/regional_policy/sources/evaluation/2014/wd_2014_en.pdf (accessed 3 June 2024).
- European Commission (2014), *The Programming Period 2014-2020: Guidance document on monitoring and evaluation - European Regional Development Fund and Cohesion Fund*, Brussels.
- European Commission (2021), “Performance, monitoring and evaluation of the European Regional Development Fund, the Cohesion Fund and the Just Transition Fund in 2021-2027”, *Commission Staff Working Document SWD* (2021) 198 final.
- European Commission (2022), *Cohesion in Europe towards 2050, Eighth report on economic, social and territorial cohesion*, Publications Office of the European Union, Luxembourg.
- European Court of Auditors (2019), “Delivering performance in Cohesion”, *Briefing Paper*, European Union, Luxembourg.
- European Court of Auditors (2021), Performance-based financing in Cohesion policy: worthy ambitions, but obstacles remained in the 2014-2020 period, Special report 24/2021, https://www.eca.europa.eu/EN/publications/SR21_24 (accessed 31 May 2023).

- European Court of Auditors (2023a), “Adapting cohesion policy rules to respond to COVID-19”, *Special Report* 02/2023.
- European Court of Auditors (2023b): “EU financing through cohesion policy and the Recovery and Resilience Facility: A comparative analysis”, *Review* 01.
- European Union (2006), “Regulation (EU) 1083/2006”, *Official Journal of the EU*.
- European Union (2013), “Regulation (EU) 1303/2013”, *Official Journal of the EU*.
- European Union (2021), “Regulation (EU) 2021/1060”, *Official Journal of the EU*.
- Fratesi, U. (2016), “Impact Assessment of European Cohesion Policy: Theoretical and Empirical Issues”, in S. Piattoni and L. Polverari, eds., *Handbook on Cohesion Policy in the EU*, Edward Elgar, Cheltenham, 443-460.
- Fratesi, U. (2024a), “Constraining and Enabling Factors of a Successful EU Regional Policy in Europe”, *Contribution to BMF Expert Network*.
- Fratesi, U. (2024b), *Regional Policy, Theory and Practice*, Routledge, Abingdon, Oxon, New York.
- Fuchs, A., K. Gehring (2017): “The Home Bias in Sovereign Ratings”, *Journal of the European Economic Association* 15 (6), 1386-1423.
- Fuest, C. (2024), “Fundamental Reflections on a More Rational Future EU Cohesion Policy”, *Contribution to BMF Expert Network*.
- Gesellschaft für Evaluation e.V. (Ed.) (2023), *DeGEval-Standards. Standards für Evaluation*, <https://www.degeval.org/degeval-standards/> (accessed 5 November 2023).
- Heinemann, F., T. Hagen, P. Mohl, S. Osterloh, M. O. Sellenthin (2009), *Die Zukunft der EU-Strukturpolitik*, Nomos, Baden-Baden.
- IPCC (2024), Structure of the IPCC, IPCC Authors and Review Editors, <https://www.ipcc.ch/about/structure/> (accessed 1 January 2024).
- Lang, V. (2024), “Umverteilungswirkungen von regionaler Strukturförderung in der EU”, *Contribution to BMF Expert Network*.
- Lang, V., N. Redeker, D. Bischof (2023), *Place-Based Policies and Inequality Within Regions*.

References

- Leino-Sandberg, P. (2024), “Cohesion Policy and the Principle of Subsidiarity – a Legal Analysis”, *Contribution to BMF Expert Network*.
- Naldini, A. (2018), “Improvements and risks of the proposed evaluation of Cohesion Policy in the 2021–27 period: A personal reflection to open a debate”, *Evaluation* 24 (4), 496-504.
- OECD (2019), *OECD Good Practices for Performance Budgeting*, OECD Publishing, Paris.
- OECD (2021), *Applying Evaluation Criteria Thoughtfully*, OECD Publishing, Paris.
- OECD (2022), Recommendation of the Council on Public Policy Evaluation, OECD/LEGAL/0478, <https://legalinstruments.oecd.org/api/print?id=686&lang=en> (accessed 3 June 2024).
- Pellegrin, J., L. Colnot, M. Pedralli (2020), The Role of Evaluation in Cohesion Policy, Study Requested by the REGI Committee, European Parliament, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/629219/IPOLE_STU\(2020\)629219_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/629219/IPOLE_STU(2020)629219_EN.pdf) (accessed 3 June 2024)
- Robinson, M., D. Last (2009): *A Basic Model of Performance-Based Budgeting*, IMF Technical Notes and Manuals, International Monetary Fund, Washington.

Appendix I: Common Provision Regulations – Chapter II Evaluations

Article 44

Evaluations by the Member State

1. The Member State or the managing authority shall carry out evaluations of the programs related to one or more of the following criteria: effectiveness, efficiency, relevance, coherence and Union added value, with the aim to improve the quality of the design and implementation of programs. Evaluations may also cover other relevant criteria, such as inclusiveness, non-discrimination and visibility, and may cover more than one program.
2. In addition, an evaluation for each program to assess its impact shall be carried out by 30 June 2029.
3. Evaluations shall be entrusted to internal or external experts who are functionally independent.
4. The Member State or the managing authority shall ensure the necessary procedures are set up to produce and collect the data necessary for evaluations.
5. The Member State or the managing authority shall draw up an evaluation plan which may cover more than one program. For the AMIF, the ISF and the BMVI, that plan shall include a mid-term evaluation to be completed by 31 March 2024.
6. The Member State or the managing authority shall submit the evaluation plan to the monitoring committee no later than one year after the decision approving the program.
7. All evaluations shall be published on the website referred to in Article 49(1).

Article 45

Evaluation by the Commission

1. The Commission shall carry out a mid-term evaluation to examine the effectiveness, efficiency, relevance, coherence and Union added value of each Fund by the end of 2024. The Commission may make use of all relevant information already available in accordance with Article 128 of the Financial Regulation.
2. The Commission shall carry out a retrospective evaluation to examine the effectiveness, efficiency, relevance, coherence and Union added value of each Fund by 31

December 2031. In the case of the ERDF, the ESF+, the Cohesion Fund and the EMFAF, that evaluation shall focus in particular on the social, economic and territorial impact of those funds in relation to the policy objectives referred to in Article 5(1).

3. The Commission shall publish the results of the retrospective evaluation on its website and communicate those results to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

Source: European Union (2021).

Appendix II: Excerpt from Recommendation of the OECD Council on Public Policy Evaluation (OECD 2022)

1. Conduct and use evaluations across government ensuring that they are carried-out in a systematic manner and that their results are used in policy and budgetary decision-making. In particular, Adherents should:

a) Designate evaluation champions to coordinate evaluations across institutions and advise on best practices to promote their quality and use.

b) Define and assign institutional responsibilities for conducting policy evaluations.

2. Foster a culture of learning and accountability by promoting demand for, and ownership of, evaluations within and beyond the executive. In particular, Adherents should:

a) Promote the role of both internal and external, national and international, knowledge brokers to strengthen the relationship between evidence from evaluations and its users, including citizens.

b) Offer opportunities for the legislative body to review and discuss policy evaluations.

III. RECOMMENDS that Adherents promote the quality of public policy evaluations. To this end, Adherents should:

1. Actively plan, design and manage evaluations so that they are timely and proportionate to the intended objectives, taking into account the needs of the primary users and the types of intended uses, and ensuring that results can be trusted by stakeholders. In particular, Adherents should:

a) Plan evaluations early by building provisions for evaluations into public interventions from the start, in order to improve their design, collect data on their implementation, and ensure that evaluation results are robust and available in a timely fashion.

b) Design and implement evaluations that are proportionate and appropriate for the likely use, by adapting the aim, scope and analysis of the evaluation, its format and resources, to the needs of its primary users and the types of intended uses.

c) Engage relevant stakeholders in the evaluation processes from the outset in order to create ownership for change and trust in evaluation results.

2. Establish quality standards and mechanisms for evaluations to generate robust and credible evaluation results that can be trusted and used with confidence. In particular, Adherents should:

a) Develop guidelines to ensure that evaluation designs, data collection processes and analytical methods, adhere to methodological best practices.

b) Adopt professional and ethical standards for evaluators to ensure that they meet high criteria for integrity and independence, as well as for knowledge of evaluative methods and culturally appropriate approaches, and that they safeguard the dignity, rights, safety and privacy of participants and other stakeholders when they conduct evaluations.

c) Promote the functional autonomy of evaluations, by safeguarding the autonomy of external evaluations through oversight of the commissioning and evaluation processes, and by providing internal evaluations team with a high degree of autonomy in the use of available resources and in deciding what studies to conduct and how.

d) Ensure that evaluations are able to withstand external scrutiny, such as through peer review, and that they can be assessed against pre-defined quality criteria.

3. Develop institutional skills and capacities to conduct, commission and use evaluations effectively and in a credible manner. In particular, Adherents should:

a) Build public sector skills for evaluation by conducting regular training, recruiting and retaining employees with the adequate skills or collaborating with academia, the private sector and other jurisdictions to improve the availability of these skills.

b) Ensure the availability of high quality, timely, accessible, disaggregated and re-usable results, performance and administrative data for policy evaluation.

c) Provide institutions with appropriate resources to manage, carry-out and use policy evaluations.

IV. RECOMMENDS that Adherents conduct public policy evaluations that impact decisionmaking. To this effect, Adherents should:

1. Establish institutional mechanisms to embed evaluation in decision-making processes, both at the organisational level and across government. In particular, Adherents should:

a) Provide high-level guidance, such as in a legal or policy framework or in a multi-annual evaluation agenda, on when to conduct policy evaluation and what type of evaluation is needed, in order to adapt the timing of evaluations to feed into decision-making processes, focus the analysis where it is most needed, co-ordinate efforts for cross-sectorial evaluations, and avoid overlaps.

b) Incorporate the use of evaluation results into decision-making including through the policy-making and budgetary processes.

c) Establish follow-up mechanisms for decision-makers to respond to the results of evaluations, by defining a course of action where relevant, and assigning responsibilities for implementing and tracking recommendations.

2. Provide easy access to evaluations and present the findings deliberately in order to improve the uptake of evaluation results. In particular, Adherents should:

a) Make the result of evaluation findings and recommendations public by default.

b) Tailor the way evaluation evidence is presented and communicated to its potential users, in terms of timing, communication channel, format and messaging, by developing a dissemination strategy.

c) Make use of evidence synthesis methodologies to aggregate evaluation findings and assess them in a systematic manner.

Authors of this Issue

Friedrich Heinemann



Friedrich Heinemann is head of the research department “Corporate Taxation and Public Finance” at ZEW – Leibniz Centre for European Economic Research, Mannheim, as well as professor of economics at the University of Heidelberg. His research interests are empirical public finance and political economy with a focus on European integration, taxation and fiscal institutions in the European Union and the euro area.

Contact: friedrich.heinemann@zew.de

Zareh Asatryan



Zareh Asatryan is deputy head of the ZEW research unit “Public Finance and Corporate Taxation”. His current research focuses on the political economy of the welfare states, the design of tax and transfer systems, and the economic impacts of fiscal policies. He holds a doctoral degree from the University of Freiburg.

Contact: asatryan@zew.de

Julia Bachtrögler-Unger



Julia Bachtrögler-Unger is an economist at the Austrian Institute of Economic Research (WIFO), where she has been working in the research group "Regional Economics and Spatial Analysis" since 2018. Her research focuses on regional and urban economics, in particular the analysis of economic policy measures in the regional context, the effects of EU cohesion policy, and regional development. Since April 2022, she has been a member of the Austrian Productivity Board.

Contact: julia.bachtroegler-unger@wifo.ac.at

Carlo Birkholz



Carlo Birkholz is a Ph.D. candidate at the University of Mannheim and a researcher in the Department of Public Finance and Corporate Taxation at the ZEW – Leibniz Centre for European Economic. He is also an external affiliate of the Norwegian Centre for Taxation (NoCeT). His broad research interests are in the fields of Political Economy, Public Finance and Development Economics.

Contact: c.birkholz@gmail.com

Francesco Corti



Francesco Corti is an Advisor to the Belgian Deputy Prime Minister and Minister for Social Affairs and Health and an Associate Research Fellow at CEPS. An expert in European social and employment policies, EU budget, and EU economic governance, he has provided research and advisory services to various EU institutions and agencies, including the European Commission, Parliament, Court of Auditors and Eurofound. Francesco holds a Ph.D. in Political Science from the University of Milan.

Contact: francesco.corti@ceps.eu

Maximilian von Ehrlich



Maximilian von Ehrlich is Professor of Economics at the University of Bern and the Director of the Center for Regional Economic Development. His research focuses on public economics, regional and urban economics. His research findings have been published in leading academic journals. He has conducted economic policy advisory projects for national and international institutions.

Contact: Maximilian.vonehrlich@unibe.ch

Ugo Fratesi



Ugo Fratesi is a Professor of Regional Economics and Policy at Politecnico di Milano, Italy, and the Editor-in-Chief of Spatial Economic Analysis. He is the author of the textbook "Regional Policy: Theory and Practice," published by Routledge. Additionally, he has authored several books and articles primarily focusing on regional growth and development, regional policy assessment, cohesion policy, territorial capital, models for forecasting regional growth, the theory and measurement of regional disparities, and the territorial aspects of human capital and innovation.

Contact: ugo.fratesi@polimi.it

Clemens Fuest



Clemens Fuest is President of the ifo Institute – Leibniz Institute for Economic Research at the University of Munich, Director of CESifo, Professor of Economics at the University of Munich (LMU), and Director of the Center for Economic Studies (CES) at LMU. His research areas are economic and fiscal policy, international taxation, and European integration. He is, among others, a member of the Scientific Advisory Board at the German Federal Ministry of Finance. Prior to his appointment in Munich, he was a professor at the Universities of Cologne (2001-2008), Oxford (2008-2013), and Mannheim (2013-2016).

Contact: fuest@ifo.de

Valentin Lang



Valentin Lang is an Assistant Professor of International Political Economy and Development at the University of Mannheim. Previously, he held positions at the universities of Zurich, Cambridge, Heidelberg, and Paris and has worked for and consulted the IMF, the OECD, and the European Stability Mechanism. His research focuses on studying the political, economic, and distributional implications of various dimensions of globalization and global governance.

Contact: lang@uni-mannheim.de

Martin Weber



Since 2010, Martin Weber has been working at the European Court of Auditors (ECA) in the field of auditing cohesion policy, first as head of unit and from 2014 onwards as the director of the department in charge of auditing cohesion policy. From 2017 until 2021, he was the director at the ECA in charge of strategy development. He also has a specific interest in performance measurement, evaluation and better law-making.

Contact: martin.weber@eca.europa.eu