

## Enhancing objectivity and decision relevance: a better framework for evaluating EU cohesion policies

Friedrich Heinemann <sup>a,b</sup>, Zareh Asatryan <sup>a,c</sup>, Julia Bachtrögler-Unger <sup>d</sup>, Carlo Birkholz <sup>a,e</sup>,  
Francesco Corti <sup>f,\*</sup>, Maximilian von Ehrlich<sup>g</sup>, Ugo Fratesi <sup>h</sup>, Clemens Fuest <sup>i,j</sup>, Valentin Lang <sup>k</sup>  
and Martin Weber<sup>l</sup>

<sup>a</sup>Department Corporate Taxation and Public Finance, ZEW – Leibniz Centre for European Economic Research, Mannheim, Germany; <sup>b</sup>Economics Department, University of Heidelberg, Heidelberg, Germany; <sup>c</sup>School of Business & Economics, University of Münster, Münster, Germany; <sup>d</sup>Research Group Regional Economics and Spatial Analysis, Austrian Institute of Economic Research, WIFO, Vienna, Austria; <sup>e</sup>Faculty of Business and Economics, University of Göttingen, Göttingen, Germany; <sup>f</sup>Centre for European Policy Studies (CEPS), Brussels, Belgium; <sup>g</sup>Department of Economics, University of Bern, Bern, Switzerland; <sup>h</sup>Department of Architecture, Built Environment and Construction Engineering (DABC), Politecnico di Milano, Milan, Italy; <sup>i</sup>ifo Institute, Munich, Germany; <sup>j</sup>Faculty of Economics, LMU, Munich, Germany; <sup>k</sup>School of Social Sciences, University of Mannheim, Mannheim, Germany; <sup>l</sup>European Court of Auditors, Luxembourg City, Luxembourg

### ABSTRACT

By international standards, the European Union's Cohesion Policy (CP) evaluation system is well developed and institutionalised. This paper analyses the remaining shortcomings and proposes recommendations for further refinement. A broad CP objective function emerges as a key challenge. Evaluation culture is not equally developed in all member states, and evaluation markets suffer from a lack of international competition. There is still limited use of counterfactual methods which have become standard in the academic literature on the effectiveness of CP, and there may be tensions between policy learning and demonstrating accountability. The paper proposes innovations to make CP evaluations more credible and impactful.

### ARTICLE HISTORY

Received 10 February 2025  
Accepted 27 January 2026

### KEYWORDS

European Union budget;  
Cohesion Policy; evaluation;  
performance budgeting;  
regional policy

### JEL

H43; H87; R58

## 1. Introduction

The European Union's (EU) Cohesion Policy (CP) aims to promote economic, social and territorial cohesion (Art. 174 of the Treaty on the Functioning of the European Union – TFEU). Evaluations are a tool for assessing whether CP delivers on these broad ambitions, serving different specific purposes (Fratesi, 2024; Pellegrin et al., 2020): evaluations support a *learning* function by helping programme managers improve programme design and implementation, and helping legislators reconsider the policy and make informed decisions on budgetary allocations. Evaluations also fulfil an *accountability* function because they inform the public about whether public money has been used responsibly. Additionally, they fulfil an *advocacy* function: the knowledge they generate can inform debates and negotiations on how to fund and develop the policy in the future.

In general, the EU budget, and CP in particular, are seen as frontrunners in performance-orientation compared with budgetary systems in most Organisation for Economic Co-operation and Development (OECD) countries (Downes et al., 2017; European Court of Auditors, 2019, p. 40). The Common Provisions Regulation (CPR) (European Union, 2021) assigns specific roles to the European Commission and member states (MS), with the primary responsibility for evaluation assigned to national and regional managing authorities that administer the CP programmes. It defines a legal requirement to evaluate programmes. Consequently, the evaluation coverage of CP programmes is extremely high, with nearly all CP programmes in the 2014–20 period being subject to at least one evaluation (Asatryan et al., 2025). Moreover, the CP evaluation system is transparent, as all EU and MS evaluations are publicly available on the Cohesion Open Data Platform.

**CONTACT** Friedrich Heinemann  [friedrich.heinemann@zew.de](mailto:friedrich.heinemann@zew.de)

\*Present affiliation: European Commission, Brussels, Belgium

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00343404.2026.2625765>

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

However, even a full formal compliance with the CPR's requirements does not guarantee that evaluations fulfil their learning function, especially if the evaluation system does not reflect the current state of knowledge on CP effectiveness and the role of regional and local conditions.

Recently, the influential High-Level Group report on the *Future of Cohesion Policy* (European Union, 2024) has provided a comprehensive overview of state-of-the-art empirical evidence on CP effectiveness. Together with the bulk of studies that apply convincing causal identification approaches, that report delivers a cautiously optimistic message: the CP as the largest place-based policy in the world has a positive impact on growth, employment and innovation, but this impact is conditional upon key enabling factors (Bachtrögler et al., 2020; Bachtrögler-Unger, 2025; Di Caro & Fratesi, 2022; Fratesi, 2024, 2025; Lang, 2025; von Ehrlich, 2024; von Ehrlich & Overman, 2020). Factors that are found to shape the (positive) effect of CP include regional institutional quality (Rodríguez-Pose & Garcilazo, 2015) and administrative capacity (Mendez & Bachtler, 2024), which improve not only policy outcomes but also evaluation practices. Other important factors include the availability of crucial production factors such as human capital (Becker et al., 2013), the amount of CP funding received (Becker et al., 2012) and territorial characteristics (Fratesi, 2025). Moreover, effects are often found to be place specific and of limited duration.

So far, the CP evaluation design does not yet adequately reflect these insights. The High-Level Group report (European Commission, 2024) recommends implementing a system for monitoring, evaluation and learning that is based on a more granular understanding of regional and local development dynamics. This could facilitate policy learning on a much finer scale rather than relying on a one-size-fits-all approach. Moreover, the report finds that the proliferation of ever new objectives, funds and instruments has increased policy complexity to the point that monitoring and evaluation have become difficult tasks.

Another limitation of CP evaluations relates to the methods used in national evaluations and how results and data are interpreted. Academic studies such as those reviewed in the High-Level Group report are typically more hesitant in attributing economic developments to the causes of CP, whereas CP evaluations contracted by managing authorities are quick in interpreting changes that occurred during a project as a causal effect of the project. Following the 'credibility revolution' in economics, more recent scientific papers have often resorted to quasi-experimental designs, which are more convincing on causal connections.

Finally, the specific perspectives of key stakeholders remain relevant in the evaluation process. Programme managers, managing authorities, national governments and European institutions (Commission, Parliament, Council) all have different roles and different interests in CP advocacy, which may impact the evaluation process and shape outcomes (Bachtler & Wren, 2006).

Against this overall backdrop, this paper confirms and spells out the High-Level Group's recommendation that the CP evaluation system of the future should better reflect what we know about CP effectiveness. In the following, suggestions are developed on how to further advance the CP's evaluation system in the next programming period. Particular attention is paid to strategies that could foster the factual independence, impartiality and quality of MS's programme evaluations given current limitations of the evaluation market's supply side (studied in-depth by von Ehrlich, 2025). Moreover, this paper examines how policy evaluations could draw inspirations from academic literature on conditional effectiveness and from more advanced methods. The proposed changes to the evaluation system aim not only to improve CP outcomes in future programming periods but also to increase trust among EU citizens.

## **2. CP evaluation: the status quo and the 2021–27 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 evaluation approach has become more formalised and specified in legislation (Pellegrin et al., 2020). Since the 2007–13 funding period, the main legal basis defining the formal rules and procedures is the CPR. For each programming period, a new CPR is adopted by both EU legislators, the European Parliament (EP) and Council, based on a Commission proposal which is negotiated and modified during the legislative procedure (European Union, 2006, 2013, 2021). Since 2015, the Better Regulation Guidelines have exerted

significant influence on promoting comprehensive evaluation throughout the policy cycle and on the specification of evaluation criteria (Pellegrin et al., 2020).

## 2.2. Crucial players

The most important players in the CP evaluation system are the European Commission, the MS and evaluators. The Commission has the responsibility for certain types of evaluations (see below). Moreover, it supports the evaluation activities of MS 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 MS evaluations and makes these evaluations accessible through the Cohesion Open Data Platform. In the MS, managing authorities are tasked to carry out the evaluations of CP programmes under their responsibility. Centralised models with national responsibility as managing authority coexist with decentralised models where regional authorities manage and evaluate the programmes (for case studies, see Pellegrin et al., 2020). Finally, the evaluators, who conduct the evaluations commissioned by the managing authorities or the Commission, are important stakeholders as well. This group comprises public and private research institutes, universities, private consultancies, individual experts and also internal evaluators from the civil service.

Other stakeholders in the CP evaluation system are the EP, the Council and the European Court of Auditors: the Commission is accountable to the EP and has to report to it in annual consultations (Art. 8 CPR), on important milestones such as mid-term reviews (Art. 18 CPR) or in special circumstances (Art. 19 CPR). It also communicates its CP evaluations to the EP (Art. 45 CPR). In addition, the EP is free to commission own evaluations or conduct assessments of CP through its own services. Next to the EP, the Council monitors the implementation of the policy. It plays a particular role as it is composed of or represents the ministries in charge of carrying out the programme evaluations in the MS. The European Court of Auditors, in its capacity as the EU's independent external auditor, carries out performance audits, not unlike evaluations, that examine the effectiveness and efficiency of programmes, operations, management systems and procedures. These audits complement its compliance audits that examine the regularity of the expenditure incurred and co-financed from the EU budget.

## 2.3. Evaluation requirements

The CPR regulation currently in place for the 2021–27 period includes a chapter on evaluations with two Articles (see Appendix A in the supplemental data online).<sup>1</sup> Compared with the preceding programming period, the evaluation rules have changed in several respects (Corti et al., 2025; European Commission, 2021; Naldini, 2018), as outlined below.

The CPR regulation mandates that MS or their managing authorities must provide an evaluation plan for their operational programmes within one year after programme adoption. This evaluation plan needs to be approved by the monitoring committee. MS or their managing authorities have to carry out two types of evaluations (Art. 44 CPR): first, they have to conduct evaluations of all programmes according to five criteria (effectiveness, efficiency, relevance, coherence and Union added value; OECD, 2022).<sup>2</sup> In the CPR negotiations with the EP and the Council, it was agreed that evaluations must apply at least one of these criteria, and that they can also incorporate other criteria (such as inclusiveness). Second, MS or managing authorities have to provide ex-post impact evaluations for each programme by mid-2029, and the Commission is responsible for carrying out an ex-post evaluation of CP overall ('retrospective evaluations') by the end of 2031.

In the 2021–27 period, MS are no longer required to conduct ex-ante evaluations to simplify the process and reduce the administrative burden (European Commission, 2021). By contrast, the Commission is now obliged to carry out mid-term evaluations by the end of 2024 (Naldini, 2018).

A further difference to previous programming periods is that the object of MS's evaluation is defined more broadly than before. In the past period, evaluations had assessed how a programme has contributed to each 'priority axis', which defines a specific policy aim such as small and medium-sized enterprise (SME) support or employment. Under the new CPR, the evaluation requirement refers to an operational programme as a whole. Evaluations may even cover several programmes at the same time. This change has

decreased the number of obligatory evaluations substantially, but also poses methodological challenges since operational programmes usually include different types of instruments with distinct aims (Naldini, 2018).

Finally, the evaluation articles in the CPR state that evaluations can be entrusted to internal or external experts who must be functionally independent. However, some explicit obligations that had been included in the preceding CPR edition have been deleted. These include the Commission's task to provide guidance on evaluations and the managing authority's responsibility to provide the necessary resources for evaluations.

## **2.4. Methods used**

The CPR does not define methodological details. Various evaluation methodologies can be employed, including 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, see Fratesi, 2024, ch. 7). For ex-post impact evaluations, counterfactual methods focus on whether the intervention has had an effect that can be causally linked to the CP intervention, for instance, by applying difference-in-difference (DID) approaches or regression discontinuity designs (RDDs). These methods have become the state of the art in the academic literature, but as they are more demanding, for example, in terms of data collection and the requirement of preparing the evaluation in parallel with policy design, they have been relatively rarely employed in CP evaluations. Theory-based impact evaluations focus on 'why' and 'how' questions with a particular interest in 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, see 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 analyses estimate both the project's financial profitability and its economic rate of return, thereby evaluating the CP intervention's benefits to society as a whole (European Commission, 2013).

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 stylised 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 also OECD, 2021; Gesellschaft für Evaluation e.V., 2023) (and see Appendix B in the supplemental data online). Fratesi (2024, ch. 7) develops how such principles could be applied to CP. Other requirements result from the role of evaluations within performance budgeting (Robinson & Last, 2009), or the emerging new methodological breakthroughs in the academic literature summarised 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 multiple objectives at the same time. The ex-ante definition of success also implies to have a notion of when to discontinue an intervention. Ideally, the extent to which a policy objective is reached can be quantified based on robust and reliable performance indicators that are available when needed to inform policy decisions. In practice, it is often difficult to specify such indicators upfront. This makes ex-post evaluation even more important.

### **3.3. Evaluation culture**

A thriving evaluation culture describes a setting where the important stakeholders share the willingness to learn. This includes the readiness to challenge existing policy approaches and programmes. This principle is in tension with stakeholder interests (Bachtler & Wren, 2006; Naldini, 2018). For example, national authorities and European institutions may have a motivation to demonstrate 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.

### **3.4. Capacity and expertise**

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

### **3.5. Impartiality and unbiasedness of evaluators**

An effective 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. An evaluator's financial dependence on few contracting authorities, related career motives or narrow links from a long-standing cooperation in small evaluation markets can impair impartiality and credibility.

### **3.6. Appropriate methods**

The use of a variety of methods in CP evaluations is well justified. Evaluations have various functions that range from an assessment of programme implementation to ex-post impact evaluation (Fratesi, 2024). These varying functions also translate into different methodological requirements. The choice of methods should follow the principle of proportionality: methodological standards should be higher if large budgets are involved or if an intervention is innovative and prior evidence is lacking.

Although a variety of measures is legitimate, this is no excuse for sticking to outdated methods. CP impact evaluations should strive for a continuously increasing level of rigour that echoes the development of the academic literature on causal inference to some extent. Proportionality and costs are relevant limiting factors, but the costs for rigorous impact analysis can be lowered if data requirements are already considered in the initial programme design and legislation.

The methodological choices should be well explained and ideally be documented beforehand. Pre-analysis plans and pre-registration have become more important in science and ideal CP evaluations would include these. While full pre-analysis plans may often be too much to ask for CP evaluation, it would be important to agree on the standards, methods and data to be used before the 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**

Any communication of evaluation results that addresses non-experts should signal the different reliability of the various methods. For example, macro-modelling simulation studies that demonstrate the CP's potential to increase growth and employment should not be presented as if they could prove these results ex-post. 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’. 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. The limits of causal inferences should also be addressed openly. For example, sound evidence for a temporary effect of a programme should not be sold to the public as if this could demonstrate a long-lasting impact. Finally, in the last programming period, MS’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 MS 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 programme 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 programme managers, the budget executive, the legislator and the public, including the media. A good evaluation culture should seek integration of the evaluation system into the decision-making processes. That includes formalised 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 programme 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 programmes would serve this purpose.

## **4. Imperfections in the current evaluation CP system**

With the yardstick of evaluation principles describe above and based on various recent reports and papers (Asatryan et al., 2025; European Commission, 2024; European Court of Auditors, 2019; Fratesi, 2024; Naldini, 2018; Pellegrin et al., 2020), and our own judgment, we see the following main imperfections.

### **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 recently accelerated with the transformation of CP towards a policy that increasingly wants to support the full EU policy agenda (Leino-Sandberg, 2025). In the 2021–27 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 and a Europe closer to citizens (Art. 5 CPR). This objective function is so broad that it would be hard to imagine which specific policy objective could *not* be subsumed under these headlines. The problem is exacerbated by a lack of transparency in how trade-offs are to be assessed. The High-Level Group report has identified this proliferation of objectives to

cause an excess complexity as a serious obstacle for monitoring and evaluation (European Commission, 2024). Multiple objectives thus immunise 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. This issue's relevance is confirmed by evaluators. Asatryan et al. (2024) run a comprehensive survey among authors of CP evaluations reported in the Cohesion Open Data Platform, and find that more than 60% of respondents regard unclear policy objectives as a bottleneck for the CP evaluation system.

#### **4.2. Heterogeneous evaluation culture**

Evaluation culture has always been dissimilar across MS. For the EU-15, Bachtler and Wren (2006) describe that the evaluation of regional policies had different traditions across MS. From case studies, 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 some administrations account more for how much was spent rather than to assess targets, let alone the causal impact. A (sole) focus on the absorption of CP funds is, however, strongly challenged by a broad literature, pointing also to the importance of administrative capacity to enhance the compliant and effective use of funds (Bachtrögler-Unger, 2025; Mendez & Bachtler, 2024).

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 with the money financed by the own country's taxpayers (Deaton, 2015).

#### **4.3. Lack of capacities and expertise**

A lack of evaluation capacities and expertise remains a significant issue in regions and MS according to current observations and judgments (Cerqua & Pellegrini, 2023; Naldini, 2018). Pellegrin et al. (2020) summarise insights from European Commission's Helpdesk assessment that two-thirds of MS impact evaluations from the 2014–20 period reviewed in 2018 had major quality deficiencies; in 2019 the situation had improved, but the Commission still rated the average quality of evaluations at 2.5 out of a maximum of 4 points, indicating a mediocre average quality.

The use of more advanced methods (e.g., theory-based and counterfactual methods) requires substantial expertise, constant training and investment into methodological knowledge, as well as data infrastructure. The required skills comprise not only methods but also the ability to draft 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). 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 MS with lower levels of human capital, and where universities and research institutes are not well-integrated into the international academic communities. In principle, an open European market for evaluations could compensate for this handicap of smaller and poorer countries. However, this is far from being the European reality today. MS's CP evaluation markets are largely separated and closed with almost full absence of international collaboration. Asatryan et al. (2025) show in a meta-study of CP evaluations that evaluators only rarely collaborate across borders.

As described in Section 2.3, the 2021–27 CPR has not only dropped explicit obligations to strengthen evaluation capacities in the MS, but also the European Commission's obligation to give MS guidance on how to carry out evaluations. Equally, it no longer obliges MS explicitly to provide the necessary resources for evaluations. This streamlining of the regulation may hinder further improvements and implies the risk that some MS 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 CPR). This rule is helpful but does not preclude that stakeholder interests have an impact on evaluation outcomes (Bachtler & Wren, 2006). These interests are a logical consequence from the above-described evaluation functions (learning, accountability and advocacy) which stand in a mutual tension. The learning function requires unbiased and impartial guidance from the evaluation. On accountability, the Commission, MS and their managing authorities share the interest to demonstrate to the national and European public a successful use of EU funds. 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 (Naldini, 2018; Pellegrin et al., 2020, p. 58).

Evaluators from consultancies, research institutes and freelancers are confronted with these motives. Although reputation concerns point in the direction of unbiased reports, these service providers will have a business interest to satisfy their customers' expectations 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. However, due to a strong national segmentation of CP evaluation markets (Asatryan et al., 2025), most evaluators provide their services to the authorities in their home country only. Hence, impartiality not only suffers from a non-diversified customer structure but also potentially from a home bias. It is known from other contexts, such as internationally operating rating agencies, that experts tend to give better assessments to their home country (Fuchs & Gehring, 2017). Asatryan et al. (2024) also show from their evaluator survey that 29% 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 therefore appears to not be 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

Research shows that very few evaluations make use of methods for causal inference, thereby lagging behind what is practiced in other fields, with development policy evaluations being among the most advanced. An analysis by Pellegrin et al. (2020, p. 49) for almost 1400 MS-level CP evaluations from the last programming period classifies 48% as qualitative and only 2.8% 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 generally is a lack of understanding how and through which specific mechanism a budgetary intervention contributes to a policy objective (Begg et al., 2023).

Another shortcoming is a lack of transparency in methodological limitations in the presentation of the 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 at both MS 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 the modelling results depend crucially on model assumptions. If implicit beneficial assumptions, for example, about a good project selection, 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 with the model simulations (Fratesi, 2024).

There is also a lack of clarity on the time dimension. If the method and the data are unable to decide whether an effect is longer lasting this needs to be clarified. Fratesi (2024) 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 stems from the narrow focus of MS evaluations on their own national or regional programmes. Because these evaluations usually do not include cross-country (and rarely cross-regional) comparisons, they cannot account for the impact of uniform national factors that do not vary within one country (or region).<sup>3</sup> Hence, by construction, MS analyses are blind for the significant role of national, and often also regional, bottlenecks. This is a serious deficiency, since these factors have been identified to crucially explain the success or failure of CP in influential reports such as the recent High-Level Group report on CP (European Commission, 2024) and the academic literature (e.g., regulation, lack of human capital, low administrative capacity, deficient institutions, etc.). For instance, Becker et al. (2013) show that the growth effects of CP are larger in regions with good institutional quality. Rodríguez-Pose and Garcilazo (2015) find that the quality of government positively influences the returns to CP investments, especially above a certain threshold of public investment (see also Becker et al., 2012, pointing to a maximum amount of effective funding in a region). Rodríguez-Pose and Ketterer (2020) provide evidence that improvements in government quality in disadvantaged regions can improve economic performance. Accordingly, together with human capital, institutional quality and administrative capacity are found to be central elements of a region's absorptive capacity as well as crucial for the success of CP implementation (Bachtrögl-Unger, 2025).

The lack of cross-country collaboration of evaluators lowers the evaluators' awareness of the role of differing national factors for the success of a European policy intervention. The methodological limits of merely national assessments also interact with evaluator incentives (see Section 4.4) not to lose favour with national authorities through a too explicit critique of deficiencies for which the national government bears responsibility.

#### **4.6. Lack of differentiated and unbiased aggregation of evaluation results**

The Cohesion Open Data Platform currently provides access to more than 2500 MS evaluations from the last two programming periods. This sheer number 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 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).<sup>4</sup> This central role of the Commission in this aggregation exercise is not without risks. As argued above, this institution is a stakeholder with a particular institutional 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 programmes. This is an overwhelming finding of relevant reports and studies, including by the European Court of Auditors (2019). This holds not only for MS but also for the Commission. For none of the operational programmes that had failed to meet their milestones did 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% of the evaluators surveyed by Asatryan et al. (2025) regard a lack of decision impact as a problem (and much more important than issues such as a lack of resources or impartiality).

Ex-ante evaluations can help in an early stage when budget decisions are taken, but also when a programme is implemented. However, contrary to the past period, the obligation for MS to conduct ex-ante evaluations has been removed from the CPR for the period 2014–20. 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 policymaking (European Court of Auditors, 2019; Naldini, 2018; Pellegrin et al., 2020).

#### **4.8. Evaluation timing not synchronised with decision-making**

Some of the revised evaluation obligations and deadlines for the 2021–27 period (see Section 2.4) may be counterproductive as they are ill-timed. For example, MS have to carry out their impact evaluations of 2014–20 programmes by June 2029 and the Commission is obliged to provide retrospective evaluations for each fund by the end of 2031. Both types of ex-post evaluations come too late to inform the Commission's preparation of the post-2027 legislative proposal and its subsequent negotiation with the EP and Council (European Court of Auditors, 2023). They even come too late to inform programme managers when they start to implement new programmes from 2028 onwards.

Obviously, timing decisions involve a trade-off. Data requirements and the particular interest in longer term effects (see Section 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 Commission to provide mid-term evaluations of each fund by the end of 2024 is to be welcomed as a positive novelty (European Court of Auditors, 2019; Naldini, 2018) 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 (2021–23). In reality, even less data will be available since CP programmes had a slow start in the 2021–27 period (Bachtrögler-Unger, 2025). Therefore, the 2024 mid-term evaluation of the 2021–27 CP programmes may have little to analyse.

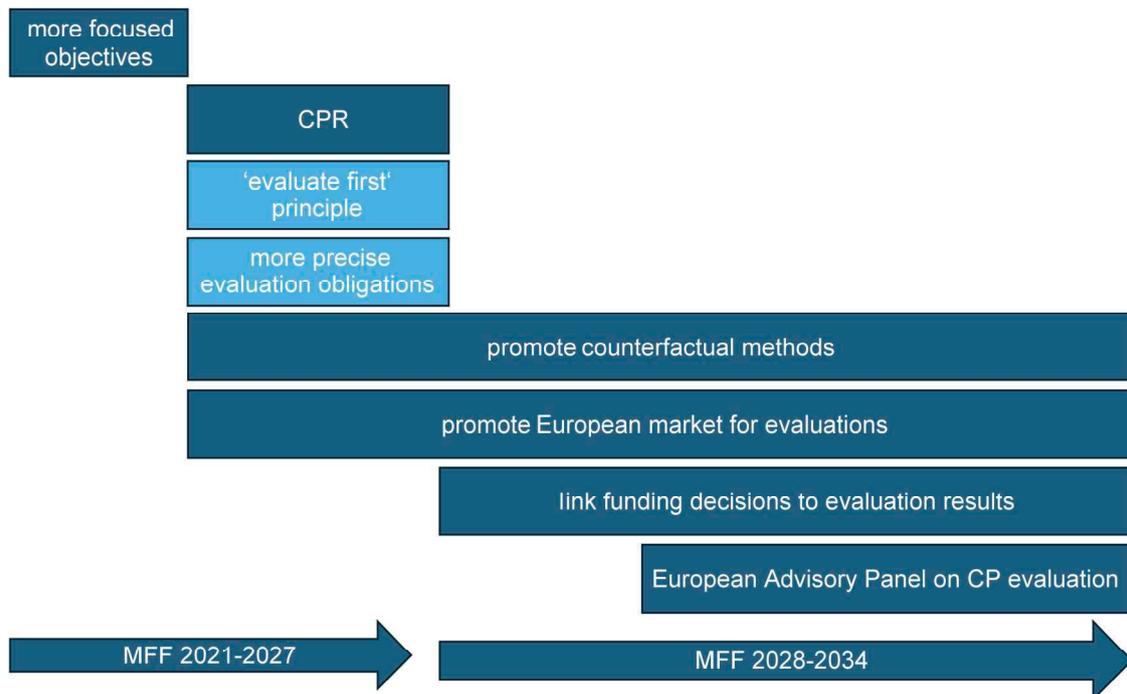
### **5. Policy recommendations**

The preceding gap analysis has revealed that a lot can be done to provide CP with a more 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. In general, genuinely independent and critical evaluations should be incentivised such that managing authorities do not perceive them as a threat but as a tool to increase the effectiveness of CP. We think that our following recommendations below can facilitate a transition towards a good equilibrium characterised by high expertise, more stringent standards, impartiality, credibility and genuine policy learning.

A roadmap for preparing the implementation of these recommendations is proposed in Figure 1. It implies a prioritisation of the recommended reforms in terms of timing, taking into account what could be implemented in the current programming period and in the preparation phase of the Multiannual Financial Framework (MFF) 2028–34 and thereafter. For example, a more focused set of objectives, more precise evaluation obligations and an 'evaluate first' principle should ideally be considered in the CPR for the next MFF.

#### **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 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. Reform proposals, such as the suggestion to refocus CP on the convergence of the poorest regions



**Figure 1.** Roadmap for implementation.

(Fuest, 2025), align with the need for a clearer division of labour among EU policies. While a comprehensive reform along these lines may prove challenging to attain, it is crucial for upcoming negotiations to recognise a fundamental trade-off: the proliferation of policy objectives for CP 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 MS have to provide sufficient resources to their evaluation process. Administrative capacities (potentially bundled across regions) are expected to be crucial not only for policy implementation but also for coordinating and organising evaluations, which are carried out by independent bodies. Equally, the 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 emphasise the need for more transparency on methodological limits of the methods used for the evaluation.

Finally, the 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 programmes for both evaluators and staff at managing authorities dealing with evaluations. Courses on causal evaluation methods have become part of standard curricula in economics programmes 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 MS evaluation capacities and sanctions that are imposed if a MS fails to reach the minimum standards could be introduced.

### **5.3. Introduce an ‘evaluate first’ requirement when preparing or updating programmes**

Like the Commission, MS should apply an ‘evaluate first’ principle when preparing or updating programmes. No longer requiring MS 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 reintroducing a focused ex-ante evaluation obligation which concentrates on particularly important programmes. The selection could be decided based on budget size or other more content-oriented criteria that identify innovative approaches with a particular need for thorough impact reflections and programme design. This could pave the way for fewer but more careful evaluations. Whatever criteria are used to choose which programmes to evaluate ex-ante, they should not leave room for interpretation 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 respect, the next CPR should reconsider its deadlines for retrospective impact analyses to inform policymakers and programmers in a timely manner before programming the next period’s policy.

### **5.4. Promote the use of counterfactual methods**

Counterfactual policy evaluations require not only methodological skills and rich data availability, but also experimental or quasi-experimental set-ups that allow one to compare a treated group to an appropriate control group. To ensure that, the use of experimental and quasi-experimental evaluation approaches needs to be planned already along with designing CP programmes. We see three possibilities, which differ in their feasibility.

From a methodological perspective, actual experiments would be the ideal approach. This ‘gold standard’ requires randomising funding (its levels, types, design features, etc.) across eligible projects and regions. In development policy, this has become common practice. Together with the World Bank, the Commission has recently engaged with fostering the implementation of randomised control trials in CP evaluation. Here, implementing agencies often collaborate with evaluators or research institutes in a way that a specific element of the development project is randomised (Gibson et al., 2017; Pomeranz, 2017). Of course, randomisation may raise ethical, legal and political problems that relate to unequal treatment. Preparations into this direction could benefit from an exchange with experimental researchers who have applied field experiments in development economics for a long time and who have been confronted with similar constraints.

Beyond randomised evaluations, quasi-experimental evaluation designs pose fewer ethical and legal problems since they exploit variation arising naturally from project selection and policy implementation. A prominent example is the RDD. A series of studies (including Becker et al., 2010, 2018; Pellegrini et al., 2013; Ferrara et al., 2017; Lang et al., 2023) have used RDDs to estimate CP effects based on the eligibility threshold for less developed (former Objective 1) regions at 75% of the EU average. The scope for such (ex-post) evaluation would expand if the EU were to add further discontinuities in their policy design. For instance, funding allocation could be based on rankings of eligible firms, projects or regions by means of a points system where funding is approved up to a certain cut-off value. This could then be exploited in RDDs comparing firms, projects or regions just above and below this threshold. Implementing such designs, however, will come with an additional administrative burden and cost.

Another quasi-experimental method that could be well suited for CP evaluations is the DID design. Here, causal identification is based on comparing the development of treated and untreated units (or a well-designed control group) that would have developed equally in the absence of the treatment. To implement this strategy, first, data collection for untreated units (projects, beneficiaries, geographical entities) and in the pre-treatment period is crucial to be able to test this parallel trends assumption. Second, evaluators would benefit from the EU randomising project’s start dates across eligible firms or regions. Such randomisation strategies would circumvent the problem of unequal treatment by only randomising the *timing* of the treatment rather than its actual implementation. Such designs could be studied with (staggered) DID models that look at an event window before and after the start. Whether it will be politically feasible to implement such randomised elements in this context is an open question, and some pushback is to be expected.

### **5.5. Implement measures to stimulate a European market for CP evaluations**

National evaluators possess a commendable understanding of their home country, including its language, institutions and political system. 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 and a potential home bias in judgment. 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. Currently, the trade-off for MS evaluations leans heavily towards an almost entirely closed evaluation market (Asatryan et al., 2025). A more open market would support independence and also could quickly import evaluator expertise into those MS which suffer from a shortage of advanced evaluator expertise.

We recommend starting a broad initiative to open the borders of the CP evaluation service market with appropriate incentives. These incentives could entail the use of an international team composition as one criterion in the calls for tenders for national evaluation contracts for programmes above a certain budget threshold. A second proposal is a peer-review stage for evaluations with the requirement that some of the reviewers must come from other MS. A cross-country approach to quality checks and peer review could help to overcome national biases and to identify deficiencies in national evaluations.

### **5.6. Explicitly link funding decisions at programme and policy level to evaluation results**

Ultimately, evaluation results should also be reflected in the funding allocation for and within programmes. This is a complex undertaking. We propose referring to the ‘evaluate first’ principle set out above. We suggest that MS must back up any programme amendments that involve a significant reallocation of funding up by an evaluation. Similarly, any proposal for a programme would need to be accompanied by an ex-ante evaluation or impact evaluation of the predecessor programme. The Commission (which must approve programmes and their amendments) would act as a goalkeeper to ensure that evaluation findings are properly considered in the programme (re-)design and related funding (re-)allocations. In that respect, it should also be given the right to unilaterally decide on funding (re-)allocations based on these evaluation findings.

### **5.7. Establish a standing European Advisory Panel on CP Evaluation**

Finally, we recommend setting up a ‘European Advisory Panel on CP Evaluation’ (EAP-EVAL). As argued above, the Commission has specific institutional interests that may hinder fully neutral CP assessments. As an independent and non-political body, the main role of the EAP-EVAL would be to advise the Commission and the managing authorities in MS in their evaluation activities.

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 organise 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). 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 programmes to preclude any risk of a conflict of interest. Such (and further) measures should strive to maximise the independence and impartiality of the EAP-EVAL, while acknowledging that there is no guarantee for institutions involved in political processes to ever achieve perfect independence and impartiality. Political economy research has repeatedly shown how international institutions that are designed to be independent

become subject to political capture and bias (e.g., Dreher & Lang, 2019). Protecting the impartiality of an EAP-EVAL will thus be a constant challenge.

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), which 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. IPCC authors are selected based on their expertise and with the intention to cover a diversity of socio-economic views and backgrounds, as well as geographical and gender balance. The 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 MS.

## 6. Conclusions

In sum, the EU has developed a comprehensive and transparent evaluation framework for CP. Nevertheless, significant gaps remain in terms of methodological rigour, evaluator independence and the actual use of evaluation findings in policy decisions. The contrast between more optimistic official assessments and more differentiated academic evaluations highlights the need for greater impartiality, clarity in objectives, transparency about the limitations of the respective evaluation methods, as well as a better integration of evaluations into the policy cycle. Excessively broad policy objectives, national biases and fragmented evaluation markets undermine the credibility and effectiveness of CP evaluations.

To address these challenges, this article proposes a comprehensive reform agenda aimed at fostering a more focused, independent and methodologically advanced evaluation system. Key recommendations include reintroducing ex-ante evaluations and obligations for MS to provide sufficient resources for evaluations, promoting counterfactual methods, establishing a European advisory panel and opening national evaluation markets to cross-border expertise. Ultimately, improving the evaluation system is essential not only for learning and accountability but also for ensuring that future CP reforms are grounded in credible evidence and contribute to policy effectiveness. Due to the numerous new and urgent demands on the EU budget, ranging from defence to migration and transformation, a more performance- and evaluation-based approach to cohesion is essential to justify the ongoing allocation of a significant budget share to this policy.

## Notes

1. This analysis focuses on CP financed by the core budget. For an analysis of how the performance orientation and evaluation differ for the Recovery and Resilience Facility, see European Court of Auditors (2023) and Corti et al. (2025).
2. In practice, CP evaluations have mostly covered the effectiveness and efficiency criteria (Pellegrin et al., 2020).
3. Still, if there is regional variation, national studies can control for the influence of institutional constraints.
4. For example, the Cohesion Reports sometimes arguably give an overly optimistic account of the impact of CP. For example, chapter 9 of the Eighth Cohesion Report (European Commission, 2022) summarises evaluations results on the impact of CP. Essentially all evaluation results included in this chapter have a positive message. 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.

## Acknowledgements

The authors thank the two anonymous reviewers for excellent comments and suggestions. Francesco Corti conducted this research while affiliated with the Centre for European Policy Studies (CEPS). Since then, he has joined the European Commission. His views as a co-author of this study cannot be construed as representing the position of the European Commission. Similarly, Martin Weber writes in a personal capacity and his views as a co-author of this study cannot be construed as representing the position of the European Court of Auditors.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This research was supported by the German Federal Ministry of Finance [grant number fe 1/22].

## ORCID

Friedrich Heinemann  <http://orcid.org/0000-0003-3501-0199>  
 Zareh Asatryan  <http://orcid.org/0009-0004-1500-2399>  
 Julia Bachtrögler-Unger  <http://orcid.org/0000-0002-6312-3492>  
 Carlo Birkholz  <http://orcid.org/0009-0002-5132-3320>  
 Francesco Corti  <http://orcid.org/0000-0003-1984-9496>  
 Ugo Fratesi  <http://orcid.org/0000-0002-0755-460X>  
 Clemens Fuest  <http://orcid.org/0009-0008-7698-8648>  
 Valentin Lang  <http://orcid.org/0000-0002-5719-418X>

## References

- Asatryan, Z., Birkholz, C., & Heinemann, F. (2024). *Evidence-based policy or beauty contest? An LLM-based meta-analysis of EU Cohesion Policy evaluations* (ZEW Discussion Paper No. 24-037).
- Asatryan, Z., Birkholz, C., & Heinemann, F. (2025). Evidence-based policy or beauty contest? An LLM-based meta-analysis of EU Cohesion Policy evaluations. *International Tax and Public Finance*, 32(2), 625–655. <https://doi.org/10.1007/s10797-024-09875-4>
- Bachtler, J., & Wren, C. (2006). Evaluation of European Union Cohesion Policy: Research questions and policy challenges. *Regional Studies*, 40(2), 143–153. <https://doi.org/10.1080/00343400600600454>
- Bachtrögler, J., Fratesi, U., & Perucca, G. (2020). The influence of the local context on the implementation and impact of EU Cohesion Policy. *Regional Studies*, 54(1), 21–34. <https://doi.org/10.1080/00343404.2018.1551615>
- Bachtrögler-Unger, J. (2025). The role of administrative capacity for an effective implementation of EU Cohesion Policy. In Z. Asatryan, L. P. Feld, & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 187–222). Mohr Siebeck.
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (2010). Going NUTS: The effect of EU structural funds on regional performance. *Journal of Public Economics*, 94(9–10), 578–590. <https://doi.org/10.1016/j.jpubeco.2010.06.006>
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (2012). Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review*, 56(4), 648–668. <https://doi.org/10.1016/j.euroecorev.2012.03.001>
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (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. <https://doi.org/10.1257/pol.5.4.29>
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (2018). Effects of EU regional policy: 1989–2013. *Regional Science and Urban Economics*, 69, 143–152. <https://doi.org/10.1016/j.regsciurbeco.2017.12.001>
- Begg, I., Corti, F., Liscai, A., Kiss-Gálfalvi, T., Grossi, T., & Rayner, L. (2023). *Social tracking methodology for the EU budget*, European Parliament, study requested by the BUDG Committee (PE 742.788 – January). [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)
- Cerqua, A., & Pellegrini, G. (2023). I will survive! The impact of place-based policies when public transfers fade out. *Regional Studies*, 57(8), 1605–1618. <https://doi.org/10.1080/00343404.2022.2136370>
- Corti, F., Pedralli, M., & Pancotti, C. (2025). The recovery and resilience facility: Key innovations and the interplay with Cohesion Policy. In Z. Asatryan, L. P. Feld, & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 247–274). Mohr Siebeck.
- Deaton, A. (2015). *Weak states, poor countries* [Blog]. Social Europe. <https://www.socialeurope.eu/weak-states-poor-countries>
- Di Caro, P., & Fratesi, U. (2022). One policy, different effects: Estimating the region-specific impacts of EU Cohesion Policy. *Journal of Regional Science*, 62(1), 307–330. <https://doi.org/10.1111/jors.12566>
- Downes, R., Moretti, D., & Nicol, S. (2017). Budgeting and performance in the European Union. *OECD Journal on Budgeting*, 17(1), 1–60. <https://doi.org/10.1787/budget-17-5jfnx7fj38r2>
- Dreher, A., & Lang, V. F. (2019). The political economy of international organizations. In R. D. Congleton, B. Grofman, S. Voigt, A. Dreher, & V. F. Lang (Eds.), *The Oxford handbook of public choice* (Vol. 2, pp. 606–652). Oxford University Press.
- 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](https://ec.europa.eu/regional_policy/sources/evaluation/2014/wd_2014_en.pdf)

- European Commission. (2014). *The programming period 2014–2020: Guidance document on monitoring and evaluation – European Regional Development Fund and Cohesion Fund*.
- 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 No. 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.
- European Commission. (2024). *Forging a sustainable future together – Cohesion for a competitive and inclusive Europe: Report of the High-Level Group on the Future of Cohesion Policy* (February). Publications Office of the European Union, Directorate-General for Regional and Urban Policy. <https://op.europa.eu/en/publication-detail/-/publication/c6e97287-cee3-11ee-b9d9-01aa75ed71a1/language-en>
- European Court of Auditors. (2019). *Delivering performance in cohesion* (Briefing Paper, June). European Union.
- European Court of Auditors. (2021). *Special report 24/2021: Performance-based financing in Cohesion Policy: Worthy ambitions, but obstacles remained in the 2014–2020 period*. [https://www.eca.europa.eu/EN/publications/SR21\\_24](https://www.eca.europa.eu/EN/publications/SR21_24)
- European Court of Auditors. (2023). *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*.
- European Union. (2024). *Forging a sustainable future together: Cohesion for a competitive and inclusive Europe: Report of the High-Level Group on the Future of Cohesion Policy*. Directorate-General for Regional and Urban Policy, European Commission.
- Ferrara, A. R., McCann, P., Pellegrini, G., Stelder, D., & Terribile, F. (2017). Assessing the impacts of Cohesion Policy on EU regions: A non-parametric analysis on interventions promoting research and innovation and transport accessibility. *Papers in Regional Science*, 96(4), 817–842. <https://doi.org/10.1111/pirs.12234>
- Fratesi, U. (2024). *Regional policy, theory and practice*. Routledge.
- Fratesi, U. (2025). Constraining and enabling factors of a successful EU regional policy in Europe. In Z. Asatryan, L. P. Feld, & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 187–222). Mohr Siebeck.
- Fuchs, A., & Gehring, K. (2017). The home bias in Sovereign ratings. *Journal of the European Economic Association*, 15(6), 1386–1423. <https://doi.org/10.1093/jeea/jvx009>
- Fuest, C. (2025). Fundamental considerations for a more rational EU Cohesion Policy in the future. In Z. Asatryan, L. P. Feld, & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 13–30). Mohr Siebeck.
- Gesellschaft für Evaluation e.V. (Ed.) (2023). *DeGEval-standards. Standards für Evaluation*.
- Gibson, M., Sautmann, A., Feeney, L., & Walsh, C. (2017). *Introduction to randomized evaluations*. J-Pal, Abdul Latif Jameel Poverty Action Lab. <https://www.povertyactionlab.org/resource/introduction-randomized-evaluations>
- Heinemann, F., Hagen, T., Mohl, P., Osterloh, S., & Sellenthin, M. O. (2009). *Die Zukunft der EU-Strukturpolitik*. Nomos.
- Intergovernmental Panel on Climate Change (IPCC). (2024). *Structure of the IPCC*. <https://www.ipcc.ch/about/structure/>, updated on 1 January 2024
- Lang, V. (2025). The distributional effects of place-based policies in the EU. In Z. Asatryan, L. P. Feld, & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 145–164). Mohr Siebeck.
- Lang, V., Redeker, N., & Bischof, D. (2023). *Place-based policies and inequality within regions*. Center for Open Science.
- Leino-Sandberg, P. (2025). Cohesion Policy and the principle of subsidiarity – A legal analysis. In Z. Asatryan, L. P. Feld & F. Heinemann (Eds.), *The future of Cohesion Policy* (pp. 75–122). Mohr Siebeck.
- Mendez, C., & Bachtler, J. (2024). The quality of government and administrative performance: Explaining Cohesion Policy compliance, absorption and achievements across EU regions. *Regional Studies*, 58(4), 690–703. <https://doi.org/10.1080/00343404.2022.2083593>
- 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. <https://doi.org/10.1177/1356389018804261>
- Organisation for Economic Co-operation and Development (OECD). (2021). *Applying evaluation criteria thoughtfully*. OECD Publ.
- Organisation for Economic Co-operation and Development (OECD). (2022). *Recommendation of the council on public policy evaluation* (OECD/LEGAL/0478).
- Pellegrin, J., Colnot, L., & Pedralli, M. (2020). *The role of evaluation in Cohesion Policy, study requested by the REGI committee*. European Parliament.
- Pellegrini, G., Terribile, F., Tarola, O., Muccigrosso, T., & Busillo, F. (2013). Measuring the effects of European regional policy on economic growth: A regression discontinuity approach. *Papers in Regional Science*, 92(1), 217–234. <https://doi.org/10.1111/j.1435-5957.2012.00459.x>
- Pomeranz, D. (2017). Impact evaluation methods in public economics. *Public Finance Review*, 45(1), 10–43. <https://doi.org/10.1177/1091142115614392>
- Robinson, M., & Last, D. (2009). *A basic model of performance-based budgeting*. IMF technical notes and manuals. International Monetary Fund (IMF).

- Rodríguez-Pose, A., & Garcilazo, E. (2015). Quality of government and the return of investment: Examining the impact of cohesion expenditure in European regions. *Regional Studies*, 49(8), 1274–1290.
- Rodríguez-Pose, A., & Ketterer, T. (2020). Institutional change and the development of lagging regions in Europe. *Regional Studies*, 54(7), 974–986. <https://doi.org/10.1080/00343404.2019.1608356>
- von Ehrlich, M. (2024). *The importance of EU Cohesion Policy for economic growth and convergence*. Contribution to BMF Expert Network.
- von Ehrlich, M., & Overman, H. G. (2020). Place-based policies and spatial disparities across European cities. *Journal of Economic Perspectives*, 34(3), 128–149. <https://doi.org/10.1257/jep.34.3.128>