Emilia NASIŁOWSKA, Monika WOŹNIAK, Zbysław ZIEMACKI
National Centre for Research and Development, Warsaw, Poland

EVALUATION OF POLISH R&D SUPPORT PROGRAMMES – PRACTICES AND PERSPECTIVES

Key words

Research and development, evaluation, innovation, NCRD, policy effectiveness, impact, counterfactual impact evaluation.

Summary

The National Centre for Research and Development (NCRD) is the implementing agency of the Minister of Science and Higher Education.

The main task of the NCRD is the management and execution of strategic research and development programmes, which lead directly to the development of innovativeness. The NCRD is obligated by law to conduct a systematic evaluation of its strategic research and development programmes and other tasks of the Centre, including an assessment of their impact on the development of science and economy.

The article refers to **case studies of two evaluations conducted by the NCRD** and presents problems and challenges connected with measuring the effectiveness and innovation of tested programmes. Properly designed *Impact evaluation* should provide an answer to basic questions: *Does public intervention work?* and *Why*? The answer for the first question can be found by using counterfactual methods, which seek to identify the net effects or impacts of interventions. This is known as **Counterfactual Impact Evaluations (CIEs)**, and it is based on comparison of results to estimate what would have occurred otherwise.

Introduction

A research and development (R&D) sector is one of the most important factors for social-economic growth. Poland faces many challenges in the field of R&D policy that require ambitious and sustained efforts. According to the European Union Innovation Scoreboard 2013, Poland is among the EU countries with the lowest level of R&D expenditure and one of the worst performers in broader innovativeness indicators [3]. However, it should be noted that, in recent years, Poland has achieved a high growth in investment in R&D. In the period of 2008-2012, Gross Domestic Expenditure on R&D almost doubled from 7.7 bil PLN in 2008 to 14.3 bil PLN in 2012. Private R&D funding is also steadily increasing, and the share of Business expenditures for R&D in total R&D expenditures increased from 28.1 in 2011 to 32.3% in 2012 [8]. Research, development, and innovation are key policy components of the EU strategy for economic growth, "Europe 2020." To meet the Europe 2020 target in the R&D intensity (R&D expenditure as a percentage of GDP), which is 3% for the EU and 1.7% for Poland, Poland needs to substantially raise its rate of increase in R&D intensity [2]. The National Centre for Research and Development (NCRD) is fulfilling these goals by implementing programmes that are focused on the transfer of research and experimental development results to the economy. Because such R&D support involves large amounts of public money, it is essential to measure the impact and effectiveness of such support by systematic evaluation. NCRD is obligated by law to conduct a systematic evaluation of its strategic R&D programmes and other tasks of the Centre, including an assessment of their impact on the development of science and economy.

1. Selected evaluation studies carried out by NCRD

1.1. Mid-term evaluation of the "Work safety improvement in mines" strategic programme

One of the main objectives of the NCRD is conducting programmes that are strategic from the point of view of research and innovation policies and serve for Poland's social and economic development. They are based on the National Research Programme (Krajowy Program Badań),[1] which specifies the strategic directions of scientific research and development. In 2012, the Centre completed a mid-term evaluation of the "Work safety improvement in mines" strategic project.

Results of studies analysing the causes and high death toll of the 2009 Wujek- Śląsk mine blast provided a basis for undertaking actions aimed

[1] The National Research Programme (Krajowy Program Badań) was established by the Resolution No. 164/2011 of the Council of Ministers of 16 August 2011.

at creating solutions to avoid such threats and tragic accidents in the future. Therefore, in a letter of 29 March 2010, the Minister of Science and Higher Education commissioned the "Work safety improvement in mines" strategic research project to the NCRD. The aim of the project is to develop organizational and technical solutions whose implementation will contribute to minimizing threats and increasing safety improvement in mines. Eight research tasks were assigned to scientific and industrial consortia consisting of scientific units, mines, and suppliers of technology, services, and products used in mining.

A permanent growth of risk connected with mining fossil fuels and minerals is caused, among others, by the facts that geological conditions are becoming ever more difficult, mining is conducted at greater depths, the concentration of extraction is increasing constantly, and that the mining sector in Poland has been underinvested in for many years. The results of scientific studies conducted as part of the project should be implemented in the form of specific provisions in mining regulations and rules followed by mining services in everyday practice connected with designing, conducting and monitoring underground work and preventing potential threats.

The aim of the project's mid-term evaluation was to assess the current progress of research tasks, which allowed the determination of the chances of meeting the project's objectives and the identification of any obstacles and threats obstructing the achievement of the main objective. If products manufactured so far were compliant with the provisions of the project contracts and the chances of increasing demand for the product in the market, including mines, was assessed. In addition, the level of innovation of the product in the current state of the art was evaluated. An opinion on whether the continuation of the project is in accordance with the goals of the national research and support for innovation policies was obtained.

Due to a small number of research tasks implemented by beneficiaries, as well as the need to learn more about the rules of cooperation between partners in the consortia, it was decided that the following qualitative methods and techniques should be used:

- In-depth interviews with managers of particular research tasks, a coordinator from the NCRD and a representative of the authority supervising the project – the Steering Committee;
- Group interviews (one with leaders of the consortia and partners and one with potential recipients of the products of research tasks); and,
- Expert panels, whose aim was to verify the preliminary conclusions and evaluate preliminary results of project implemented so far, e.g. in terms of their innovation.

The main objective in evaluating an R&D project is to determine the level of the innovation of the research results. In the case of the mid-term evaluation of the "Work safety improvement in mines" project described herein, this was made especially difficult since most of the research was still in progress and its

final effects were not yet available. It was thus necessary to base the evaluation on the beneficiaries' declarations concerning the expected products of R&D and expert opinions, which were discussed during an expert panel. However, there is no comparative hard data and relevant knowledge of global solutions to unequivocally and indisputably determine if the product in question is unique.

Another difficulty in evaluating an R&D project is the need to evaluate solutions in a narrow field of science, which the evaluation team usually does not specialize in, and the necessity to cooperate with an expert in the given field. Only after learning about the project, its environment and results of research tasks, is it possible to appropriately lead such cooperation and obtain a fair evaluation of the project from the expert. This needs to be done by the evaluating team, so that they can ask relevant research questions, as well as by the scientist or practitioner asked to cooperate, so that they can respond with a well-thought-out answer. As for the commercialization of the results of given research tasks, the opinions of end users should be sought in order to allow the researchers to verify the demand for the given product.

The largest challenge of the evaluation was to work with the ill--programmed logic of the project, since the main objective and specific objectives were not specified. It was thus impossible to refer to objectives in the evaluation of product, the results, and influence indicators, and to specify programme risks, as well as to determine if the project is compliant with the national research and innovation policy. Above all, this made it especially difficult to evaluate the effectiveness of the project, since there were no clearly determined evaluation criteria. It was impossible to establish whether it should be evaluated with reference to the profits from the commercialization of the products developed, or the results of introducing given machines, or whether it should be evaluated according to the methods and regulations of mines from the point of view of workplace safety as considered by workers or owners. In addition, due to the specific nature and narrow domain of the project in question, there was no data needed for benchmarking costs incurred in research. No business plans or economic analyses were made by beneficiaries, which did not help in the evaluation of effectiveness, but drafting them was not one of the requirements in filing grant applications.[2]

2. Assessment of the Commissioned Research Projects' effectiveness and utility

The evaluation of the Commissioned Researched Projects (CRP), conducted in 2013, was the first ex-post research on programme supporting scientific research carried out by the NCRD. The evaluation from methodological point of

[2] More detailed information on the evaluation of the "Work safety improvement in mines" strategic programme can be found at: http://ncbr.gov.pl/gfx/ncbir/pl/defaultopisy/839/1/1/prezentacja_raportu_koncowego_ncbr_bezpieczenstwo_w_kopalniach.pdf.

view was a mix of qualitative and quantitative methods (including statistical and bibliometric analysis). The methodology was proposed by the NCRD and enriched by the improvements suggested by the evaluation's contractor (an independent research firm). There were two main interrelated evaluation goals. The first aim was the assessment of the CRP's impact on the socio-economic development of Poland, and the second goal was the analysis of management procedures and the commissioning process of research projects.

CRPs have been executed since the mid-1990s[3]. In the beginning, CRP operated under the supervision of the Ministry of Science and Higher Education until 2007, when the NCRD took over the responsibilities for financing the CRP scheme. The legal basis under which the CRP operated was changed in 2004 as a result of the modifications in the system of science and research financing and the establishment of the National Framework Programme[4] (NFP) in 2005. The NFP defined the priority area for scientific research, and the CRP scheme was the main instrument of achieving those goals. Behind the rationale for establishing NFP (and in consequences the CRP) was the pursuit for supporting those scientific areas which would have the most significant influence on the stimulation of sustainable economic development and the improvement of the quality life in Poland in the future. The CRPs were large interdisciplinary projects carried out by consortia having from 2 to 19 partners (mostly scientific entities).

The evaluation included 86 out of 167 projects co-financed within the CRP scheme. The sample of projects had to be chosen mainly because of difficulties with access to documentation and information about all the projects. The lack of exhaustive information about the CRP was caused by two main interrelated factors: the time that passed from the projects' completion and the approach toward the programming of intervention, which prevailed in the past. As almost all of the projects subjected to the evaluation had ended a few years before evaluation started, there were great difficulties with collecting contact data and information about the CRP scheme (including scheme assumptions, monitoring data and reports, contacts to projects' coordinators and experts, who were engaged in assessing and choosing projects and who were responsible for creating the CRP scheme). Vaguely stated, NFP's goals and the insufficient monitoring system of CRP caused that the assessing effectiveness of the scheme became a true challenge, since the approach toward programming intervention according to PCM rules was not very popular in the past. The large diversity of project subjects and disciplines resulted in different kinds of project outcomes and results that could not be easily compared and assessed. What made the

[3] CRPs were executed by virtue of the Act on the establishment of the Committee of Scientific Research and the Committee of Scientific Research's resolution no 1/94 on the criteria and mode of granting funds form national budget for commissioned research projects.

[4] The National Framework Programme was the first attempt at concentrating significant funds for the scientific researches recognized as the most important in the future of Poland and united Europe http://www.pptb.pl/Krajowy_Program_Ramowy.pdf).

assessment even more complicated was the fact that the basic researches were prevailing among research projects, and tracking down their results and economic impact was hardly possible. Due to the same reasons, the evaluation of project's innovativeness (with reference to both single projects and the CRP as the whole) was not an easy task. The trial of the measurement of the economic impact on the macro level was also only partly successful. Despite the fact of finding some statistically significant correlation between the CPR's presence and increase of some macroeconomic indicators, a full-blown analysis could not be conducted. The difficulties resulted from the high-level aggregation of available data and the long-term effects of fundamental studies, which could not always be estimated financially. The important factor that hampered gaining reliable assessment was also relatively little experience of the evaluators, who despite being experts in evaluation of programmes co-financed by the UE, had a little practice in the evaluation of programmes devoted to R&D support.

To sum up, conducting the evaluation of the CPR indicates that most of difficulties connected to carrying out ex-post research stemmed from three sources:

1. Programme design/architecture – a lack of assumptions for monitoring and evaluation in programme/scheme, lack of a proper programme monitoring system resulting in lack of databases, and a weak institutional continuality;
2. Difficulties in the assessment of R&D projects results (especially identification of basic researches' results) – difficulties in assessing and comparing the scientific, social, economic, and social effects of projects; and,
3. A weak evaluation capacity in terms of the assessment of R&D projects – the use of a standard approach, the lack of sophisticated methodology and little experience in R&D project evaluation, a lack of preparation for the assessment of R&D project results in the form of human capital and knowledge, and publications (difficulties in comparing and assessing their utility).

However, most of above-mentioned difficulties can be avoided in the future, because the evaluation capacity of firms is getting stronger and the methodology is developing to meet needs of the assessment of even very sophisticated support programmes.

3. Perspective – looking for more evidence – Counterfactual Impact Evaluation

The NCRD is conducting a large variety of programs supporting R&D activities. Currently, there are more than 40 programmes implemented in the NCRD; therefore, evaluation is very complex and the selection of proper methodology is a real challenge. There are many different types of evaluation studies. Some of them focus on programme processes and implementation (formative evaluations), and others look at the effects or impacts of programmes,

typically called summative evaluations [9]. The second group seems to be very useful and effective in identifying the precise effects and impacts of programs implemented by the NCRD. Good understanding of what works or what does not work, and why it does or does not work, for whom does work, and in which contexts this takes place, is crucial for policy makers and for proper programing of future programs [5].

Properly designed *Impact evaluation* should provide the answer to the basic questions: *Does public intervention work?* and *Why*?

The answer to the first question can be found by using counterfactual methods that seek to identify the net effects or impacts of interventions. Studies that include these methods can be called **Counterfactual Impact Evaluations (CIE)**, which are based on the comparison of results to estimate what would have occurred otherwise. In practice, it is made by a comparison of two groups; one with treatment (observed) and another without treatment (control group). The difference in outcomes between the treated group and the control group is the estimate of impact. The CIE indicates if intervention makes any difference and if the difference is caused by the intervention itself or caused by something else and what would happened without intervention.

To answer the question *why* intervention works and produces effects (intended and unintended), we need to use ***Theory-Based Impact Evaluation (TBIE),*** which refers to a logical framework. A theory-based impact evaluation focuses on programme theories, i.e. the assumptions of policy makers and stakeholders on the preconditions, mechanism, and context for an intervention to work. Theory-based impact evaluations test these assumptions against the observed results following the different steps of the intervention logic and examine other influencing factors [7].

It should be noted that both questions cannot exist in complete separation from each other and both methods complement one another especially in evaluations realized after intervention (ex-post). The CIE shows a causal link between interventions and their results, while theory-based evaluation methods are useful to identify causal mechanisms that are also very important for programing and making decisions concerning future interventions [1].

The **CIE** approach is promoted by many international organisations, including the European Commission and the World Bank. The major benefits of it includes easy to interpret results, an essential ingredient for cost-benefit and cost effectiveness calculations, and it can be broken down in separate estimates for subgroups, provided that the subgroups were defined in advance [1].

A fundamental issue in the CIE is a matter of causality and answer the question about the causal link between intervention and observed results. Some effects can be caused by other factors not connected to our intervention. Programme supporting companies' R&D investments can be used as an example. To estimate impact, a researcher could simply compare levels of R&D expenditures in supported companies before and after intervention.

A growth in R&D expenditures could be considered the effect of the implemented programme. However, in fact, there are many other factors affecting R&D expenditures in the companies, and they are independent from intervention, i.e. global economic situation, local businesses, or access to other funds.

In order to avoid such bias caused by these factors, it is necessary to compare beneficiaries with non-beneficiaries (control group). To do this, two samples/groups should be selected – one with beneficiaries and the other with non-beneficiaries. However, the biggest challenge in the CIE approach is **a selection bias,** which means that there are primary differences between beneficiaries and non-beneficiaries before intervention that influences comparison results after it [10].

A good practice to avoid a selection bias is the **experimental design approach** where beneficiaries are selected randomly – Randomised Controlled Trials (RCT). This method assumes that two groups are randomly selected from the same population, and they are similar with only difference in that one group received treatment. Thus, any difference between beneficiaries and non-beneficiaries is caused by the treatment. Unfortunately, this method has many limitations and can be applied in very few and rather simple interventions, and it is not the case of R&D programs where it is unacceptable to select R&D projects randomly.

More applicable are quasi-experimental (non-experimental) methods, which also can be used to construct samples. The most common among them are propensity difference-in-difference, score matching (PSM), and regression discontinuity design. The following is a very brief description of these methods:

Difference-in-difference – The first step in this methodology takes the difference between the pre-test and post-test outcome values of a treatment group and a non-treatment-control group, because the control group does not have to be similar to the treatment group [10]. An example of this method in R&D is investigating subsidies for companies aimed at increasing R&D expenditures in these companies. First, we have to compare R&D expenditures trends between the control and the treatment group before the intervention and check if it is similar. Then, we compare these trends between treated and control groups after receiving treatment. Impact is estimated by subtracting the pre-intervention difference in outcomes from the post-intervention difference.

Propensity Score Matching (PSM) – The design attempts to select a control group that is as similar as possible to the treatment group, based on observable characteristics. In this method, it is crucial to identify as many variables that might have an effect the selection process as possible. Propensity in this method means the ability/probability to apply characteristics that are estimated by analysing identified variables. Then cases with similar propensity scores among beneficiaries and non-beneficiaries are compared.

The process at its heart has the following steps: (1) Identify key variables which are thought to predict membership in the treatment group, (2) use logistic

regression to generate a scoring system, based on these variables, to predict the likelihood of belonging to the treatment group, (3) match each member of the treatment group with a control group which has a similar score; and (4) estimate the effect as a difference between means [6].

Regression discontinuity design – The method is based on the cut-off point or threshold, comparing units slightly above threshold (beneficiaries) and slightly below (non-beneficiaries). Assuming that compared units around the threshold are similar, the effect of treatment can be measured by comparing the difference between these units.

To sum up, the main goals of **quasi – experimental methods** are to obtain an unbiased estimate of the change the intervention under consideration has brought about. Thus, there is no random selection in this method, and they require far more attention to methods accounting for potential differences between treatment group members and potential controls that are likely to affect the decision to participate and the results. The key is the proper selection of a plausible control group. A failure to select an adequate control group and account for remaining differences between the two groups in the analysis weakens the credibility of estimates and can confound attempts to rule out alternative explanations for any observed effects [4].

4. Limitations of CIE

Gaining reliable outcomes from CIE requires a reasonable number of treated units and of broadly similar non-treated units and other factors like behavioural motive, replicable nature, and homogenous treatment. Some of the most common factors that affect applicability of CIE are presented below.

Table 1. Factors affecting applicability of CIE

Type of policy	Support for R&D projects	Investment support	Renewable energy	Urban renewal	Transport infra-structure	Human capital investment
Behavioural (vs. redistributive) motive	++	++	++	+	+	+
Replicable nature (vs. idiosyncratic)	++	++	+	–	–	++
Homogenous treatment (vs. composite)	++	+	+	– –	–	+
Large numbers of eligible units	+	++	+	–	– –	++
APPLICABILITY OF CIE	HIGH	HIGH	MIXED	LOW	LOW	HIGH

Legend: ++ positive contribution; + moderate contribution; – limited obstacle; – – serious obstacle

Source: *Alberto Martini Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy,* 6th European Conference on Evaluation of Cohesion Policy, Warsaw, November 30, 2009, p. 14

The most important requirement is good data, and in case of the CIE, it is not only for the treated units, but also for compared non-treated units. Even if the applicability of the CIE is high, it will be impossible to perform a counterfactual analysis without sufficient data [7]. An access to good data is crucial for every kind of evaluation, and lack of information is the most common source of problems for proper performance of evaluation.

Conclusions

Innovations and R&D play an increasingly important role in economic and social growth. Therefore, the proper support of R&D projects has become a crucial aim for Polish policy makers and for the NCRD, which is a special agency created for improving innovativeness in Poland. However, evidence based policy demands a good support in the form of reliable information that stems from (among others) evaluations.

However, there is little experience in the evaluation of programs supporting R&D projects in Poland. The results of evaluations conducted by the NCRD and the Centre's experience show that, in order to get reliable information, the evaluation should become an essential part of intervention programming. Including evaluation into intervention logic enables one to design in advance a proper monitoring system, which is crucial for the future assessment of results. The need of reliable information about intervention results can be provided by the employment of the CIE approach. However, since the CIE methods are very data-hungry and demand experienced evaluators, they have not been used frequently in the past.

The articles considered and the authors' past experiences[5] with evaluations of programs that support R&D indicate that the counterfactual approach is an interesting and useful tool for assessing programs results and recommend it as an approach that should be vastly used in future evaluations of program impacts.

References

1. Martini A.: Counterfactual impact evaluation: what it can (and cannot) do for cohesion policy, 6th European Conference on Evaluation of Cohesion Policy, Warsaw, November 30, 2009.
2. Europa 2020 in Poland, European Commision's recommendations http://ec.europa.eu/europe2020/europe-2020-in-your-country/polska/country-specific-recommendations/index_en.htm.
3. European Commission, Innovation Union Scoreboard 2013, http://ec.europa.eu/enterprise/policies/innovation/files/ius-2013_en.pdf.

[5] More detailed information about the NCRD programs' evaluations and their findings can be found one the NCRD's website: http://www.ncbir.pl/o-centrum/ewaluacja/ewaluacja-2013/.

4. European Commission Design and commissioning of counterfactual impact evaluations A Practical Guidance for ESF Managing Authorities, October 2012.
5. European Commission Design and commissioning of counterfactual impact evaluations A Practical Guidance for ESF Managing Authorities, October 2012 pp. 15.
6. European Commission EVALSED: The resource for the evaluation of Socio-Economic Development, July 2012 pp. 127 http://ec.europa.eu /regional_policy/sources/docgener/evaluation/guide/guide2012_evalsed.pdf.
7. European Commission, Impact Evaluation Centre http://ec.europa.eu/ regional_policy/information/evaluations/impact_faq_en.cfm.
8. Główny Urząd Statystyczny w Szczecinie, informacja sygnalna Działalność badawcza i rozwojowa w Polsce w 2012 r., październik 2013 r.
9. Newcomer K.: Basics of Design for Evaluation of Cohesion Policy Interventions in Evaluating the effects of regional interventions. A look beyond current Structural Funds' practice K. Olejniczak, B. Ledzion, S. Bienias, Warsaw 2011.
10. Jalil M.M.: Practical Guidelines for conducting research, Donor Committee for Enterprise Development, February 2013.

Ewaluacja polskich programów wsparcia B+R – praktyka i perspektywy

Słowa kluczowe

badania i rozwój, ewaluacja, innowacje, NCBR, efektywność polityk, wpływ, ewaluacja wpływu oparta na metodach kontrfaktycznych.

Streszczenie

Narodowe Centrum Badań i Rozwoju (NCBR) jest agencją wykonawczą Ministerstwa Nauki i Szkolnictwa Wyższego. Głównym zadaniem NCBR jest zarządzanie i realizacja strategicznych programów badawczych, które przyczyniają się do wzrostu poziomu innowacyjności polskiej gospodarki. Ustawa o NCBR obliguje Centrum do prowadzenia systematycznej ewaluacji strategicznych programów badawczo-rozwojowych oraz innych zadań Centrum, w tym do oceny ich wpływu na rozwój nauki i gospodarki.

Poniższy artykuł odnosi się do **stadium przypadku dwóch badań ewaluacyjnych przeprowadzonych przez NCBR** oraz podejmuje kwestie problemów i wyzwaniach związanych z pomiarem efektywności i innowacyjności badanych programów. Właściwie zaprojektowana ewaluacja wpływu powinna udzielić odpowiedzi na podstawowe pytania, czy interwencja publiczna odniosła zamierzony skutek? i dlaczego? Odpowiedź na pierwsze pytanie można określić za pomocą metod kontrfaktycznych, dzięki którym można oszacować efekty inter-

wencji netto lub w wymiarze ich wpływu. **Ewaluacja wpływu oparta na metodach kontrfaktycznych** polega na porównaniu faktycznych wyników interwencji oraz szacunków efektów działań przeprowadzonych w sposób alternatywny.