



Peer Review Experimental methods for impact evaluation of social inclusion policies

28-29 November 2023, Madrid, Spain
Summary report

Written by ICF
October 2023



EUROPEAN COMMISSION

Directorate-General for Employment, Social Affairs and Inclusion

Directorate D — Social Rights and Inclusion

Unit D1 — Social Policies, Child Guarantee, SPC

Contact: Gilberto PELOSI

E-mail: empl.d1@ec.europa.eu

European Commission

B-1049 Brussels

Manuscript completed in December 2023

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. More information on the European Union is available on the Internet (<http://www.europa.eu>).

© European Union, 2023



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

Peer Review
Experimental methods for
impact evaluation of
social inclusion policies

Table of Contents

1	Introduction.....	1
1.1	Purpose of the Peer Review	1
1.2	The EU policy context.....	1
1.3	Impact evaluations.....	1
2	The Spanish Policy Lab: an illustration of the potential of RCTs to evaluate social policies	2
2.1	The Spanish social and policy context.....	2
2.2	The Policy Lab and 34 pilot projects	3
2.3	Design and implementation of RCTs across pilot projects.....	5
2.4	Early lessons learnt and future of the Policy Lab	6
3	RCT and quasi-experimental evaluation cycle.....	6
3.1	Preconditions	6
3.2	Design phase.....	9
3.3	Implementation phase	12
3.4	Follow-up and scaling-up.....	12
4	Key policy messages	13

1 Introduction

1.1 Purpose of the Peer Review

The Ministry of Inclusion, Social Security, and Migration in Spain hosted on 28-29 November 2023 a Peer Review on the potential of using experimental and quasi-experimental methods to evaluate social inclusion policies. In addition to government representatives from Spain, the Peer Review brought together government representatives from Belgium, Bulgaria, Czechia, Malta, Portugal and Slovenia, as well as representatives of the European Commission and independent experts.

The Peer Review focused on the Policy Lab, or Laboratory of Inclusion and Activation Strategies, an initiative in Spain aiming at supporting evidence-based social policies. The Policy Lab accompanies the implementation of the Spanish national minimum income scheme launched in 2020. In order to test new complementary interventions, a range of 34 educational, employment, and psychosocial pilot programmes involving regional and local governments and the third-sector organisations were launched. These pilot programmes are funded by [NextGenerationEU](#). All pilot programmes are being evaluated through Randomised Control Trials (RCT), an experimental design used in research that utilises a treatment group and a control group to assess the impact of a particular intervention. The Policy Lab, which oversees the design, implementation, and evaluation of the pilot measures, brings together a range of scientific partners to ensure that evaluations follow the highest scientific standards.

The pilot programmes are currently in their evaluation phase. The objective of the Peer Review was to share the perspective of the Host Country on encountered obstacles and lessons learned as part of the design and implementation phase, as well as encourage debate on the use of experimental and quasi-experimental methods to evaluate social inclusion policies and get insights from peer countries.

1.2 The EU policy context

The [European Pillar of Social Rights](#) aims to build a fairer and more inclusive EU. In line with principle 14 of the European Pillar of Social Rights, income support schemes, which provide an essential safety net and help lift people out of poverty, should go hand-in-hand with measures aiming at active inclusion in society and incentives to (re)integrate into the labour market for those able to work.

The 2023 [Council Recommendation on adequate minimum income](#) calls on Member States to modernise their minimum income schemes by reaffirming the active inclusion approach and focusing on adequacy, coverage and take-up of benefits, as well access to inclusive labour markets, essential services and individualised support. The Recommendation also focuses on strengthening governance, monitoring and reporting mechanisms for income support policies and related labour market activation measures, considering the role of stakeholders involved, such as regional and local authorities, social partners, civil society organisations and social economy actors, to pave the way for a more holistic approach to social inclusion policies.

From the standpoint of the European Commission, supporting policymaking grounded in empirical evidence is of paramount importance and social policy experimentation via EU funds is a key strategy in this regard.

1.3 Impact evaluations

Rigorous evaluation of public policies is a critical component of good governance. It is essential for understanding which policies are most effective, promoting accountability,

effectiveness, and efficiency, enhancing transparency, and building public confidence, facilitating continuous adaptation and improvements in policy design.

Counterfactual impact evaluation aims to assess the causal effects and effectiveness of an intervention by comparing the outcomes for participants who received a “treatment” with a counterfactual group that did not – allowing to understand what would have occurred in the absence of an intervention. The main challenge of this approach is to find a way to perform the comparison in such a way that the distribution of outcome for the control group serves as a good counterfactual for the distribution of outcome for the treated group in the absence of treatment.

In the last years, experimental sciences have understood that, if well-designed and implemented, the most rigorous method to infer causality from a policy to an outcome is by performing Randomized Control Trials (RCTs). RCTs are experimental approaches where eligible participants are randomly assigned to a treatment group or to a control group, the latter receiving no intervention or business as usual. As participants from both groups are statistically comparable and the only difference between these two groups is the exposure to the intervention, with a large enough sample, significant variation in outcomes between groups can be directly attributed to the policy with strong evidence.

When RCTs are too costly, difficult to implement or not feasible, quasi-experimental designs can be an option to artificially replicate the RCT paradigm. Quasi-experimental approaches infer the counterfactual by examining a comparison group that does not benefit from the policy but shares similar characteristics with participants before the intervention begins.

2 The Spanish Policy Lab: an illustration of the potential of RCTs to evaluate social policies

This section provides an introduction to the Policy Lab; more information can be found in the Host Country paper prepared for this Peer Review¹.

2.1 The Spanish social and policy context

Spain is characterised by a decentralised institutional system, where competences are shared across multiple administrative levels. The Ministry of Inclusion, Social Security, and Migrations in Spain (MISSM) is responsible for social security and pensions, as well as the coordination of social inclusion policies. The implementation of social inclusion policies falls under the primary responsibility of the 17 autonomous regional governments, which often delegate this implementation on the local municipal governments. In Spain, there is long tradition in dealing and collaborating with third-sector organisations in the field of social inclusion policies.

The levels of poverty and inequality in Spain increased in the context of the COVID-19 pandemic, increasing the risks of social exclusion and calling for additional policy interventions. While social spending in Spain increased significantly from 25% of GDP in 2010 to 30% in 2020, it remained primarily concentrated on unemployment, pensions and healthcare, while spending on social exclusion was only 1.0% in 2019, much below the EU average.

Before 2020, Spain was one of the few EU countries without a national minimum income scheme, as only regional level schemes were in place which varied in size and scope.

¹ García Hernández, A., Martínez-Bravo, M. (2023). [The Inclusion Policy Lab: Towards a New Era of Policy-Making in Spain](#). Host country discussion paper for Peer Review on Experimental methods for impact evaluation of social inclusion policies.

Recommendations put forth by the EU emphasised the need to improve the coverage of minimum income policies across the territory.

Against this background, the Spanish Minimum Income Scheme (MIS) was introduced in May 2020. The MIS benefit is managed by the Spanish National Social Security Institute (*Instituto Nacional de la Seguridad Social*). This benefit focuses on households and the cash transfer programme guarantees a monthly income dependant on household composition from around EUR 560 for a household composed by a single individual to EUR 1 368 for a single-parent household with four or more children. Regional governments can complement the MIS with their pre-existing schemes. By December 2022, the MIS had reached close to 284 000 households and the annual cost of the scheme was of EUR 1 919 million. However, non-take-up rates remain high, at 57%. To facilitate take-up, the MISSM has created a registry of MIS mediators (third-sector organisations), who can support potential beneficiaries in complying with administrative requirements.

To go beyond a simple cash transfer, the law on the MIS introduced a companion policy called 'social inclusion pathways', focusing on the development of policies to foster social inclusion. These pathways aim to address the root causes of poverty and social exclusion. In order to assess which measures are most effective to support social inclusion, it was determined from the onset that the social inclusion pathways would be evaluated in the most rigorous way, via RCTs. The Policy Lab is the first instance of experimenting with RCTs on a substantial scale in Spain.

A challenge for Spain, common to many other European countries, is that the evaluation of public policies is fragmented and heterogeneous, with little emphasis on quasi-experimental and experimental approaches. Indeed, policy evaluation across different policy fields is carried out by a wide range of institutions and mostly focused on evaluation of implementation, rather than evaluation of impacts.

In the context of the launch of the Policy Lab, two important milestones must be noted, paving the way for the institutionalisation of policy evaluation in Spain. First, the Recovery, Transformation and Resilience Plan adopted in 2021 led to creation of a unit on public spending evaluation at the [Independent Authority for Fiscal Responsibility \(AIREF\)](#). In addition, the 2022 Law on the evaluation of public policies established a framework with general provisions for evaluating public policies, although without standard criteria.

2.2 The Policy Lab and 34 pilot projects

The Secretariat General for Inclusion Goals and Policies and Social Welfare, within the MISSM, is the institution in charge of the Policy Lab. The Policy Lab encompasses the design, implementation and evaluation of 34 innovative social inclusion pathways which aim at increasing the take-up of the MIS and complementing it with assistance in areas such as education, digital skills, employment, social support, and health to help promote social and economic inclusion.

The Secretariat General for Inclusion Goals and Policies and Social Welfare launched public consultations for regional and local governments and third-sector organisations to submit their proposals to be funded by NextGenerationEU Recovery Plan, including a requirement that the pilot pathways should be evaluated through RCTs. Two sets of pilot projects were selected through Royal Decrees: 16 in October 2021 and an additional 18 in May 2022.

The procedures used allowed to reach a broad geographical coverage, with pilot projects implemented across all Spanish regions benefiting from a national-level coordination and technical support, in line with institutional arrangements in the field of social inclusion policies. Specific mechanisms for coordination were created to support the Policy Lab including a monitoring committee as well as technical meetings.

The pilot projects targeted national minimum income beneficiaries and other disadvantaged groups. Their approaches varied, from comprehensive interventions to specific measures like improving digital skills and outreach efforts to encourage uptake of minimum income scheme.

Box 1. Overview of the social inclusion pathways within the Policy Lab

- **Objective:** launching 34 pilot projects on social inclusion in Spain accompanying the minimum income scheme and generating robust evidence on their effectiveness.
- **Coordinating organisation:** Spanish Ministry of Inclusion, Social Security, and Migrations (MISSM) – Secretariat General for Inclusion Goals and Policies and Social Welfare.
- **Implementing partners:** 16 regional administrations, 4 local administrations and 14 third-sector organisations.
- **Scientific partners:** Centre for Monetary and Financial Studies (CEMFI) and Abdul Latif Jameel Poverty Action Lab (J-PAL Europe).
- **Funding:** NextGenerationEU Recovery Plan (calls for proposals). Total investment EUR 212 million, covering both implementation and evaluation of measures.
- **Groups targeted:** Beneficiaries of the national and regional minimum income schemes and other disadvantaged groups, including ethnic minorities, homeless people, women who are victims of domestic violence or sexual exploitation, unemployed people, migrants and persons with disabilities.
- **Intervention areas:** education (88%), social support (71%), employment (65%), digital skills (56%), health (18%), housing (15%), caregiving (12%), MIS take-up (12%), energy poverty (9%), entrepreneurship (3%).
- **Geographical coverage:** all regions in Spain.
- **Dates and length:** pilot projects launched in October 2021 and May 2022; average duration of interventions of 8 months (from 1.5 to 18 months) in the period between October 2021 to November 2023.
- **Evaluation methods:** Randomised Control Trials.
- **Number of participants:** More than 100 000 direct participants.

Source: Adapted from Host country discussion paper for Peer Review.

The implementing partners of the pilot projects based on their knowledge of local level needs and target groups, crafted the proposals, bearing in mind the requirement to carry out RCTs. The implementing partners implemented the pilot interventions and were tasked to collect data for evaluation purposes.

The MISSM played a strong role in the design phase and supported implementing partners in the preparation of proposals. RCTs are resource-intensive in terms of scientific expertise and require involvement of researchers from the start. To ensure that the evaluations would meet the highest academic standards, the MISSM partnered with two institutions to provide scientific support, CEMFI (Centre for Monetary and Financial Studies) and J-PAL Europe (Abdul Latif Jameel Poverty Action Lab). CEMFI provided a team of 18 evaluators and J-PAL Europe assigned dedicated personnel working along the Ministry and CEMFI researchers. In addition, some implementing partners brought their own research teams.

2.3 Design and implementation of RCTs across pilot projects

During the initial stage of the preparation of proposals, significant efforts were devoted to providing training and bringing all stakeholders to a common understanding of the methodological approach, refine the proposals for interventions and designs of the evaluations. This involved, for example, limiting the focus of some of the pathways to specific interventions to facilitate the evaluation of their impact.

The initial level of awareness of RCTs among implementing partners was generally low. Third-sector organisations were experienced in data collection processes and evaluations but not with RCTs; similarly, regional and local authorities had little or no previous experience with experimental evaluation methods.

Each of the pilot projects within the Policy Lab had a different scope, target groups, sample size, etc. Implementing partners recruited participants for their pilots. To identify potential participants, implementing partners were able to draw from the national registry of beneficiaries of the MIS, owned and managed within the MISSM, as foreseen in the law introducing the minimum income scheme.

The sample size differed across projects. Attrition occurs when participants leave the sample, either by dropping out of the intervention or declining to respond to final surveys. Implementing partners were made aware that attrition should be anticipated to guarantee a sufficient final sample size. Attrition is a particular concern in interventions targeting individuals in hard-to-reach populations. In particular, those participants in the control group were less motivated to answer final surveys because they are not receiving the intervention. Their participation depends also on other factors such as engagement, follow-up, involvement of partners. Within each project, strategies were developed to engage with participants and to mitigate the risk of attrition, for instance by providing monetary incentives to survey respondents.

To help address ethical issues linked to the use of experimental methods, the MISSM established an Ethical Review Board, composed of academics. They scrutinised all pilot projects during their design phase and made recommendations for ethical safeguards, paying attention to aspects such as communication and transparency with participants, consent, and data protection issues.

Participants were asked to participate on a voluntary basis in each of the experiments and were required to sign an informed consent form. Potential participants were clearly informed that their decision did not condition their eligibility for minimum income schemes. After the randomisation phase, participants were informed about the type of support they would receive, without communicating explicitly if they were assigned to the treatment or control group. Although there were exceptions where assignment to treatment and control groups was communicated. The MISSM was responsible for the randomisation. The level of randomisation was set differently across projects, taking into account the sample size as well as specific characteristics of the project. A higher level of randomisation was used among others in projects where participants have a close interaction, to ensure they receive a similar form of treatment and avoid comparisons.

For each pilot project, a set of measurable outcomes was established. Most projects used both objective and subjective (well-being) indicators for participants. Objective indicators included administrative data from social security, incorporating participants' labour history and employment status, used with their informed consent for scientific purposes. Exchanging administrative data necessitated a formal agreement between the Policy Lab's Directorate and the Social Security Directorate within the MISSM.

2.4 Early lessons learnt and future of the Policy Lab

The introduction of the Policy Lab in Spain is a promising example of a top-down strategy. It seeks to promote capacity in experimental methods in social policy evaluation and foster partnerships, with active participation from the scientific community. The Spanish experience highlights the role of political will advocating for rigorous evaluation and fostering a shared appreciation of its value. It also demonstrates the substantial effort needed to develop the necessary knowledge base and partnerships.

The value of the Policy Lab lies in its model of governance, involving multiple stakeholders selected for their specific expertise, including implementing partners from the third-sector, local and regional government, and academic experts providing technical support at each stage of the evaluation. This approach requires, however, considerable time and human resources to ensure stakeholder preparedness and building strong partnerships and communication channels.

Although the use of NextGeneration funds constrained the timeframe for designing experiments, the diversity of pilot schemes under the Policy Lab highlights the need for careful design of measures and RCTs. This design should be based on scientific criteria, tailored to each intervention's specifics. The proposals for each intervention were carefully considered and designed separately by implementing partners in cooperation with the MISSM. Throughout the design and implementation phases, ethical concerns were minimised through the support of an Ethical Review Board.

The data collection phase for most projects is underway. The publication of evaluation results and key learnings from the most successful pilot project is expected in the first quarter of 2024. If results are inconclusive, further evaluation may be necessary to investigate the reasons for the lack of impact, and/or if longer-term effects can be observed. A thorough review of the results from the 34 pilots will guide future policymaking and potentially lead to scaling-up of some projects. The Spanish authorities have included in European Social Fund Plus (ESF+) Programmes the possibility to scale up pilot projects with proven impacts.

Regarding governance, the MISSM is planning how to move from the Policy Lab into a systematic and structured process of policy impact evaluation. This is based on the conviction that social inclusion policies and their evaluation should remain a priority, regardless of the political context.

3 RCT and quasi-experimental evaluation cycle

RCTs and quasi-experimental policy evaluations are powerful tools for assessing the effectiveness of policies and to identify causal links. However, there is an overall scarcity of established policy mechanisms for conducting robust and rigorous policy evaluations. (Quasi-)experimental policy designs are seldomly carried out in the EU countries, especially in the area of social policy. Nonetheless, the Peer Review discussions emphasised a growing awareness among EU countries of the importance of adopting more rigorous evaluations in social policy. The Peer Review participants delved into the key aspects of introducing, designing, carrying out and analysing the results of RCTs in social policy, as well as of scaling-up successful pilot interventions. They explored common challenges, brainstormed mitigation measures and identified success factors.

3.1 Preconditions

Rigorous evaluations are emerging only recently as a systematic practice in EU Member States, and they are overall still novel. In this situation, the lack of awareness, preparedness and competences concerning (quasi-)experimental policy evaluations among public administrators and implementing partners emerged as a widespread critical challenge in all

Peer Countries, hindering the implementation of such designs on a large scale in Europe. The limited tradition of RCTs in Europe is compounded by a degree of institutional resistance to experimental evaluations within public administrations. However, valuable lessons have been collected from discussing these challenges.

Need for new social policy measures

The presentations from peer countries highlighted the emergence of social vulnerabilities in Europe, due to the COVID-19 pandemic, the Russian war against Ukraine and the consequent refugee crisis and the spike in energy prices. The worsening of social conditions triggered renewed attention from governments and public administrations on the effective provision of social policies. These efforts extend beyond income support, drawing on the European Pillar of Social Rights and the growing emphasis on social inclusion policies. The significance of comprehensive social policies, which integrate income support with activation measures, is increasingly acknowledged as essential in addressing vulnerabilities effectively.

In this context, the peer countries shared their experiences with innovative and holistic social integration programmes. In Slovenia, various social activation projects were developed between 2017 and 2022, with the support of ESF+ funding. Notably, the Ministry of Labour, Family, Social Affairs and Equal Opportunities oversaw a nationwide pilot social activation project. The project operated in parallel to existing social and employment programmes and involved multiple actors. Project Unit included coordinators facilitating candidate interviews, programme support, and stakeholder coordination. Local centres of social work identified the candidates. Programme providers, selected through public tenders, offered diverse programmes with social, employment, and skills training components.

In Belgium the state of policy evaluation is marked by a historically weak demand for such assessments, despite increasing recognition of their importance. Shortcomings in policy evaluation include a lack of quality assurance, unclear expectations regarding evaluation types, a lack of integration into the policy cycle, and insufficient coordination of evaluation research, resulting in limited cross-sectoral policy analysis. However, despite these challenges, there is a growing emphasis on evidence-informed policy, also enabled by EU funding. The inter-federal policy support centre in Belgium addresses poverty, existential insecurity, and social exclusion. The centre produces biennial evaluation reports. The Poverty Barometer monitors the Federal Plan for Poverty Reduction's strategic objectives, and significant investment in BELMOD2 (i.e., a microsimulation model based on the Belgian EUROMOD model, running on a detailed administrative input) supports simulations and ex-ante evaluations related to social protection adjustments. However, rigorous evaluation research remains ad hoc, lacking a systematic approach.

Portugal is currently working on the revision of solidarity benefits, with the aim of creating a single social benefit covering various benefits, simplifying the legal and procedural regime, facilitating access and combating non-take-up (work is currently under development with OECD's support). Finally, in Czechia, the large wave of incoming Ukrainian refugees prompted the government to design and implement different social inclusion schemes.

Political will and legislative tools

The Spanish experience with the Policy Lab highlighted that political will, coupled with the establishment of appropriate legal foundations, are essential elements to experimentation in social policy evaluations (see Section 2.1).

Among the other participating countries, there was consensus that the attention to more rigorous policy evaluation in the field of social policy is increasing in the EU, with more data

being collected. However, examples of legal provisions mandating robust evaluations remain scarce.

In Slovenia, the evaluation of social policies is not systematic and not based on experimental or quasi-experimental designs. Often, there appears to be a lack of distinction between monitoring of policies, systematically performed, and impact evaluation of policies seldomly carried out. Overall, the evaluation of social inclusion policies is mostly performed in pilot projects, especially those funded by the ESF/ESF+.

Availability of funding

RCTs and quasi-experimental policy evaluations in the field of social policies are time consuming and resource intensive. The experiences shared by Belgium, Slovenia and Spain highlighted that the availability and adequacy of funding and resources are essential to enable social experimentation in the policy evaluation and to ensure the robustness of the design and implementation.

EU funding can support these initiatives. The ESF/ESF+ supported the Belgian CREActief project (see Box 2), Housing First pilot in Czechia and the Slovenian pilot social activation project. The Employment and Social Innovation (EaSI) programme financed the MISSION project in Belgium (see Box 2). NextGenerationEU enabled the Spanish Policy Lab. EU Member States have also the [Technical Support Instrument \(TSI\)](#) at their disposal. These experiences also showed that the process of applying for EU funding can be difficult and adequate time should be allocated to this.

Box 2. Belgium's experiences with experimental methods of impact evaluations

CREActief (2020-2022)

- **Government level:** Flemish initiative
- **Funding:** ESF
- **Brief description of the project:** CREActief introduced a structured trajectory for vulnerable job seekers, incorporating creative methodologies and psychoeducation to foster personal growth and bridge the gap to the labour market.
- **Evaluation method:** The project underwent evaluation with a sample of 46 individuals in the intervention group and 32 in the control group. The trajectory, consisting of an intake, group sessions, and a retrospective, aimed to strengthen individuals personally and reduce their distance from the job market. Despite some non-significant effects due to a small sample size, the evaluation affirmed the desired person-strengthening impact of CREActief.
- **Lessons learnt:** CREActief faced challenges amid the COVID-19 crisis, impacting the organisation of group sessions and necessitating adjustments to the recruitment strategy. Despite these challenges, the project's effectiveness was demonstrated, leading to validation by ESF managing authorities and securing additional funding for scaling up CREActief.

MISSION (2016-2019)

- **Government level:** Local initiative
- **Funding:** EaSI
- **Brief description of the project:** The 'Mission' project was implemented by the Public Centre for Social Welfare in Kortrijk (PCSW Kortrijk) and aimed to address the challenges faced by vulnerable families with children in navigating the fragmented local support landscape. It introduced an 'outreaching case management' methodology to improve the utilisation of local assistance, enhance access to financial support, and ameliorate living conditions for such families.

- **Evaluation method:** The project employed a randomised controlled trial (RCT) with baseline, six-month, and twelve-month measurements to assess the effectiveness of the outreaching case management approach.
- **Lessons learnt:** Challenges included identifying needy families not already known to aid organisations and participant dropouts. Despite these challenges, the project exemplifies successful social innovation aligned with a scientifically rigorous evaluation.

Source: Adapted from Peer Country paper and presentation for Belgium

Capacity building

RCTs and quasi-experimental designs require internal and/or external expertise for planning, implementation, and interpretation of counterfactual impact evaluation results.

Building capacity for RCTs is a multifaceted undertaking that requires careful planning, dedication, and collaboration across administrative units and external expertise. The provision of training by evaluation experts, covering diverse skills such as data analysis, effective communication of sensitive results, and research management, and early involvement of scientists were identified as key strategies. Moreover, incorporating capacity building activities into human resources processes was highlighted, emphasising the need for time, resources, and ongoing commitment.

A trade-off between external expertise and in-house capabilities emerged as a critical consideration. The Spanish Policy Lab, for instance, benefits from its collaboration with CEMFI and J-PAL, whose staff work part-time alongside public officers. In some countries, state-funded research institutes support public administration in research and evaluation efforts (for example in Slovenia), while others form partnerships with academia (Belgium) or utilise cross-departmental or cross-governmental analytical units (Bulgaria). The participation of external researchers and contractors is also ensured through procurement processes. In Malta, for example, the best value method in procurement is used considering factors such as quality and expertise in addition to price in the selection processes.

3.2 Design phase

During the working groups, the discussion focused on different critical elements of the design phase, such as:

Encompassing multi-level collaboration

One of the success factors identified by Spain in the Policy Lab is the inclusion and collaboration with a multitude of stakeholders, from the engagement of high-level scientific advisors to the involvement of third-sector organisations throughout all the phases of the policy experiments and scale-up. On the one hand, the presence of scientific experts can help defining and maintaining ethical standards and guide the design and analysis phase. On the other hand, third-sector organisations can be recruited, for example via public procurement tenders, as implementing partners that could carry out the projects. Because third-sector organisations are often linked to their territory, they are best placed to identify potential beneficiaries and to support the design phase with insights from the specific territorial needs.

The presence of a multitude of stakeholders requires significant coordination and communication efforts to ensure commitment throughout the entire phase of the experimentation and beyond. The discussions focused on the best practices to maximise the collaboration, such as the clear assignment of roles and responsibilities, the continuous communication of the developments of the project, the transparency from all the parties involved and the sharing of common objectives.

Ethics and communication

RCTs require careful ethical considerations and cannot be used to evaluate interventions where the right to the intervention is legally granted to everyone. A phase-in design can be used to minimise ethical concerns, implying that the control group received the intervention as part of a later phase of the intervention.

The Spanish Policy Lab and experience with the inclusion pathways stressed that ethical considerations should be included in the policy experiments from the design phase, to ensure the protection and respect towards the beneficiaries, with the following lessons learnt:

- From the design phase, it is crucial to include the point of view of beneficiaries to improve the understanding of their needs and identify possible hurdles that could affect recruitment and ongoing participation.
- Potential participants need to be informed that they are part of an experiment and that they may or may not receive support. It is important to use a transparent, but also familiar and reassuring language to ensure the engagement of beneficiaries.
- Implementing partners in the third-sector, local and regional governments may resist the implementation of these experiments, because their core mission is normally to try to assist everyone in need, rather than targeting their action to a restricted group of beneficiaries. In this context, it is important to clearly communicate the aims of the experiments and clarify in advance that experimentation is a critical phase to enable the scale-up of successful projects.
- The selection of implementing partners should be carried out keeping also ethics in mind. They should have the independence and neutrality standards to recruit beneficiaries.
- The presence of ethical protocols and/or an Ethical Board, like in the Spanish case, is advised to ensure the implementation and respect of ethical standards. Such boards could include representatives of beneficiaries to reflect their point of view.

Sample size

Attrition during pilot projects is quite common, as exemplified in the Belgian MIRIAM-project (see Box 3). Therefore during the design phase, decisions must be made, concerning the target sample size. The participating countries identified a trade-off between aiming at a large sample, which is more resource-intensive but also can ensure better scientific results, and smaller and easier to recruit sample size.

Box 3. Belgium's experiences with small sample sizes and attrition

MIRIAM (2016-2017)

- **Government level:** Federal initiative
- **Brief description of the project:** The MIRIAM project, initiated in 2015 with support from the State Secretary for Poverty Reduction, aimed to empower single mothers by providing intensive, gender-sensitive guidance through Public Centres for Social Welfare (PCSW), reducing poverty, and addressing social isolation.
- **Evaluation method:** The project underwent a two-year evaluation starting in its initial phase, utilising an experimental design conducted by the Karel de Grote University College. Case managers guided single mothers, and the impact was measured through three surveys with validated scales.
- **Lessons learnt:** The experimental design showed promise but faced challenges in demonstrating unequivocal impact due to variations in approach among PCSWs and a small sample size. Adjustments in the second research cycle, including increased number of participants and simplified measurements, proved successful in observing targeted changes in the intervention group. Challenges remained,

such as differences in approach, measurement instrument validity, and guiding non-native speakers. The project transitioned from experimental to structural, highlighting the complexity and challenges of impact evaluations in this policy context and emphasising the importance of a good design.

Complexity of the intervention

Other challenges for the design of RCTs is that they work best at checking the effect of single interventions, and that they can be more complex, time-consuming, and costly to implement than mere monitoring approaches. However, RCTs can provide strong evidence of causality and are valuable tool to test new approaches before their scaling-up.

The challenges in the assessment of the pilot social activation project in Slovenia included the project's complexity, linked to the variety of actors involved in the implementation, and the inability to access personal data of participants due to data protection rules. An application was developed to collect anonymised data on participants, aiding in the comparison of outcomes. The evaluation focused on programme implementation, identifying good and bad practices, programme exits, the functionality of the social activation system, and cooperation among key actors. The emphasis was on assessing gains and benefits for participants and improving cooperation between actors. A mixed-methods approach, combining quantitative and qualitative data, was employed. Surveys were conducted for various stakeholders. Focus groups with policymakers and field visits to programme sites were also conducted. Despite the challenges in structuring the collected data, the researchers produced five interim reports and two final reports with different focuses. The evaluation aimed to determine the project's success at three levels: individual participants, local-level actors and stakeholders, and the broader social system. The findings indicated positive effects for participants, particularly in social goals such as empowerment, positive problem-solving approaches, increased self-confidence, and enhanced competencies. The project also improved cooperation between local centres of social work and employment offices.

Indicators and access to administrative data

Another critical moment in the design phase is the definition of the indicators and evaluation requirements for the projects. To capture the outcomes of an intervention, a combination of a few key subjective indicators and objective indicators should be used. Moreover, the implementing partners should be aware in advance of the evaluation requirements, in order to already collect the appropriate data.

Having access to administrative data in the design and evaluation phase emerged as a critical point for the success of the experiments and the analysis of the pilot projects. Administrative data can help to identify potential targets of the policies, but also to track socio-economic characteristics of participants over time to gain insights over the long-term effects of the implemented schemes. For this reason, access to administrative data is crucial also in the implementation and follow-up phases. Two main types of obstacles have been identified:

- The legal access to administrative data, which is limited due to restrictive data privacy rules and lack of regulatory framework that facilitates data requests.
- The interpretation of administrative data, that are often not designed for scientific purposes. The efforts carried out by Eurostat to enable links between the EU surveys and administrative data and to harmonise public registers are welcome and will constitute a solution in the future.

3.3 Implementation phase

Participating countries had the opportunity to brainstorm on the challenges and solutions to different elements of the implementation phase, as such:

Recruitment

The experiences shared by Slovenia and Spain revealed the difficulties in reaching the target group, keeping the participants active and engaged throughout the whole duration of the experiment/project. Despite meeting the success criteria of the programme, the social activation pilot project in Slovenia highlighted motivation challenges, as evidenced by high drop-out rates (45% of enrolled participants dropped-out). Over the project duration, 20 633 candidates were identified, 5 051 enrolled, and 2 234 completed the programme.

Attrition

Attrition can adversely affect the robustness of evaluation of policies. Therefore, it is crucial to identify strategies that maximise participation and engagement among participants. The following reasons can explain the high level of dropouts. For the control group, this may be linked to the lack of motivation to stay in a project without receiving a benefit and/or support. For both groups, high mobility rates of vulnerable populations can also explain the lack of continuity with social programmes. Other elements, such as stigma or lack of immediate positive effects can also contribute to high attrition rates.

Response rate to surveys

Offering cash or non-cash incentives to both groups to participate in follow-up interviews, surveys and monitoring activities can constitute a possible best practice to mitigate this risk. In addition, the third-sector organisations involved as implementing partners can play a crucial role to continuously communicate with the participants, ensuring their engagement and participation.

Flexibility and continuous engagement

The implementation of social projects is subjected to uncontrollable external factors that can affect its implementation. Examples from Belgium (see Box 2) described the difficulties encountered during the COVID-19 pandemic in collecting data and following-up the participants. In this context, some flexibility in the design and implementation of intervention is key to ensure adequate responses to unforeseeable circumstances. The presence of scientific advisors on the team can once again support this goal.

3.4 Follow-up and scaling-up

Different obstacles concern the final evaluation and scale-up phase, such as the difficulties in evaluating heterogeneous projects in different social and territorial contexts, the time and budgetary constraints, or the lack of clarity on the decision-making authority.

A successful evaluation strategy should include meticulous preparation for result interpretation, placing a strong emphasis on additional research and continuous follow-up. External validation, through collaboration with independent entities or experts, can contribute to the credibility of the findings. Employing theory-based approaches, rooted in established social and behavioural theories, provides a conceptual framework for understanding the mechanisms underlying the intervention's impact.

When contemplating the scaling up of pilot schemes, careful consideration must be given to the design of a robust analytical framework. This framework serves the critical purpose of assessing whether the intervention, which demonstrated success in a smaller sample or specific context, is expected to yield similar outcomes when expanded to a larger target group and applied in a different setting. Anticipating potential adjustments and adaptations

to the intervention in advance is essential to ensure its relevance and effectiveness in diverse circumstances.

Incorporating cost-benefit analysis into the scaling-up strategy becomes paramount. Evaluating the economic feasibility of the policy intervention helps policymakers and stakeholders understand the financial implications and potential return on investment. This step is crucial for making informed decisions about resource allocation and for aligning the policy with broader economic goals. By combining rigorous evaluation methodologies, theoretical foundations, and economic analysis, scaling up social policies can be approached with a comprehensive understanding of both their impact and feasibility. Nonetheless, further lessons will be derived in the upcoming months, with the advancements of the Spanish Policy Lab activities.

From the lessons learnt shared by the peer review countries, it emerged that the availability of financial resources is crucial not only in the design phase but also for the scale-up of projects. This is exemplified by the pilot projects implemented in Belgium, such as CREActief, which was funded by the ESF and, upon positive evaluation, received additional funding for its scale-up. Other EU funding schemes can also support scaling-up of pilot projects, such as the ESF+ and NextGenerationEU. Moreover, the lessons learnt from the Belgian experience also highlighted that these schemes provide funding subjected to rigorous policy evaluations, that can hence foster the integration of scientific evidence-based policy evaluation mechanisms in the national social policy systems.

4 Key policy messages

The key insights gained from the presentations and working group discussions during the Peer Review, structured along different phases of the experiments, as well as some encompassing elements, are summarised below.

Building robust social policy evaluation: key preconditions

- Despite growing social needs and mounting attention to identifying well-designed and effective social policies, in many EU countries there is not a solid tradition of scientifically rigorous policy evaluation in the field of social inclusion. Throughout the EU, there is a lack of experience and awareness about the potential benefits of using experimental and quasi-experimental designs to evaluate and implement social inclusion policies, with some institutions also exhibiting resistance to change. Lessons learned from the experience of the Spanish Policy Lab point to the importance of political will, supported by an appropriate legal framework, to create the appropriate conditions for systematically including these activities to support effective policymaking and help assess the most effective interventions to be scaled up, leading to an optimal allocation of resources.
- Availability and adequacy of funding and resources are essential to enable social experimentation in the policy evaluation and to ensure the robustness of the design and implementation. The European Union offers a variety of support measures for carrying out experimental and quasi-experimental policy evaluation in the field of social inclusion, such as , especially [the Employment and Social Innovation \(EaSI\)](#) strand, as well as the [Technical Support Instrument \(TSI\)](#) and [NextGenerationEU](#). Nonetheless, the process of applying for funding can be time and resource intensive and contingency plans should be envisaged.
- Raising awareness and capacity development of government officials are crucial for helping policymakers make well-informed decisions based on evidence. Involving policymakers and administrators early in the process, and providing training in analytical skills (e.g., data analysis and effective communication of results), is key to establishing a strong culture of evidence-based policymaking. Furthermore, giving

public administration access to specific expertise and fostering collaboration with scientific advisors enhances overall effectiveness, fostering a culture of learning and continuous improvement. Finally, mutual learning events and other occasions to share best practices between different Member States can represent an essential learning platform. However, it should be stressed that the capacity-building process requires time, resources, and continuous commitment.

Efficient collaboration between policy makers and stakeholders with external actors is required from design of experiments to the interpretation of evaluation results

- In the development and implementation of RCTs and other quasi-experimental policy evaluation designs, many actors are involved, from the different levels of public administration to implementing partners, including third-sector organisations and members of the scientific community, all of them with different starting points and interests. To ensure consistency, all stakeholders should be involved and trained at an early stage. It is important to have different mechanisms in place to ensure an effective collaboration, communication, and management of expectations between the parties. This large ecosystem of stakeholders involved is expected to yield positive results beyond the experiments themselves, but also to cross-fertilise and strengthen a culture of multi-level collaboration aimed at targeting the most vulnerable population.
- Throughout all the phases of the design, implementation and evaluation of the policy experiments, the support and guidance of the scientific community are essential to ensure the robustness of the approach and validity of the findings, from respecting the ethical standards to defining the indicators to selecting the sample.

A careful and accurate design of the experiments and evaluation methods is crucial to their success

- RCTs are a potent tool to assess the causal impact of a policy; yet they are not universally applicable to every social policy intervention and come with inherent limitations, such as logistical challenges, ethical concerns, a relatively narrow focus, and difficulties in extrapolating findings. The decision to use RCTs or other counterfactual impact evaluation methods depends on the unique context, the nature of the policy under scrutiny and the availability of resources.
- The design phase in RCTs and quasi-experimental approaches is crucial for crafting schemes that address specific scientific questions. Particularly in social policies, which often involve comprehensive programmes with multiple tools, accurately isolating the impact of interventions is challenging. To tackle this, involving researchers from the beginning and ensuring all partners understand the methodology are essential steps.
- Differentiating between a treatment group and a control group (which does not benefit from the tested measure) entails significant ethical considerations. The drafting of ethical protocols and/or the support of an ethical board have been identified as good practices. Setting up the right level of randomisation, fostering transparency and adopting appropriate communication style with participants, ensuring their consent at all stages of the experiment and data collection processes are key elements of a holistic ethical approach.
- Beneficiaries should also be involved from the design phase, to understand their needs and identify potential barriers and obstacles for recruitment and continuous participation. Reaching out to participants for the experiments is challenging; some practical considerations in this regard must be carried out during the design phase. Selecting potential beneficiaries can be supported by third-sector organisations, which are aware of social needs at the local level.

- When setting the size of the sample, attention should be paid to the trade-off between sourcing a large sample size, which is more costly and difficult, but also more scientifically robust, and a smaller one.
- Common quality standards for evaluation should be set at the design phase, including regarding the list of result indicators (both quantifiable and qualitative, including subjective) used to evaluate the intervention. Attention should be paid to ensuring access to the administrative data needed for evaluation purposes. In some countries, access to administrative data remains difficult, because of inflexible legal frameworks, strict privacy rules and lack of infrastructures for the anonymisation of sensitive personal information, and necessary steps must be prepared in advance, including protocols to require consent from participants.

The implementation phase requires appropriate mitigation strategies

- During the implementation phase, some flexibility should be allowed to permit the required adaptations in how the experiments and evaluations are run, while maintaining the overall approach to ensure consistency in the robust results. Maintaining the commitment of partners to deliver the results is essential and recalls the importance of appropriate monitoring and communication channels, both formal and informal, between involved stakeholders, throughout implementation of experiments.
- The stability of the sample size during the experiment, essential to gather evidence-based results, is a key challenge for interventions targeting vulnerable target groups. Indeed, social policy interventions suffer from an elevated level of attrition, as both participants from the treatment and control group may drop out of the project. Countermeasures and incentives should be in place to ensure participants stay in the project, ensuring the scientific robustness of the findings, despite attrition. Implementing partners including third-sector organisations are an integral part of the strategy to engage and motivate participants, thanks to their proximity to the target group. Another identified good practice is to provide an incentive, cash or non-cash, to survey respondents after the end of the observation period.

Scaling up of pilot interventions: a path to rigorous evidence-based policy making

- Thorough preparation for result interpretation, emphasising additional research and follow-up, external validation, and employing theory-based approaches are essential elements of a winning evaluation strategy.
- When considering scaling up pilot schemes, it is important to design a robust analytical framework to assess whether the same intervention is expected to yield similar outcomes, when generalised to a larger target group and applied in a different context. It is also useful to consider any adjustments in advance. These considerations should be complemented by cost-benefit analysis, to estimate the economic feasibility of the policy.
- Similarly, meticulous preparation is essential for scaling up successful experiments, including legal and financial frameworks, to avoid unnecessary delays. In this context, the European Union can provide financial support through different instruments. Notably, the requirements to obtain this financial assistance often entail rigorous monitoring and evaluation of policies effectiveness, which can support the growth of a culture of rigorous policy evaluation.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from: <https://op.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.



Publications Office
of the European Union