

**Theory-based
evaluation**
in
complex
environments

Theory-based evaluation in complex environments

Editors:

Jacek Pokorski, Zuzanna Popis, Teresa Wyszzyńska
(Polish Agency for Enterprise Development),
Katarzyna Hermann-Pawłowska (IDEA of Development Foundation)

Authors:

Michael Quinn Patton – Chapter I
Katarzyna Hermann-Pawłowska, Paulina Skórska – Chapter II
Marcin Kocór, Barbara Worek - Chapter III
Emanuela Sirtori, Elena Vallino, Silvia Vignetti – Chapter IV
András Kaszap, Jan-Philipp Kramer, Rafał Trzciński – Chapter V
Maciej Koniewski, Seweryn Krupnik – Chapter VI
Dominika Wojtowicz, Łukasz Widła-Domaradzki – Chapter VII
Maciej Gajewski, Andrzej Regulski – Chapter VIII
Anna Bruska, Katarzyna Lotko-Czech – Chapter IX

The views of the authors of the chapters contained in the publication express their own opinions and do not reflect the official position of PARP.

Copyright by the Polish Agency for Enterprise Development, Warsaw 2017

Photo on the inside of the cover:

Old city of Wroclaw with Town hall seen from church tower on colourful sunset
Copyright by toimeyk - Fotolia.com

ISBN 978-83-7633-334-2

Edition I

Circulation of 750 copies

Free copy

Purchase of the selected original works co-financed from the European Regional Development Fund under the Smart Growth Operational Programme, 2014-2020.



Delivery of original works, technical and language editing, layout and cover design by:
IDEA of Development Foundation, in cooperation with CG2 Sp. z o.o. and PARP

Translated by:

GTC AMG Sp. z o.o.

Preparation for printing, printing and binding by:

Grafpol Agnieszka Blicharz-Krupińska

Table of contents

Introduction	5
I. Michael Quinn Patton - <i>Developmental Evaluation</i>	7
II. Katarzyna Hermann-Pawłowska, Paulina Skórska - <i>A comprehensive review of the theory-based evaluation concepts</i>	21
III. Marcin Kocór, Barbara Worek - <i>Qualitative comparative analysis in theory-based impact evaluation</i>	35
IV. Emanuela Sirtori, Elena Vallino, Silvia Vignetti - <i>Testing intervention theory using Bayesian Network Analysis: evidence from a pilot exercise on SMEs support</i>	55
V. András Kaszap, Jan-Philipp Kramer, Rafał Trzciński - <i>The use of contribution analysis as a theory-based evaluation approach in the ex post evaluation of support to large enterprises under Cohesion Policy programmes 2007-2013</i>	73
VI. Maciej Koniewski, Seweryn Krupnik - <i>How to better understand the results of counterfactual analyses? Selection of beneficiaries for qualitative research using statistical matching, illustrated by an example of the impact evaluation of the Operational Programme Innovative Economy 2007-2013</i>	95
VII. Dominika Wojtowicz, Łukasz Widła-Domaradzki - <i>Stratified Propensity Score Matching – quasi-experimental research design for theory-based impact evaluation with lacking dependent variable</i>	115
VIII. Maciej Gajewski, Andrzej Regulski - <i>Theory-based evaluation in studies of the innovative project selection system in Operational Programmes for 2014-2020</i>	135
IX. Anna Bruska, Katarzyna Lotko-Czech - <i>Experience in the application of theory-based evaluation at regional level</i>	151

Introduction

We are pleased to present you a publication devoted to the theory and practice of evaluation of public interventions – *Theory-based evaluation in complex environments*.

The subject of the publication was inspired by challenges related to evaluating public initiatives, in a complex and ever-changing environment. An example here may be *inter alia* the challenges in implementing innovation-oriented programmes, in the face of various economic crises and changes in attitudes of consumers as well as in organizations and business models. Taking into account this type of determinants, both the programme theory and the logic of carried out evaluation activities should be continually adjusted. This is perfectly highlighted by Prof. Michael Quinn Patton in the text entitled “Developmental evaluation”, opening the publication.

Development of evaluation activities is also stimulated by the European Commission. In the present market reality the EC indicates the need for the more rigorous methodological approach in evaluation (including an innovative and experimental approach). It would allow to explain the actual mechanisms of intervention and to obtain reliable information about the real effects of policies. One of the approaches recommended for the evaluation of Cohesion Policy of the European Union (EU) for 2014-2020 (and for the *ex post* evaluation of support within the perspective of 2007-2013) is the theory-based evaluation (TBE). Its application should allow to explain the changes taking place as a result of intervention, but also contribute to improving the quality and usefulness of evaluation itself.

This direction of the development of evaluation entails a necessity to adapt the existing principles of implementing the programmes by establishing a stronger link between evaluation processes and the programming and implementation of interventions. It enforces cultural changes in the public sector. This issue is referred to in the text by Maciej Gajewski and Andrzej Regulski presenting an example of evaluating the systems used for selecting innovative projects. Linking evaluation of interventions to programming and implementation processes remains the challenge at the level of nationwide activities, but particularly in the evaluation of regional programmes (as it is well illustrated by the text by Anna Bruska and Katarzyna Lotko-Czech on experiences of using TBE at the regional level). It should be stressed that the approach based on the theory of change is not only useful for the evaluation itself but, most of all, may be helpful in implementing development policies, and particularly their continuous adaptation to the ever-changing context.

An argument for the selected subject of the publication was a niche in the area of national papers in this regard – both ideological as well as implementation papers

devoted to specific case studies and indicating how the theory-based approach should be applied in practice.

In comparison, in the case of the counterfactual impact evaluation (CIE) complementary to the theory-based evaluation and no less intensely promoted in the evaluation of impact of the EU funds – the literary achievements (particularly, PARP expert works) may be considered relatively substantial.

The publication organizes numerous concepts of theory-based evaluations (see text by Katarzyna Hermann-Pawłowska and Paulina Skórska), describes the example of methodology perfectly addressing the TBE ideas (see text by Marcin Kocór and Barbara Worek concerning the QCA technique i.e. helpful in discovering the causal mechanisms of intervention), and then shows the specific applications of theory-based evaluations or theory-based impact evaluations (TBIE).

In the publication, we have focused on showing the experiences of national experts, in particular resulting from TBIE implemented for PARP (see texts by Maciej Koniewski and Seweryn Krupnik - "How to better understand the results of counterfactual analyses? Selection of beneficiaries for qualitative research using statistical matching..." and by Dominika Wojtowicz and Łukasz Widła-Domaradzki - "Stratified Propensity Score Matching – quasi-experimental research design for theory-based impact evaluation with lacking dependent variable"). They are complemented with the ideas and achievements from other countries, also concerning PARP indirectly (two texts referring to the examples of the *ex post* evaluation of support for the EU Cohesion Policy for 2007-2013 for Small and Medium-Sized Enterprises and for large entities – respectively, the text by Emanuela Sirtori, Elena Vallino and Silvia Vignetti presenting the use of the BNA technique in testing the theory of intervention, and the chapter, prepared by András Kaszap, Jan-Philipp Kramer and Rafał Trzciński, presenting the use of the contribution analysis in TBE). We hope this will help to a greater extent, disseminate the valuable experiences of PARP in this regard – on the one hand, resulting from the use of innovative approaches in evaluation, on the other, concerning the innovation-oriented object of evaluation i.e. evaluation of various types of support for innovation and enterprise under the EU Cohesion Policy.

We wish you an interesting reading and effective TBE implementation and use of TBE results. We would like to thank all the authors and IDEA of Development Foundation which supported us in preparing this publication. We hope that the experiences and conclusions you will find in this book will inspire you to improve the actions implemented in the area of evaluation and development policies.

The editorial team

Developmental Evaluation

Introduction

Developmental evaluation supports innovative intervention *development* to guide adaptation to emergent and dynamic realities in complex environments¹. Developmental evaluation is *utilization-focused*² in that it focuses on a specific intended use – development – for specific intended users: social innovators adapting their interventions in complex dynamic environments. Evaluation for equity and the fostering of human rights, as part of achieving meaningful development results, often occurs in complex adaptive systems. A complex system is characterized by a large number of interacting and interdependent elements in which there is no central control. Complex environments for social interventions and innovations are those in which what needs to be done to solve problems is uncertain, where key stakeholders are in conflict about how to proceed. This is typically the situation when fostering human rights. What has worked in one place may not work in another. Context matters. Variations in culture, politics, resources, capacity, and history will affect how development initiatives unfold and how attention to equity and human rights is incorporated into those initiatives. In such situations, informed by systems thinking and a sensitivity to complex nonlinear dynamics, developmental evaluation supports increased effectiveness of interventions, social innovation, adaptive management, and ongoing learning.

The developmental evaluator is often part of a development team whose members collaborate to conceptualize, design, and test new approaches in a long-term, on-going process of continuous development, adaptation, and experimentation, keenly sensitive to unintended results and side effects. The evaluator's primary function in the team is to infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection and real-time decision-making in the developmental process.

Improvements versus Developments

There are many approaches to evaluation. Each, including developmental evaluation (DE), fulfills a specific purpose and adds a particular kind of value. As noted above, DE has proven especially relevant and attractive to those interested in systems change and social innovation in complex dynamic systems. These are systems where people are trying to bring about major social change by fighting poverty; homelessness; in-

¹ M. Q. Patton, *Essentials of Utilization-Focused Evaluation*, Sage, Thousand Oaks, CA 2011.

² *Ibid.*

equity; human rights abuses; community and family violence; and helping people with AIDS, severe disabilities, chronic diseases, and victims of natural disasters and war. A deep commitment to fostering human rights and supporting equity undergirds many of these interventions and systems-change initiatives. Canadian colleagues Frances Westley, Brenda Zimmerman, and I have studied successful social innovations. We reported what we found in a book entitled *Getting to Maybe: How the World Is Changed*³. To be a change agent is to think boldly and envision grandly. Complexity theory shows that great changes can emerge from small actions. This involves a belief in the possible, even the “impossible”. Moreover, major social innovations don’t follow a simple linear pathway of change. There are ups and downs, roller coaster rides along cascades of dynamic interactions, unexpected and unanticipated divergences, tipping points and critical mass momentum shifts, and things often get worse before they get better as systems change creates resistance to and pushback against the new.

Traditional evaluation approaches are not well-suited for such turbulence. Traditional evaluation aims to control and predict, bring order to chaos, by carefully specifying and measuring fidelity of implementation and attainment of predetermined priority outcomes. In contrast, developmental evaluation accepts turbulence and uncertainty as the way of the world, as social innovation unfolds in the face of complexity. Developmental evaluation adapts to the realities of complex non-linear dynamics rather than trying to impose order and certainty on a disorderly and uncertain world. DE does this by tracking and documenting emergent and dynamic implementation adaptations and results.

Many of those working to foster human rights tell me that they have experienced evaluation methods that are entirely unrelated to the nature of their initiatives. Identifying clear, specific, and measurable outcomes at the very start of an innovative project, for example, may not only be difficult but also counter-productive. Under conditions of great uncertainty, outcomes can emerge through engagement, as part of the process for change rather than prior to such change efforts. So-called “SMART objectives”⁴, imposed prematurely, are not smart – and can, in fact, do harm by limiting responsiveness and adaptability. Developmental evaluation is designed to be congruent with, and to nurture developmental, emergent, innovative, and transformative processes.

Developmental Evaluation and Complexity Theory

Complexity as a construct is a broad tapestry that weaves together several threads relevant to innovation and evaluation: non-linearity; emergence; dynamic systems; adaptiveness; uncertainty; and co-evolutionary processes⁵. Developmental evaluation, likewise, centers on situational sensitivity, responsiveness, and adaptation, and is an

³ F. Wesley, B. Zimmerman, M. Q. Patton, *Getting to Maybe: How the World Is Changed*, Random House, Canada 2006.

⁴ SMART objectives: specific, measurable, attainable, realistic, timely

⁵ M. Q. Patton, *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, Guilford, New York 2011 <http://www.guilford.com/excerpts/patton.pdf> (access on 25.05.2017)

approach to evaluation especially appropriate for situations of high uncertainty, where what may and does emerge is relatively unpredictable and uncontrollable. Developmental evaluation tracks and attempts to make sense of what emerges under conditions of complexity, documenting and interpreting the dynamics, interactions, and interdependencies that occur as innovations and systems-change processes unfold.

Complex adaptive systems

Complexity writings are filled with metaphors that try to make complex phenomena understandable to the human brain's hard-wired need for order, meaning, patterns, sense-making, and control, ever feeding our illusion that we know what's going on. We often don't. But the pretense that we do is comforting - and sometimes necessary for some effort at action. So complexity theorists talk of flapping butterfly wings that change weather systems and spawn hurricanes; individual slime molds that remarkably self-organize into organic wholes; ant colonies whose frantic service to the Queen mesmerize us with their collective intelligence; avalanches that reconfigure mountain ecologies; bacteria that *know* the systems of which they are a part without any capacity for self-knowledge; and 'black swans' that appear suddenly and unpredictably to change the world. Complexity science offers insights into the billions of interactions in the global stock market; the spread of disease throughout the world; volatile weather systems; the evolution of species; large scale ecological changes; and the flocking of migrating birds. Complexity theorists explain the rise and fall of civilizations, and the rise and fall of romantic infatuation. That's a lot of territory. It can and should include attention to the rise and fall of evaluations.

Dealing with the unexpected

There is a lot of lip service in evaluation about looking for unanticipated consequences and assessing side effects; in reality, these are typically token elements of evaluation designs and inadequately budgeted, which are rarely given serious time and attention because of the overwhelming focus on measuring attainment of intended outcomes and tracking the preconceived performance indicators. You have to go out into the real world, do fieldwork, engage in open inquiry, talk to participants in programmes, and observe what is going on as interventions and innovations unfold to detect unanticipated consequences. I find that evaluators typically approach the unexpected and unanticipated in a casual and low-priority way, essentially saying, we'll look for unanticipated consequences and emergent outcomes if we have time and resources after everything else is done. But, of course, there seldom is time or resources. But the probabilities for unexpected impacts become quite high under conditions of complexity and so, developmental evaluators make expecting the unexpected fundamental to the work at hand.

Developmental Evaluation and Learning

Developmental evaluation supports learning to inform action that makes a difference. This often means changing systems, which involves getting beyond surface learning to a deeper understandings of what is happening in a system. Social innovators and social entrepreneurs, especially those working on issues of human rights and equity, are typically trying to bring about fundamental changes in systems, to change the world. To do so, they have to understand how the system they want to change is operating and to make the changes that change the system itself, by getting beyond temporary and surface solutions. This involves *double-loop learning*.

For decades three stories have been endlessly repeated: one about the stream of ambulances at the bottom of the cliff instead of building fences at the top; one about the numerous dead bodies coming down the river but all we do is build more impressive services for fishing them out; and one about giving someone a fish versus the value of teaching that person how to fish. In reviewing these stories, distinguished Australian action research scholar and practitioner Yolande Wadsworth⁶, has commented that they are reminders about our repeated tendency to go for the short-term quick fix (rather than to examine, come to understand, and take action to change how a system is functioning), that creates the very problems being addressed. Double-loop learning involves systemic solutions and is supported by evaluation attuned to looking for system explanations, and offering systemic insights.

In single-loop learning, people modify their actions as they evaluate the difference between desired and actual outcomes, and make changes to increase attainment of desired outcomes. In essence, a problem-detection--and-correction process, like formative evaluation, is *single-loop* learning. In double-loop learning, those involved go beyond the single loop of identifying the problem and finding a solution to a second loop that involves questioning the assumptions, policies, practices, values, and system dynamics that led to the problem in the first place, and then intervening in ways that involve the modification of underlying system relationships and functioning. Making changes to improve immediate outcomes is single loop learning; making changes to the system to prevent the problem or embed the solution in a changed system, involves double-loop learning. *Triple-loop learning* involves learning how to learn, and is embedded in the processes of developmental evaluation.

Developmental Evaluation in the Context of Development Evaluation

Developmental evaluation is easily confused with development evaluation. They are not the same, though developmental evaluation can be used in development evaluations.

⁶ Y. Wadsworth, *Building in Research and Evaluation: Human Inquiry for (truly) Living Human Systems*, Allen & Unwin, Sydney, Australia 2011

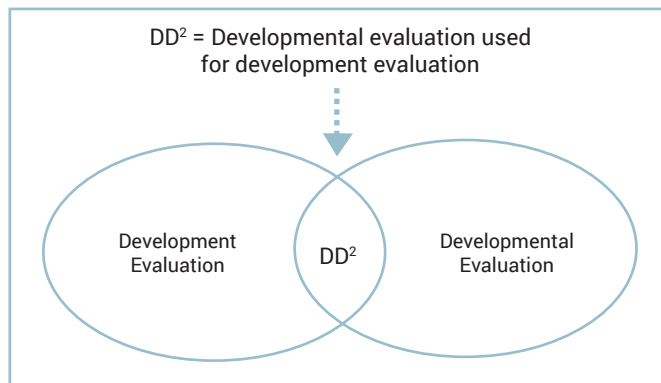
Development evaluation is a generic term for evaluations conducted in developing countries, usually focused on the effectiveness of international aid programmes and agencies. *The Road to Results: Designing and Conducting Development Evaluations*⁷ is an exemplar of this genre, a book based on The World Bank's highly successful International Programme for Development Evaluation Training (IPDET) which the book's authors founded and direct, and on which their book is based.

Developmental evaluation, as defined and described in the *Encyclopedia of Evaluation*⁸, has the purpose of helping develop an innovation, intervention, programme, or systems change. The evaluator uses evaluative methods to facilitate ongoing programme, project, product, staff and/or organizational *development*. The evaluator's primary function in the team is to facilitate and elucidate team discussions by infusing evaluative questions, data and logic, and to support data-based decision-making in the developmental process.

An evaluation focused on development assistance in developing countries could use a developmental evaluation approach, especially if such developmental assistance is viewed as occurring under conditions of complexity with a focus on adaptation to local context. Developmental evaluation can be used wherever social innovators are engaged in bringing about systems change under conditions of complexity.

The *al* in *developmental* is easily missed, but it is critical in distinguishing development evaluation from developmental evaluation.

Figure 1. Developmental evaluation used for development evaluation



⁷ L. G. M. Imas, R. C. Rist, *The Road to Results: Designing and Conducting Development Evaluations*, The World Bank, Washington, D.C 2009.

⁸ S. Mathison (ed.), *Developmental evaluation. Encyclopedia of evaluation*, Sage, Thousand Oaks, CA 2005, p. 116.

When I first labeled and wrote about *developmental evaluation* 15 years ago⁹, development evaluation was not a distinct and visible category of evaluation practice and scholarship. Evaluations in developing countries were certainly being conducted, but an identifiable body of literature focused on evaluating development assistance had not attracted general professional attention. One of the most important trends of the last decade has been the rapid diffusion of evaluation throughout the world including, especially, the developing world, as highlighted by formation of the International Development Evaluation Association. Confusion about the distinct and sometimes overlapping niches of development evaluation and developmental evaluation is now, I am afraid, part of the complex landscape of international evaluation. I hope that this chapter helps to sort out both the distinctions and the areas of overlap.

Examples of developmental evaluation in development contexts

- Working with agricultural scientists to take an integrated systems approach to 'orphan crops' would involve working with agronomists; soil scientists; plant breeders; water specialists; extension personnel; health; nutrition; gender researchers; and farmers, to conceptualize agricultural innovation as a complex adaptive system and identify real time indicators of the systems interactions and dynamics as the new farming approaches start to affect use of agricultural inputs, production techniques, farm labour, and farm family dynamics. This kind of holistic intervention involves changes in how traditionally distinct agricultural and nutritional scientists engage with farmers (separately rather than together), and would affect farm family decision-making and interactions.
- A microfinance intervention examined through a developmental evaluation lens would look at the infusion of capital as triggering a leverage point in a complex adaptive system. It would have implications for a variety of business calculation and decisions; interdependencies among loan recipients; relationships with consumers; and family finances and interpersonal dynamics. Watching for and adapting to emergent outcomes beyond simple use of small loan funds would be built into the evaluation design and real time feedback, as the microcredit system developed.

Examples of developmental evaluation in equity-focused evaluations

Developmental evaluations focusing on the marginalized and excluded populations help to adapt to rapidly changing conditions. Here are some examples.

- People living in poverty exist on the edge of subsistence. Sudden changes in food availability can move an entire population from subsistence to famine. Food insecurity can result from weather (severe drought or flood), political unrest (food transport is disrupted), and economic changes (increases in food prices). Sometimes all three factors – weather, political, and economic disruptions – occur simultaneously, creating a mutually reinforcing downward spiral on increasing des-

⁹ M. Q. Patton, *Developmental Evaluation. Evaluation Practice* 15, 1994, p. 311-320.

peration. Such situations requires real time data about what is happening to the people affected and how well-intentioned interventions are actually performing.

- Marginalized and excluded populations are especially susceptible to contagious diseases. For example, polio immunization campaigns have to be adapted to specific development contexts. Where polio eradication efforts have floundered, as in parts of Nigeria, Pakistan, and India, new outbreaks can break-out and spread rapidly in areas where the disease was thought to have been eradicated. For example, a developmental evaluator would help monitor rumors about resistance a vaccination campaign, as when Muslim mothers hid their children from vaccinators because of rumors that the vaccine was a Hindu plot to sterilize their children. Detecting and correcting such rumors in real time, as they emerge, can save lives.
- A human rights campaign anywhere in the world may have to be significantly adapted as street demonstrations calling for democratic reforms in Tunisia and Egypt (2011) change the global context within which human rights initiatives are undertaken. Marginalized, disempowered, and excluded populations can become homeless refugees when political turmoil accelerates and spreads.
- Responding to a humanitarian disaster, such as the earthquake in Haiti (2010), requires real time data about how local pockets of people are being affected; which roads are passable; where heavy rains after the earthquake are threatening the stability of remaining buildings; where there are outbreaks of cholera; where food, clean water, and medications are most desperately needed; and so on and so forth. Efforts to coordinate an international humanitarian response are inherently *developmental* because the disaster context is complex and emergent. The evaluation should also be developmental in support of ongoing humanitarian relief decision-making. Marginalized, disempowered, and excluded populations are often especially vulnerable in disaster situations because they tend to live in highly vulnerable areas that lack basic infrastructure. This makes delivering timely assistance all the more challenging. Developmental evaluation can track both developing vulnerabilities and developing interventions.

Dynamic versus static impact evaluation designs

As these examples illustrate, developmental evaluation views development interventions as dynamic and emergent in complex adaptive systems. Both the intervention and the evaluation are dynamic and adaptive. This stands in stark contrast to impact evaluations that use randomized controlled trials (RCTs) as a methodological framework. RCTs conceptualize interventions as occurring in closed systems, and study the intervention as a static and mechanical cause aimed at preconceived effects in a simple linear model of cause-effect. Such designs aim to standardize interventions and to control variation, which limits the utility and generalizability of findings¹⁰. In

¹⁰ For more on the mechanical and linear assumptions of RCTs, see chapter 12 in: M. Q. Patton *Utilization-Focused Evaluation*, 4th ed., Thousand Oaks, CA 2008.

contrast, developmental evaluations assume that development more often occurs in complex dynamic systems and puts a premium on understanding context, real time adaptability, and ongoing development, rather than generating high-fidelity and highly prescriptive practices. These differences go beyond methodological preferences and debates. They involve fundamentally different views about the nature of development, the contexts within which development occurs, how change occurs, and epistemological differences about what constitutes actionable knowledge.

Developmental evaluation and accountability

The traditional approach to accountability is to evaluate whether resources are used as planned, and whether targeted outcomes are attained. This is a static and mechanical approach to accountability that assumes designers know, three or five years in advance, what important outcomes to target and how to go about achieving those desired outcomes. Departing from planned implementation is considered implementation failure. Targeting new and emergent opportunities is considered 'mission drift.' The mantra of traditional, static accountability is plan your work, work the plan, and evaluate whether what was planned was achieved. But that's not how high performance organizations approach either development or accountability.

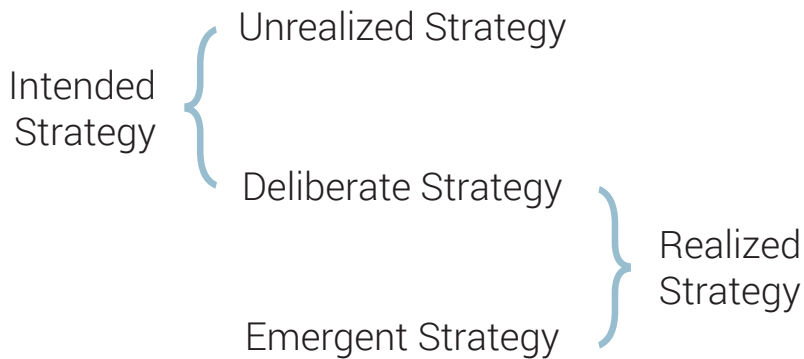
Henry Mintzberg is one of the world's foremost scholars on strategic thinking, organizational development, and the characteristics of high performing business. He has found that, implementing strategy is always a combination of deliberate and unplanned processes. In studying hundreds of companies over many years, he found that there is no such thing as a perfectly controlled, deliberate process in which intentions lead to formulation of plans, implementation, and the full realization of intended results. The real world does not unfold that way. As the graphic below shows, realized strategy (where you end up after some period of time) begins as intended strategy (planning), but not all of what is intended is realized. Some things get dropped or go undone because planning assumptions proved faulty in the face of real world processes; this he calls "unrealized strategy." What remains of the intended strategy he calls the deliberate strategy, which intersects with emergent strategy to become realized strategy. Emergent strategy comes from seizing new opportunities, which is another reason some things that were planned remain undone as new and better opportunities arise¹¹. In essence, a high performance organization that is paying attention to the world in which it operates does not expect to rigidly follow a highly prescriptive plan. The plan is a starting point. Once implementation begins, the plan has to be – and should be – adapted to what is observed and learned, in interaction with the complex adaptive system of real world dynamics.

Mintzberg's insights about strategy implementation in the real world contrast significantly with the classic accountability-oriented approach to evaluation in which pro-

¹¹ Chapter 1 in: H. Mintzberg, *Tracking Strategies*, Oxford University Press, New York 2007.

programme implementation and results are measured and judged based on what was planned to be done and achieved (intended outcomes). Under such an accountability framework, an innovative and adaptive programme that seizes new opportunities and adjusts to changing conditions will be evaluated negatively. Developmental evaluation, in contrast, expects that some of what is planned will go unrealized, some will be implemented roughly as expected, and some new things will emerge. *Developmental evaluation tracks and documents these different aspects of strategic innovation* – and their implications for further innovation and development. Accountability resides in carefully, systematically, and thoroughly documenting these developmental shifts, making transparent the data on which changes are made, and tracking the implications of deviations from the original plan - both deviations in implementation and in emergent outcomes.

Figure 2. Mintzberg on Strategy



Source: H. Mintzberg, *Tracking Strategies*, Oxford University Press, New York 2007.

Complexity-based developmental evaluation shifts the locus and focus of accountability. Accountability in developmental evaluation means documenting adaptations and their implications, not evaluating rigid adherence to planned implementation and pre-conceived outcomes. Why? Because complexity-sensitive developmental evaluation assumes that plans are fallible, based on imperfect information and assumptions that will be proven wrong, and that development occurs in dynamic contexts where even good plans will have to be adapted to changing realities. Thus, rather than becoming a barrier to adaptation, as occurs in traditional rigid accountability measures in which programmes are deemed to have failed if they depart from what was planned, developmental evaluation assumes a dynamic world with departures from initial plans. Developmental evaluation places the emphasis on understanding, supporting, and documenting adaptations and their implications.

Developmental Evaluation as Utilization-Focused

What brings me to complexity is its utility for understanding certain evaluation challenges. Complexity concepts can be used to identify and frame a set of intervention circumstances that are amenable to a particular situationally-appropriate evaluation response, what I am calling here - developmental evaluation. This makes dealing with complexity a defining characteristic of the developmental evaluation niche. Principles for operating in complex adaptive systems inform the practice of developmental evaluation. The controversies and challenges that come with ideas on complexity will also, and inevitably, afflict developmental evaluation. The insights and understandings of complexity thinking that have attracted the attention of, and garnered enthusiasm from, social innovators will also envelope developmental evaluation – and be the source of its utility.

Developmental evaluation is meant to communicate that there is an option in an approach to conducting evaluations that specifically supports *developmental adaptation*. In so doing, I place this approach within the larger context of *utilization-focused evaluation*¹². *Utilization-focused evaluation* is evaluation done for and with specific primary intended users for specific, intended uses. Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything that is done, from beginning to end, will affect use. ‘Use’ is about how real people in the real world apply evaluation findings and how they experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on achieving *intended use by intended users*. In developmental evaluation, the intended use is development, which I have here argued is a distinct and important evaluation purpose. The primary intended users are development innovators and others working to bring about major change.

Situation recognition and developmental evaluation

Astute situation recognition is at the heart of utilization-focused evaluation. There is no one best way to conduct an evaluation. This insight is critical. The design of a particular evaluation depends on the people involved and their situation. The Development Assistance Committee standards¹³ provide overall direction, a foundation of ethical guidance, and a commitment to professional competence and integrity, but there are no absolute rules an evaluator can follow to know exactly what to do with specific

¹² M. Q. Patton, *Utilization-Focused Evaluation, 4th ed.(...)*, Patton M. Q., Utilization-Focused Evaluation, in: M. Segone (ed.), *From Policies to Results: Developing Capacities for Country Monitoring & Evaluation Systems*, UNICEF 2009, pp. 252-277, M. Q. Patton, *Essentials of Utilization-Focused (...)*

¹³ DAC (Development Assistance Committee), *Quality Standards for Development Evaluation, 2010*, <http://www.oecd.org/dataoecd/55/0/44798177.pdf> (access on 25.05.2017)

users in a particular situation. Recognizing this challenge, situation analysis is one of the “essential competencies for programme evaluators”¹⁴.

The ideal is to match the type of evaluation to the situation and needs of the intended users to achieve their intended uses. This means – and I want to emphasize this point – *developmental evaluation is not appropriate for every situation*. Not even close. It will not work if the conditions and relationships are not right. The point here is that every evaluation involves the challenge of matching the evaluation process and approach to the circumstances, resources, timelines, data demands, politics, intended users, and purposes of a particular situation. Such matching requires astute situation recognition. Developmental evaluation is appropriate where the situation is understood to involve interventions and innovations in complex adaptive developmental situations¹⁵.

Figure 3. Use of evaluation



Written by Mark M. Rogers and illustrated by Lawson Sworch

“I can honestly say that not a day goes by when we don’t use those evaluations in one way or another.”

Source: Written by Mark M. Rogers and illustrated by Lawson Sworch

¹⁴ CES (The Canadian Evaluation Society), *Competencies for Canadian Evaluation Practice*, 2010. http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf (access on 25.05.2017)

¹⁵ M. Q. Patton, *Developmental Evaluation: Applying Complexity (...)*

About the author

Michael Quinn Patton

Michael Quinn Patton is the founder and CEO of Utilization-Focused Evaluation, an independent organisational development and program evaluation organisation. After receiving his doctorate in sociology from the University of Wisconsin–Madison, he spent 18 years on the faculty of the University of Minnesota (1973-1991), including five years as the Director of the Minnesota Center for Social Research and ten years with the Minnesota Extension Service.

Patton has written many books on the art and science of program evaluation, in which he emphasizes the importance of designing evaluations to ensure their usefulness, rather than simply creating long reports that may never get read or never result in any practical changes. He has written about evaluation, and worked in the field beginning in the 1970s when evaluation in the non-profit sector was a relatively new development.

Patton is one of only two recipients of both the Alva and Gunnar Myrdal Award (awarded by the Evaluation Research Society and subsequently by the American Evaluation Association) for “outstanding contributions to evaluation use and practice” and the Paul F. Lazarsfeld Award for lifetime contributions to evaluation theory (awarded by the American Evaluation Association). The Society for Applied Sociology honoured him with the 2001 Lester F. Ward Award for Outstanding Contributions to Applied Sociology. He was President of the American Evaluation Association in 1988 and Co-Chair of the 2005 International Evaluation Conference in Toronto sponsored jointly by the American and Canadian evaluation associations. He sits on the Editorial Advisory Board for The Foundation Review.

Bibliography:

1. CES (The Canadian Evaluation Society), *Competencies for Canadian Evaluation Practice, 2010* http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf (access on 25.05.2017).
2. DAC (Development Assistance Committee), *Quality Standards for Development Evaluation, 2010* <http://www.oecd.org/dataoecd/55/0/44798177.pdf> (access on 25.05.2017).
3. Imas L. G. M., Rist R. C., *The Road to Results: Designing and Conducting Development Evaluations*, The World Bank, Washington, D.C 2009.
4. Mathison S. (ed.), *Developmental evaluation. Encyclopedia of evaluation*, Sage, Thousand Oaks, CA 2005, p. 116.
5. Mintzberg H., *Tracking Strategies*, Oxford University Press, New York 2007.
6. Patton M. Q., "Developmental Evaluation", *Evaluation Practice* 15, 1994, p. 311-320.
7. Patton M. Q. *Utilization-Focused Evaluation, 4th ed.*, Sage, Thousand Oaks, CA 2008.
8. Patton M. Q., "Utilization-Focused Evaluation", in: Segone M. (ed.), *From Policies to Results: Developing Capacities for Country Monitoring & Evaluation Systems*, UNICEF 2009, pp. 252-277.
9. Patton M. Q., *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford, New York 2011 <http://www.guilford.com/excerpts/patton.pdf> (access on 25.05.2017).
10. Patton, M. Q, *Essentials of Utilization-Focused Evaluation*, Sage, Thousand Oaks, CA 2011.
11. Wadsworth Y., *Building in Research and Evaluation: Human Inquiry for (truly) Living Human Systems*, Allen & Unwin, Sydney, Australia 2011.
12. Wesley F., Zimmerman B., Patton M. Q., *Getting to Maybe: How the World Is Changed*, Random House, Canada 2006.

A comprehensive review of the theory-based evaluation concepts

Introduction

The objective of this chapter is to present the results of the most popular and broadly discussed evaluation approaches based on theory of the programme (*theory-based evaluation, TBE*). The multiplicity of theories in the literature calls for systematic overview and synthesis of terminology issues through identification of common components of various definitions and concepts, which make this approach original. Then, the chapter focuses on the analysis of selected TBE concepts, according to their approach to the so-called “black box” of the program, the problem of attribution and the manners of reconstructing and testing theories. Finally, the chapter discusses the extent to which this approach is applied in practice and opportunities provided by application of TBE in examination of public polices, particularly those related to complex programmes financed from EU funds.

The objective of the theory-based evaluation is to provide an answer to the following question: why and how a given intervention causes specific, intended and unintended, effects (outcomes). In other words, the approach does not focus on measuring the magnitude of effects, but rather on identification of the mechanism of the change - on defining why and how the intervention works¹⁶. For this purpose, it reaches to the inside of public intervention, trying to open the so-called “black box” of the programme. The notion of a “black box” refers to this part of intervention logic, which is between the “input” (resources involved in the programme) and the “output” (expected outcomes of the intervention)¹⁷. Stame defines the “black box” as the space between the actual input and expected output of the programme¹⁸.

Approaches to theory-based evaluations emerged as an answer to limitations of the previously dominant method-oriented approaches. This is because quasi-experimental studies have not brought the solution to the problem of opening up the “black boxes” of programmes or public interventions¹⁹. Thanks to these research schemes, one can at the most answer the question if a programme works, that is if the effect of an intervention is positive or negative and what is its magnitude. These approaches alone

¹⁶ European Commission, *Monitoring and evaluation of European cohesion policy: concepts and recommendations*, Brussels 2011.

¹⁷ R. Pawson, N. Tilley, *Realistic Evaluation*. Sage, London 1997.

¹⁸ N. Stame, “Theory-based Evaluation and Types of Complexity”, *Evaluation* 2014, vol. 10, no 1, p. 58.

¹⁹ *Ibid.*, p. 58.

do not facilitate going further in the analysis of public policies, explaining why and how a programme works. Therefore, they are suitable for summarising effects (summative evaluation), but only have a minor potential for formative evaluation (modification of a new intervention). As a result provided information has limited utility in the context of taking decisions on future designs of programmes and public policies²⁰.

The criticism of evaluation approaches based solely on quantitative methods for assessment of intervention outcomes, which started in 1980s, led to increasing interest in theory-based evaluation. Though some authors²¹ trace origins of this approach back to the works devoted to testing programme theory from 1930s²², further developed in the subsequent three decades, a clear paradigmatic shift in the approach to evaluation materialised together with the publication of Chen and Rossi from 1989²³. Since then the interest in TBE has been growing, which is demonstrated by the increased number of scientific articles on the subject, as well as by the proliferation of terms and various strategies of TBE practical use.

When referring to the theory-based evaluation, diverse terms are used, the definition scope of which is not always distinct²⁴. The most popular include: theory-driven evaluation, programme-theory evaluation, theory-guided evaluation, theory of action, theory of change, program logic, logical frameworks, realist / realistic evaluation, program theory-driven evaluation science²⁵.

The diversity of terminology forces a question, whether we are dealing with one approach or many different ones. And in consequence, whether it is possible to identify "a common denominator" for concepts described by various authors, which would set TBE apart from other evaluation approaches. Though definitional and operational homogeneity of TBE approach is sometimes questioned²⁶, common features of TBE evidencing consistence of this concept can be pointed to, ie.:

- a) opening up "the black box" of the programme in order to answer the question why and how intervention works (definition of mechanism of change),
- b) taking into account the evaluation context, including organisational environment, in which main stakeholders (actors) operate, as well as a broader cultural and political context of the programme,

²⁰ *Ibid.*

²¹ C. L. S. Coryn et al., "A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009", *American Journal of Evaluation* 2011, vol. 32, no. 2, p. 200.

²² This refers to the work of Tyler on formulating and testing the theory of the programme for evaluation purposes. In: *Ibid.*

²³ H.-T. Chen, P. H. Rossi, "Issues in the theory-driven perspective", *Evaluation and Program Planning* 1989, vol. 12, no 4.

²⁴ P. J. Rogers, C. H. Weiss, "Theory-based evaluation: Reflections ten years on: Theory-based evaluation: Past, present, and future", *New Directions for Evaluation* 2007, no 114.

²⁵ C. L. S. Coryn et al., *op. cit.*

²⁶ *Ibid.*

- c) methodological neutrality, ie. TBE does not give preference to any specific methods, but makes the choice of methods dependent on needs,
- d) identification of causal chains, linking expected/ensuing outcomes to intervention in order to solve the problem of attribution (to what extent a given intervention shall contribute/have contributed to a planned/observed change).

According to authors developing theoretical foundations of this approach, the research design based on TBE should consist of two components:

- a) conceptual - with the objective to formulate the programme theory, ie. causal model of the programme, in line with which the intervention is to achieve assumed outcomes;
- b) empirical - testing the theory, that is to verify whether, how and why the intervention causes observable changes (outcomes).

According to Rogers²⁷, if the programme logic has been reconstructed properly (conceptual component), but it is not used as the model driving the process of evaluation (empirical component), we are not dealing with theory-based evaluation.

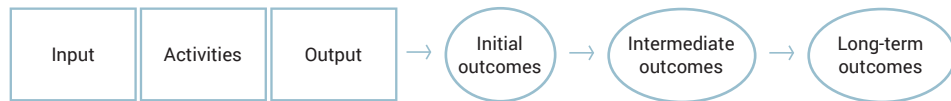
Programme theories are typically described or presented in a graphical form. It usually contains in its basic version linear model of the theory incorporating at least three components: input/resources - actions/processes - product/results/impacts. More developed complex model takes into account a number of factors, resulting from the context, in which the programme is embedded (e.g. needs, problems). In the example of the linear presentation of the theory shown below, its individual components should be understood as follows²⁸:

- a) inputs are resources (human, financial, physical) necessary to implement the programme,
- b) activities are actions undertaken to achieve the envisaged objective (e.g. organisation of a training),
- c) outputs are immediate results of actions undertaken (e.g. the number of persons trained),
- d) initial outcomes are changes to knowledge and skills,
- e) intermediate outcomes are behavioural changes,
- f) long-term outcomes are a solution of a social problem or meeting the needs of a given social group covered by the intervention.

²⁷ P. J. Rogers, *Program theory evaluation: Not whether programs work but how they work*, in: D. L. Stufflebeam et al. (ed.), *Evaluation models: Viewpoints on educational and human services evaluation*, Kluwer, Boston 2000, as cited in: Coryn et al., *op. cit.*

²⁸ C. L. S. Coryn et al., *op. cit.*, p. 201.

Diagram. 1. Graphic presentation of programme theory components - an example



Source: On the basis of S.L. Donaldson's diagram²⁹ presented in Coryn et al.

Methodological neutrality of TBE means there is no strict hierarchy of methods used in evaluation. As White has pointed out: "*Evaluations should be issues-led not methods-led*"³⁰. According to this author, the selection of methods should depend on evaluation questions and potential of various methods to give relevant answers. In practice, TBE typically combines the use of both qualitative and quantitative methods.

The outcomes of the comprehensive review

The outcomes of the comprehensive review presented in this chapter refer to two fundamental issues: the attitude of individual TBE concepts toward "the black box" and its content (1) and to the ways in which programme theory is reconstructed and tested. The beliefs about the nature of "the black box" determine the role and tasks of an evaluator. First of all, it determines the way in which the research is designed and which methods shall be selected for an evaluation study. The review focuses on the most popular and broadly discussed approaches, which are present in synthetic compilations and guidelines on theory-based evaluation.

Going back to the roots of TBE approaches, in their concept of theory-driven evaluation, Chen and Rossi³¹ concluded that one of the primary problems of evaluation studies was the fact that they focused on relationships between inputs and impact of intervention, failing to take into account the content of the programme itself. For this reason these are „black box evaluations“. However, the transition to the new paradigm in evaluations studies proposed by these authors, taking into account the theory of the programme, is not easy. According to authors, programmes are difficult to evaluate, as the process of their development is not based on social theories. As a result, objectives of programmes, envisaged outcomes and actions are not clearly explicit defined. Therefore, the evaluation study itself is reduced to nothing more than enumerating stakeholders, describing programmes and sometimes counting outcomes³², while the programme's "black box" is simply empty - the theory is missing³³. In this context, the

²⁹ S. L. Donaldson, *Program theory-driven evaluation science*, Lawrence Erlbaum, New York 2007, p. 25.

³⁰ H. White, "Theory-based impact evaluation: principles and practice", *Journal of Development Effectiveness* 2009, vol. 1, no 3, p. 162.

³¹ H.-T. Chen, P. H. Rossi, *op. cit.*

³² *Ibid.*, p. 299.

³³ N. Stame, *op. cit.*

role of the evaluation is to provide the missing theory, to define how the intervention might lead to specific outcomes. In order to develop the programme theory it is proposed to use not only available research results and social theories on which public programmes should be built, but also the perspective of programme's authors and the analysis of the programme implementation process.

Theory-based evaluation approach described by Weiss³⁴ emphasises the role of practice: in this perspective programmes have their own assumed theories, but they are distorted by the decision making process, which in fact is of political (strategic) nature. Therefore, a good programme theory explains how this decision making process happens in practice. According to the definition of Weiss from 1995, the theory of change is a set of assumptions pertaining to "mini-steps", which lead to attainment of a long-term goal of the programme. It interlinks between programme actions and results, which appear at each of the steps taken towards the objective (final outcome). Moreover, the researcher pointed out that "the black box" of a programme is full of different theories of change³⁵ - these are assumptions and beliefs of various stakeholders on how the programme should work. Thus, the fundamental task of an evaluator is to cooperate closely with a broad spectre of stakeholders in reconstruction of the theory of change. The role of evaluation is to confront and integrate these various views in order to build on synthetic theory of change. The first step on this path is to discover the implementation theory, that is to describe the steps of programme implementation. Only in the next stage the programme theory should be indicated, which uncovers the actual mechanism, which took place under intervention. When reconstructing the theory of the programme, the task of the evaluator is to first define the long-term vision of the programme, the final goal, and to link it to an existing problem to be solved. Then, it is necessary to define expected outcomes of the programme, results and outputs, which represent "mini-steps" towards attainment of the targets. Finally, programme stakeholders point to resources, which can be committed to delivery of planned interventions. As one can see, the process of establishing the theory is reversed - when reconstructing the theory, the evaluator goes back step by step from the goal (the final outcome) to the earliest changes launched (or planned to be launched) by the intervention.

The third approach in this discussion is the so called realistic evaluation proposed by Pawson and Tilley³⁶. The realistic approach to evaluation assumes that in the proper

³⁴ C. H. Weiss, *Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families*, in: J. Connell et al. (ed.), *New approaches to evaluating community*, The Aspen Institute, Washington DC 1995, as cited in: European Commission, *Evalsed Sourcebook: Methods and Technics*, 2013, the document available at: http://ec.europa.eu/regional_policy/en/newsroom/news/2013/11/evaluation-guidance-evalsed-guide-and-sourcebook-updates (access on 7.04.2017)

³⁵ C. Weiss, "Theory-Based Evaluation: Past, Present, and Future", *New Directions for Evaluation* 1997, no 76.

³⁶ R. Pawson, N. Tilley, *op. cit.*

theory of intervention, the outcome is the result of context and mechanism. The programme itself has a potential to generate specific outcomes, but eventually people (programme stakeholders), who are embedded in a context (e.g. organisational) and subject to a programme (public intervention), are either able or unable to initiate a specific mechanism of change. Here, the explanation of “the black box” dilemma is following: it consists of persons acting in a given context and according to a certain mechanism³⁷. Because a programme brings different outcomes to different stakeholders and in differing contexts, it is important for the decision making processes to establish in the course of evaluation, who works for whom and in which context³⁸. This facilitates proper replication of programmes in future.

In this case, the research design assumes that first the evaluator establishes how a mechanism might work, and then confronts it with opinions of stakeholders taking part in development and implementation of the programme. The eliciting the programme theory starts with an analysis of programme documentation and with interviews with programme authors in order to identify the programme theories. Then such theories are subject to codification through application of various techniques (i.e. concept mapping, development of problem and solution trees, scenario building) and as a result they are presented in a graphic form (diagram). The next step is the formalisation of the programme theory, i.e. translating it into hypotheses suitable for empirical testing. The final step is empirical research using qualitative and quantitative methods, which facilitates testing and refinement of the actual programme theory.

As one can see from the above compilation, the theory of change and realistic evaluation differ in the way they conceptualise the programme theory. In case of the theory of change, the reconstruction of the theory is effected through building a consensus between stakeholders, in order to subsequently test outcomes envisaged by them. The realists in turn assume a cumulative process of eliciting, testing and refining the theory. In the course of this process relations between interventions and their outcomes are sought, while the mechanism focused on establishing these relations is generative.

Another TBE approach focuses primarily on strategies for reconstruction of the theory of the programme. Leeuw and Vaessen treat “the black box” of the programme as a set of behavioural and social mechanisms influencing behaviour patterns of programme recipients³⁹. Thus, programme theories consist of a set of behavioural, social and institutional assumptions related to how programme actions and resources shall bring the expected change. Leeuw and Vaessen claim that when establishing the programme theory, assumptions should be identified pertaining to how and why programme ac-

³⁷ N. Stame, *op. cit.*

³⁸ European Commission, *Evalsed Sourcebook: Methods and Technics*, Brussels 2013, pp. 52–53.

³⁹ F. L. Leeuw, J. Vaessen, *Impact Evaluations and Development: NoNIE Guidance on Impact Evaluation*. The Network of Networks on Impact Evaluation, Washington DC 2009.

tions shall bring expected results. Leeuw also points to three autonomous strategies related to the sequence of steps necessary to establish these assumptions and methods for reconstruction of mechanisms of change, i.e.: policy-scientific approach, strategic assessment approach and elicitation method. Each of these strategies uses different methods for collection and analysis of data⁴⁰.

The policy-scientific approach assumes that reconstruction of the theory is based on the analysis of formal and informal documents and transcriptions of interviews in order to identify assumptions related to objectives of programmes and selected ways of their attainment. On this basis the assumed mechanisms, which are to solve a problem, are reconstructed. Argumentation analysis applied in philosophy and logic is used to assess the cohesion of reconstructed intervention logic. In this approach the evaluation of the programme theory may be effected through confronting various theories, empirical testing of the programme theory or through the process of continuous refining the theory with the use of different techniques of data collection⁴¹.

The strategic assessment approach in turn, focuses on reconstruction of assumptions underlying the programme design. For this purpose groups of stakeholders are formed, relatively homogeneous in terms of their characteristics and beliefs related to intervention. Within these groups an analysis of main stakeholders participating in the programme is carried out, and then assumptions leading to the success of the programme are identified related to the behaviours of stakeholders in the course of intervention implementation. The reconstructed assumptions are then subject to ranking, in order to identify the most important ones. Finally, a debate between the groups based on argumentation consisting of reconstructed assumptions is to facilitate understanding of strategies of different groups and their synthesis at the final stage of work on reconstructed assumptions. Here, the path to establish assumptions underlying the intervention leads through the consensus between different groups of stakeholders.

In case of the elicitation method, the basis for reconstruction of the theory are mental models, or in other words, cognitive maps of persons acting in organisations implementing programmes. These maps refer to persons' cognitions on organisation of phenomena in space, which affect their behaviours. Hence the key task of the evaluator here is to reconstruct stakeholders' cognitions of the organisation in which they work and its environment. This enables future comparison to evidence resulting from scientific research on similar organisations. Such approach first of all makes it easier for organisations to learn from previous experiences from programme implementation.

⁴⁰ F. L. Leeuw, "Reconstructing Program Theories: Methods Available and Problems to be Solved", *American Journal of Evaluation* 2003, vol. 24, no 1.

⁴¹ European Commission *Evalsed Sourcebook ...*, *op. cit.* p. 58

The last of TBE approaches discussed here - the contribution analysis - focuses not on the process of reconstructing the assumptions underlying intervention's theory, but on the assessment of their validity. The foundation of this approach is the belief that even if research results often show that there have been outcomes of intervention, it is not equivalent to the intervention bringing these outcomes about⁴². Therefore in this approach it is crucial to gather evidence confirming that it has been the programme, which contributed to the observed change. The basis for determining the contribution of the programme is the reconstruction of the initial theory of change through analysis of causes and effects of intervention, followed by verification of its credibility through gathering of existing evidence. Only such preliminary testing of the theory, in the light of existing evidence, allows its development for critical assessment. The next step is gathering of primary data, where additional evidence is needed. Finally, empirically collected data allow to construct a history of contribution, which is the more credible the stronger the evidence is. The history of contribution shows the existence of connections between the intervention and the observed changes. In this approach, "the programme black box" consists of actions, which lead to the sequence of outcomes connected by cause and effect relations. Therefore, the approach focuses on the ways to assess the credibility of relations between causes for launching the intervention and its outcomes.

The discussion on various TBE approaches to definition of programme theory and assessment of contribution of intervention to achieved outcomes, led to the question to what extent the theory-based evaluation is a distinct approach from the evaluation using (quasi)experimental methods. Some authors perceive the approaches as differing, because of different bases used for establishing causality⁴³. Others deny this dichotomy and propose solution connecting TBE with the approach using experimental research. White has proposed a theoretical framework, called theory-based impact evaluation (TBIE)⁴⁴, integrating both approaches. The six principles of TBIE include: (1) mapping out the causal chain (programme theory), (2) understanding context, (3) anticipating heterogeneity, (4) rigorous impact evaluation using a credible counter-factual scheme; (5) rigorous counterfactual analysis; and (6) using mixed methods. The use of the approach based on combination of theory and the counterfactual approach is perceived by experts and stakeholders⁴⁵ as the best approach to analysis of the impact of public policies (*theory-based counterfactual impact evaluation*) in the framework of *ex post* evaluation.

⁴² J. Mayne, "Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly", *The Canadian Journal of Program Evaluation* 2001, vol. 16, no 1, p. 3.

⁴³ E. Stern et al., *Broadening The Range of Designs and Methods for Impact Evaluations*, Report of a study commissioned by the Department for International Development, DFID 2012.

⁴⁴ H. White, *op. cit.*

⁴⁵ *Ibid.*, European Commission, *Evalved Sourcebook ...*, *op. cit.*

Practical application of the approach

TBE approach is promoted by numerous international organisations, such as for example the World Bank, the International Initiative for Impact Evaluation, the United Nations Evaluation Group⁴⁶, the European Commission⁴⁷. Nevertheless, practical application of TBE does not reflect the theoretical interest devoted to this issue in the literature⁴⁸. Moreover, the comprehensive review of research using TBE in 1990-2009⁴⁹ indicates that only in 13% of evaluations researchers formulated programme theories on the basis of programme observation, while only 36% of evaluations actually measured outcomes proposed in programme theory. Furthermore, Nakrošis claims that so far there have been no specific examples of good practices and potential benefits of application of theory-based evaluation in programmes financing regional development in EU Member States⁵⁰.

When it comes to theory-based impact evaluation itself, it seems that still not many studies are able to fulfil the assumptions of this approach in practice⁵¹. Despite the growing interest and rapidly increasing number of scientific articles devoted to this topic, there is no consensus on how to apply this approach in practice⁵².

In case of programmes and projects co-financed from European Union funds, TBE is considered by the European Commission as a useful evaluation tool at all stages of programme/project "life cycle"⁵³. At the stage of *ex ante* evaluation, managing institution should take care of specifying the programme theory. The indication of the programme logic should consist in explicit description of an envisaged mechanism of change underlying the intervention, through a clear definition of:

- a) how the expected change shall contribute to achievement of local and European objectives (so-called *programme theory*);
- b) in what way planned actions are to contribute to achievement of expected outcomes (so-called *implementation theory*).

In the course of *ongoing* evaluation, the objective of application of programme theory-based evaluation is to assess, whether the theory of implementation is executed effectively, thus leading to achievement of mid-term outcomes.

⁴⁶ C. L. S. Coryn et al., *op. cit.*

⁴⁷ European Commission, *Evalsed Sourcebook ...*, *op. cit.*

⁴⁸ C. L. S. Coryn et al., *op. cit.* p. 217

⁴⁹ *Ibid.*

⁵⁰ V. Nakrošis, "Theory-based evaluation of capacity-building interventions", *Evaluation* 2014, vol. 20, no 1.

⁵¹ H. White, *op. cit.*

⁵² V. Nakrošis, *op. cit.*

⁵³ M. Riché, *Theory Based Evaluation: A wealth of approaches and an untapped potential*, European Commission 2012.

In turn, the application of TBE approach at the level of ex post evaluation facilitates taking a closer look at the context and mechanism of change, in order to define the way of achieving (or failing to achieve) particular objectives of the intervention.

According to Stame⁵⁴ theory-based evaluation provides the framework to successfully evaluate the programmes implemented in multi-level management systems. In case of evaluations of such comprehensive programmes, an assumption should be made that we are not dealing with a single mechanism of change, but with several ones impacting one another. Therefore, according to Hedström and Ylikoski⁵⁵ and Rogers⁵⁶, as well as Rogers and Funnell⁵⁷, in conditions of complexity of evaluated matter attention should be paid to relations and dependences between various mechanisms. In this way it can be avoided to make mistakes when applying TBE. Other problems, which an evaluator applying TBE could face, include improperly reconstructed theories⁵⁸ or the trap of “absence of a theory”⁵⁹. Improper reconstruction of theories may result from the problem of *quasi* enriching of theories⁶⁰. This is the situation, when architects of policy insist on taking into account in the theory reconstructed by evaluators the assumptions, which in fact were not present, when a programme was developed. In case of “the absence of theory” we are dealing with graphic presentation of actions and outcomes without definition of logical links between them. Another problem could be failure to use the reconstructed theory in empirical studies, that is carrying out research irrespective of the reconstructed theory⁶¹. It is crucial, particularly when evaluating public policies, as often in the framework of interventions implemented there are no elaborated and well-structured programme theories (i.e. formulated in writing in the formal programme documentation). In such situation theories are often reproduced *post factum*, and in extreme situation even created *ad hoc* (“invented”) at the level of evaluation study. Therefore, such studies using TBE may be susceptible to the problems described above related to (re)creation of the programme theory and should be subject to special attention.

⁵⁴ N. Stame, *op. cit.*

⁵⁵ P. Hedström, P. Ylikoski, “Causal Mechanisms in the Social Sciences”, *Annual Review of Sociology* 2010, vol. 36, no. 1.

⁵⁶ P. J. Rogers, “Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions”, *Evaluation* 2008, vol. 14, no. 1.

⁵⁷ S. C. Funnell, P. J. Rogers, *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey Bass 2011, as cited in: European Commission. *Evalsed Sourcebook ...*, *op. cit.*

⁵⁸ N. J. Tilley, “Evaluation and evidence-(mis)led policy”, *Evaluation Journal of Australasia* 1999, vol. 11, no. 2.

⁵⁹ S. C. Funnell, P. J. Rogers, *op. cit.*

⁶⁰ European Commission *Evalsed Sourcebook ...*, *op. cit.*

⁶¹ S. C. Funnell, P. J. Rogers, *op. cit.*

Summary

In summary, the programme-theory evaluation has been attracting interest of theoreticians and practitioners for decades. This is because of the complementarity of TBE with approaches oriented primarily towards estimation of intervention effects and the high practical value of the approach (usefulness in assessment of effectiveness of public interventions). However, there is no integrated holistic concept for this paradigm, which is illustrated by the multiplicity of theoretical and methodological approaches to this topic. This in turn translates into relatively rare, so far, application of TBE in research practice.

Theory-based evaluation provides a conceptual framework and tools aiming to opening “the black box” of comprehensive programmes, which include - for example - programmes financed from EU funds. It makes it possible to reconstruct cause and effect relations between inputs (funds committed to the programme) and expected outcomes of actions planned. Its usefulness should be primarily seen through the lens of information it provides to designers and executors of public policies and programmes. TBE facilitates reconstruction of mechanisms of change envisaged in the programme, and in consequence feeds the decision making processes with information of how and why a given intervention works/fails to work. Thanks to application of this approach at the stage of programming, implementation and assessment of outcomes, it is possible to plan better future interventions, and to effectively take remedial action in the course of programme implementation.

About the authors

Katarzyna Hermann-Pawłowska

Analyst, evaluator, expert in the field of strategic planning, EU funds and public participation. She is a member of the Management Board and Strategic Project Manager in IDEA of Development Foundation. Europeanist by education, she graduated from the Adam Mickiewicz University in Poznań and the College of Europe in Natolin, she obtained the title of PhD in Economic Sciences at the Warsaw School of Economics. She has many years of experience in planning and implementing public interventions gained during work in public administration, including the Ministry of Regional Development. She participated in preparing the management system for the EU Cohesion Policy in Poland for 2007-2013 and was responsible for building the appropriate administrative capacity to manage and implement EU funds in that period. In addition, she carried out development projects in Tunisia, evaluation studies in Africa as well as research and watchdog projects geared towards promoting the idea of knowledge-based policies. She is an author of articles and publications in the field of process of Europeanisation of Polish public administration, multi-level governance, strategic planning, use of knowledge from the evaluation. She conducts training courses on, *inter alia*, EU Cohesion Policy, strategic planning or public participation.

Paulina Skórska

She graduated from Sociology and Psychology at the Jagiellonian University in Cracow. She is a scholar and graduate of the *Educational measurement* programme. She completed the schools in the field of the use of counterfactual methods – *Summer School on Counterfactual Impact Evaluation* (2013) and *Fundamentals and Methods for Impact Evaluation of Public Policies* (2014). She is a graduate of the *ICPSR Summer Program in Quantitative Methods of Social Research* (2014, 2015) at the University of Michigan. She is the Director for research and statistical data analysis in IDEA of Development Foundation. She has experience in designing and conducting evaluation and social studies, particularly for public administration. Since 2013, she has been the expert on the methodology and statistical data analysis at the Educational Research Institute. For several years, she has been cooperating with the Jagiellonian University's Centre for Evaluation and Analysis of Public Policies, as a coordinator and an expert in projects in the field of public policies aimed at the regional development, strategic planning and development of the education system. Since 2014, she has been advising on building an education evaluation system for the government agency of the Kingdom of Saudi Arabia (*Public Education Evaluation Commission*). She gives training to public servants and conducts academic classes on the methodology of social research.

Bibliography:

1. Chen H.-T., Rossi P. H., "Issues in the theory-driven perspective", *Evaluation and Program Planning* 1989, vol. 12, no 4.
2. Coryn C. L. S. et al., "A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009", *American Journal of Evaluation* 2011, vol. 32, no. 2.
3. Donaldson S. L., *Program theory-driven evaluation science*, Lawrence Erlbaum, New York 2007.
4. Funnell S. C., Rogers P. J., *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey Bass 2011.
5. Hedström P., Ylikoski P., "Causal Mechanisms in the Social Sciences", *Annual Review of Sociology* 2010, vol. 36, no. 1.
6. European Commission, *Monitoring and evaluation of European cohesion policy: concepts and recommendations*, Brussels 2011.
7. European Commission, *Evalsed Sourcebook: Methods and Technics*, Brussels 2013. Downloaded from http://ec.europa.eu/regional_policy/en/newsroom/news/2013/11/evaluation-guidance-evalsed-guide-and-sourcebook-updates.
8. Leeuw F. L., "Reconstructing Program Theories: Methods Available and Problems to be Solved", *American Journal of Evaluation* 2003, vol. 24, no 1.
9. Leeuw F. L., Vaessen J., *Impact Evaluations and Development: NoNIE Guidance on Impact Evaluation*, The Network of Networks on Impact Evaluation, Washington DC 2009.
10. Mayne J., "Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly", *The Canadian Journal of Program Evaluation* 2001, vol. 16, no 1.
11. Nakrošis V., "Theory-based evaluation of capacity-building interventions", *Evaluation* 2014, Vol. 20, no 1.
12. Pawson R., Tilley N., *Realistic Evaluation*. Sage, London 1997.
13. Riché M., *Theory Based Evaluation: A wealth of approaches and an untapped potential*, European Commission 2012.
14. Rogers P. J., "Program theory evaluation: Not whether programs work but how they work", in: D. L. Stufflebeam et al. (ed.), *Evaluation models: Viewpoints on educational and human services evaluation*, Kluwer, Boston 2000.
15. Rogers P. J., "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions", *Evaluation* 2008, vol. 14, no. 1.
16. Rogers P. J., Weiss C. H., "Theory-based evaluation: Reflections ten years on: Theory-based evaluation: Past, present, and future", *New Directions for Evaluation* 2007, no 114, s. 63–81.
17. Stame N., "Theory-based Evaluation and Types of Complexity", *Evaluation* 2014, vol. 10, no 1.

18. Stern E. et al., *Broadening The Range of Designs and Methods for Impact Evaluations. Report of a study commissioned by the Department for International Development.* DFID 2012.
19. Tilley N. J., "Evaluation and evidence-(mis)led policy", *Evaluation Journal of Australasia* 1999, vol. 11, no. 2.
20. Weiss C. H., "Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families", in: J. Connell et al. (red), *New approaches to evaluating community*, The Aspen Institute, Washington DC 1995.
21. Weiss C. H., "Theory-Based Evaluation: Past, Present, and Future", *New Directions for Evaluation* 1997, no 76.
22. White H., "Theory-based impact evaluation: principles and practice", *Journal of Development Effectiveness* 2009, vol. 1, no 3.

Qualitative comparative analysis in theory-based impact evaluation

Introduction

The objective of this chapter is to present the qualitative comparative analysis (QCA) as a technique which may constitute a significant supplement of research methods used in the theory-based impact evaluation. Owing to the specificity of this method, underlying assumptions, as well as the research procedure system, it may be successfully applied wherever we are interested in answering the question what factors determine the success of the public intervention, what role is played by this intervention in achieving desired changes or, finally – how the context and external conditions affect the results obtained as a result of taking this intervention. Undoubtedly, the nature of the questions, which the QCA results allow to answer, gives this technique an important position in the range of methods worth using in the impact evaluation.

The qualitative comparative analysis is an example of a technique that combines the quantitative and qualitative approach. Based on the in-depth analysis of individual cases and on their good knowledge, it goes to the comparative analysis aimed at identifying the configuration of factors, which may be described as the conditions necessary or sufficient for the occurrence of a given effect. This, seemingly simple, combination of the advantages of qualitative methods (specific to idiographic methods⁶² in-depth knowledge of a case) and the rules of formal logic (Boolean algebra)⁶³ allows to obtain the results which by far overcome the limitations of qualitative and quantitative methods. Such a limitation, in case of qualitative methods, is their inconclusiveness and

⁶² Idiographic methods are geared towards a detailed description of an individual case, while seeking the complete reconstruction of the causes and determinants of the occurrence of a given phenomenon. Usually, qualitative research methods are used in them. Often, qualitative methods as such are also referred to as idiographic methods. The idiographic approach is contrasted with the nomothetic approach, oriented towards the general explanation of a given category of phenomena with the use of few causal factors (cf. E. Babbie, *The Practice of Social Research*. PWN. Warsaw. 2004).

⁶³ Boolean algebra is general algebra applied in mathematics, theoretical informatics and digital electronics. Its name is derived from the name of mathematician, philosopher and logician George Boole. In Boolean algebra, we are dealing with binary data, which may assume only one of two possible values: 1 – true, 0 – false. On these data, we may perform three fundamental and baseline for QCA logical operations – negation/denial (NOT); alternative/logical addition/union (OR); conjunction/logical multiplication (AND) (cf. B., Rihoux, Ch., Ragin. C., (ed.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks, CA: Sage 2009).

excessive concentration on the data description while in case of quantitative methods – mechanical approach to analysed cases and ignoring their specificity⁶⁴.

As in Poland so far the QCA has been applied very rarely in evaluation studies⁶⁵, it is worth making the readers more familiar with the general assumptions of this method, presenting the analytical procedure related to its application and exemplary results it brings. The implementation of such objectives determines the structure of this chapter, which has been divided into three major parts. In the first part, we present the theoretical assumptions of the QCA, including the concept of causation based on the co-occurrence of many factors, in the second part, we describe the QCA as a research approach, in the third part, we present the research procedure related to the application of this method. We hope that the information contained in this chapter will encourage the readers to carry out independent explorations and to attempt to apply the QCA in evaluation studies⁶⁶.

Programme theory and causal complexity – common points of the theory-based impact evaluation and the qualitative comparative analysis

Before we get to presenting the specificity of the qualitative comparative analysis and the stages of the research procedure related to the application this method, it is worth paying attention to the role it may play in the theory-based impact evaluation. According to the European Commission's recommendations, the questions to be taken into account when assessing the effects of public policies may be divided into two general categories: the first applies to the quantification of these effects, the second – to their explanation⁶⁷. These two general types of the questions correspond to two general

⁶⁴ The creator of the qualitative comparative analysis, Charles Ragin, places it among the case-oriented methods, differentiating them from the variable-oriented methods. The former are characterised by the fact that they are somehow close to the cases, allowing to see their complexity and specificity. The latter, to a small extent focus on the cases (persons, programmes, organisations, enterprises, etc.), being interested only in the variables and relationships among them. Cf. Ch. C. Ragin, H. S. Becker, (ed.) *What is a case. Exploring the foundation of social inquiry*. Cambridge University Press. Cambridge, New York, 1992.

⁶⁵ The only attempt we are familiar with as regards the application of the QCA in the impact evaluation in Poland was made in the SPIN project – Model of transfer of innovation in Małopolska, implemented by the Centre for Evaluation and Analysis of Public Policies of the Jagiellonian University. As input data, the results of case studies were planned to be used. However, due to a small number of these cases (4) and a large number of the variables, the application of this technique was abandoned and the less formalised comparative analysis of cases was applied instead. More on this in: D. Szklarczyk et al. *Configurational Analysis in the Evaluation of Complex Public Programs: Application in the Area of Knowledge Transfer*. (in:) E. S.C. Berger, A. Kuckertz (ed.) *Complexity in Entrepreneurship, Innovation and Technology Research*. Springer International Publishing, 2016. pp. 371-395.

⁶⁶ In such explorations, very helpful may be the website <http://www.compass.org/>, containing both the full bibliography of the QCA method, source texts and free software used in the analysis.

⁶⁷ European Commission, *Evalsed. Evalsed Sourcebook: Method and Techniques* 2013, p. 47. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf (access on 24.02.2017)

types of research methods. The questions whether the given intervention brings the assumed effects in specific conditions, are answered, first and foremost, by counterfactual methods, geared towards isolating the specific intervention effect from all changes observed in a given group, organisation or in a given area. The evaluation approach applying these methods is referred to as the *counterfactual impact evaluation* (CIE). The second category of methods related to the theory-based impact evaluation (TBIE), is primarily oriented towards understanding how the intervention leads to the occurrence of intended and unintended effects, and how the effects of the intervention may differentiate the context in which it takes place. In case of these methods, the most important objective of the research is to answer the question: why and based on what mechanisms the intervention brings the specific effects. Of key importance is to identify the theory of change, which underlies this intervention and to assess the effects of its implementation in a given context⁶⁸.

What links these two approaches – *counterfactual impact evaluation* and *theory-based impact evaluation* – is a central position which causality occupies in them.⁶⁹ The way of understanding and analysing causality in each of these approaches is, however, different⁷⁰. In case of the CIE, of key importance is to identify a cause and effect relationship between the intervention and its effect and this identification is summarised in the answer to the question whether and to what extent the achieved effects may be attributed to the programme (*attribution*). On the other hand, in case of the TBIE the first place is taken the answer to the question whether the analysed intervention is or is not one of the causes of observed changes (*causal contribution analysis*)⁷¹. According to Stern and other researchers, we may even talk about the gradual transition from concentration on attribution-oriented questions to contribution-oriented questions.⁷² The

⁶⁸ *Ibid.*

⁶⁹ P. J. Gertler, S. Martinez, P. Premand, L. B. Rawlings, Ch. M. J. Vermeersch, *Impact Evaluation in Practice*. The International Bank for Reconstruction and Development / The World Bank. Washington 2011, p. 8.

⁷⁰ Causation itself and its understanding in the evaluation studies are currently widely discussed. Cf. e.g.: T. D. Cook, M. Scriven, C. L. Coryn, S. D. Evergreen, „Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven“. *American Journal of Evaluation* 2010, 31, pp. 105-117, S. I. Donaldson, C. A. Christie, M. M. Mark, (ed.). *What counts as credible evidence in applied research and evaluation practice?*, Sage, Thousand Oaks, CA 2009, M. Scriven, „A summative evaluation of RCT methodology: an alternative approach to causal research“, *Journal of Multidisciplinary Evaluation* 2008, 5, pp. 11-24, E. Stern, O. W. Andersen, H. Hansen, „Editorial: Special issue: What can case studies do?“ *Evaluation*, 2013. 19 (3), 213-216., E. Gates, L. Dyson, „Implications of the Changing Conversation About Causality for Evaluators“, *American Journal of Evaluation*. 2016, Volume: 38 issue: 1, page(s): 29-46.

⁷¹ Cf. J. Mayne, „Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly“, *Canadian Journal of Program Evaluation* 2001, 16(1), pp. 1-24, F. Leeuw, „Reconstructing Program Theories: Methods Available and Problems to be Solved“, *American Journal of Evaluation* 2003, 24(1), pp. 5-20.

⁷² E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, B. Befani, *Broadening the range of designs and methods for impact evaluations (Report of a study commissioned by the Department for International Development, Working paper 38)* 2012.

interest of the researchers is also increasingly oriented towards the analysis of causal mechanisms, a good example of which is the concept of realist evaluation developed by Pawson and Tilley and focused on developing and verifying the theory of how the specific mechanisms work in diverse contexts⁷³. These changes in accents have their implications for the evaluation methodology – what is observed is the increasing interest in methods and research approaches other than experimental, allowing to analyse the cause and effect relationships. Therefore, what is used on an increasingly wider scale are both non-experimental quantitative methods – such as structural modelling or econometric methods and qualitative methods allowing to explain the cause and effect mechanisms within a single case (e.g. process tracking) or within many cases (e.g. qualitative comparative analysis)⁷⁴.

As the latter is the subject of our interest, it is worth indicating the factors that make it particularly useful in the theory-based impact evaluation. These factors include:

1. The central role of the theory understood as an indication of the cause and effect relationships between the causal factors and effects of the programme. What is obvious, the theory underpins the TBIE and is in the centre of the QCA. Without developing an adequate causal model, it is difficult to carry out the analysis. Using the QCA, we may therefore test the programme theory underlying the given intervention.
2. Specific both to the TBIE and QCA striving, first of all, for identifying the mechanisms determining the success or failure of a given project, and not for measuring the size of the effects.
3. Recognising the importance of the context and local determinants as the factors determining the success or failure of the intervention. This involves questioning a possibility of the existence of the universal, context-independent programme theory. In the QCA, it is visible in seeking to identify various paths leading to the occurrence of the specific effects. Depending on the context and local determinants, these paths may be different.
4. Paying attention to a possibility of achieving the assumed effects using different activities or factors. In this context, in the QCA we talk about equifinality – indicating that various paths may lead to achieving the same effects.
5. Paying attention not only to the programme theory, but also to the implementation theory – the QCA may be used as a technique allowing to check how various factors characteristic of the way of implementing the intervention may affect its effects.

⁷³ R. Pawson, N. Tilley, *Realistic evaluation*. Sage, London, England 1997, R. Pawson, N. Tilley, *Realist evaluation*, http://www.communitymatters.com.au/RE_chapter.pdf (access on 25.02.2017).

⁷⁴ E. Gates, L. Dyson, "Implications of the Changing Conversation...", p. 3.

Multiple conjunctural causation - the basic assumption of the qualitative comparative analysis

As Gates and Dyson write, individual methods for assessing the effects and impact of public policies are based on the underlying various ways of understanding causation⁷⁵. According to these authors, we may indicate at least five various ways to formulate causation:

1. successionist framework that underlies regularity and counterfactual logics;
2. approach based on narrative stakeholder accounts;
3. generative accounts of processes and mechanisms;
4. approach analysing causal packages and contributory accounts;
5. approach taking into account nonlinear, multidirectional, and dynamical accounts of relations as found in complex systems.⁷⁶

Understanding causality underlying the QCA may be included into the third and fourth category of approaches, as mentioned above. The basic assumption of this method is the concept of the causal complexity – multiple factors which may or may not be the conditions necessary or sufficient for inducing the specific outcomes. This approach allows to analyse the role of individual elements of the intervention or configuration of these elements for achieving its effects. We may also use them to analyse the impact of individual instruments on the occurrence of specific effects – e.g. the impact of investment subsidies for enterprises and development subsidies on the increased innovation of these enterprises. It may also be helpful to identify those factors which affect the effectiveness of implementing the given intervention. Its application may thus show that, for example, the intervention is effective in these cases where its introduction was preceded by appropriate preparation of recipients and ineffective when there was no such preparation. The QCA may also be useful when we want to show, how the given mechanisms work in various contexts, leading to the occurrence of certain effects. The correct application of this analysis allows to answer such questions as: how, why, for whom and in what conditions the given intervention will lead to the occurrence of specific results. This way of thinking, specific to the third above-mentioned concept of causality and close to the realist evaluation, assumes that there are many causal conditions linking the intervention with its effect.

Such an approach to understanding of causality is a basis for the QCA. This method uses the concept of “causal complexity” or – as called by Ragin and Rihoux – multiple conjunctural causation⁷⁷. The essence of causality formulated in this way is to pay attention to the fact that various causal conditions or various causal paths may lead to the occurrence of the same outcomes. Significant is the principle of equifinality, which

⁷⁵ *Ibid.*, p. 7.

⁷⁶ *Ibid.*

⁷⁷ B. Rihoux, C. C. Ragin, (Eds.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, Sage, Thousand Oaks, CA 2009, p. 8.

means that each of these paths is equally important, i.e., that the mechanisms leading to the occurrence of the same outcomes may be different. In this understanding of causality of importance is both the concept of multiple and conjunctural - multiple refers to the number of paths, and conjunctural to the fact that each path contains a combination of conditions. A summary of the essence of the assumptions of multiple conjunctural causation, underlying the qualitative comparative analysis, is presented in Table 1.

Table 1. The most important assumptions of the concept of multiple conjunctural causation

The most important assumptions of the concept of multiple conjunctural causation

1. The most often, it is a combination of causally relevant conditions that generates the outcome ($AB \rightarrow Y$)
2. Several different combinations of conditions may produce the same outcome ($AB + CD \rightarrow Y$)
3. Depending on the context, a given outcome may result from a condition when it is present and also absent ($AB \rightarrow Y$ but also $aC \rightarrow Y$). In this example, [A] combined with [B] produces the occurrence of the outcome [Y], but its absence [a] combined with [C] also produces the outcome [Y]

Source: Rihoux B., Ragin C. C., (Eds.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, Sage, Thousand Oaks, CA 2009, p. 8.

Using the qualitative comparative analysis, the researcher does not build a single causal model which best fits the data, as it is the case for most statistical techniques, but he or she identifies the number and nature of various causal models, which occur in the compared cases⁷⁸. In addition, the qualitative comparative analysis is concentrated on explaining how the given outcomes arises and not – as it is the case of regression-based techniques – on the analysis of the net impact of the given causal condition on the occurrence of this outcome.

Qualitative comparative analysis research approach

Although the QCA developed by Ragin⁷⁹ has been applied worldwide in various research areas from the end of the eighties, in Poland it is still not popular. In the Polish literature, it is difficult to find any publications containing a mere description of the

⁷⁸ C. C. Ragin, *The comparative method. Moving beyond qualitative and quantitative strategies*, University of California Press. Berkeley, Los Angeles, London 1987, p. 19.

⁷⁹ *Ibid.*

method or examples of its application in the studies⁸⁰. This gap is undoubtedly worth filling, because this method may be a valuable complement to analytical approaches aimed at explaining the causal mechanisms of the occurring phenomena. In the further part, we will show in a synthetic way the basic methodological principles of designing research and analysis based on the assumptions of the QCA method, while explaining the most important terms used in this approach.

As it has been mentioned earlier, the qualitative comparative analysis is generally used to infer about the complex causes of specific phenomena. However, unlike the classical causal inference methods, for example, linear regression, modelling is based not on the systems of variables and interrelations among them, but on the set theory and analysis of the relationships among objects that belong to these sets and their impact on the specified output. The objective is to build an explanation of the specific phenomenon using a configuration of causal conditions which may lead to this phenomenon. Not without reason did we mention linear regression, since the QCA shares the logic of conduct with it, while adopting different resolutions in many issues. Therefore, in order to help the reader understand the QCA, we will sometimes refer to the comparisons with linear regression, as a technique which is relatively well known.

For the purposes of explaining the general principles of the QCA we will take an example in which we want to assess the effectiveness of various action programmes aimed at reintroducing economically inactive people into the labour market. The result of such actions would be, in simple terms, the success or failure⁸¹ and this would determine the affiliation to the output set – i.e. the effect of the programme would assume two values. In the course of the carried out analyses, we would like to, on a basis of certain criteria characteristic of the individual programmes, specify what determines their success. Here, we obtain the first fundamental difference when compared to regression – the latter allows to model only one specific condition of the output variable (this would be probably the success of the programme) – while the QCA is an asymmetric method and allows to infer about both outputs of the programme. This means that we may also check what determines the failure of the programme. In the QCA approach, it is even recommended to check what factors lead to each result, as it may turn out that in accordance with the principles of causal complexity, various configurations of

⁸⁰ This does not mean that this method is not known and applied in Poland at all. Occasionally, it is applied, e.g. in the studies related to management, an example of which is the publication: A. Kwiotkowska, "Qualitative comparative analysis as a methodological concept in Management Sciences", in: J. Lichtarski, S. Nowisielski, G. Osbert-Pociecha, E. Tabaszewska-Zajbert (ed.), "New directions in business management – leading trends". *Scientific work of the Wrocław University of Economics* 2014, No 340, pp. 65-77.

⁸¹ We do not go here into details of how the success or failure of the entire programme is assessed. In fact, an indicator of one or the other would be the percentage of economically inactive persons introduced into the market, while the cut-off point setting the success would have to be determined on a basis of specific criteria. We will deal with this problem when discussing the so-called calibration of the set-membership.

conditions will lead to various outcomes. In our case, various may be the causes of the failure or success of the programmes for activating people in the labour market. Then, we may assume that the analysed programmes would be assessed in terms of three characteristics – whether the persons, to which they were addressed, were offered career counselling, whether the beneficiaries were persons economically inactive for a long time (more than 24 months) or for a shorter time and whether under the programmes internships with employers were offered to these persons. As an effect, it is analysed how these three conditions, or their various configurations, contribute to the success or failure of the programme. One of the first steps in the QCA modelling is to draw up so-called “truth table”. It shows all combinations of the conditions, along with the achieved outcome. In general, in such a basic table “ones” mean the membership in the given set (the occurrence of the given condition) and “zeros” – the absence of such membership.

Table 2. Exemplary “truth table” taking into account the effects of five programs of support for economically inactive persons in a form of returning to the labour market (fictitious data)

Programme	Counselling	Period of inactivity	Internship	Effect of the programme
1	0	0	1	1
2	0	1	1	1
3	1	0	0	0
4	1	1	0	0
5	1	0	1	1

Source: own study.

In the above-presented (table) simplest solution, the membership in sets designated by the attributes of the individual factors and the programmes outcome is determined dichotomically (0-1) – the given programme offered counselling to the participants (1) or not (0), was addressed to the long-term inactive persons (1) or not (0), internship was offered (1) or not (0) and ended with the success (1) or failure (0). The objective of the inference is to define two relationships between the conditions and the output, analysed on a basis of the corresponding sets – whether the given condition (or a combination of the conditions) is sufficient or necessary for the occurrence of the outcome. We conclude that some condition or their combination constitutes a sufficient relation for the occurrence of the output when each occurrence of the factor corresponds to the occurrence of the phenomena in question (if x is y). In our example, we deal with this situation in the relationship between the internship and the success of the programme – whenever the internship was offered, the programme succeeded. In case of the remaining conditions, we cannot say that they were sufficient for the

programme to succeed. On the other hand, the condition is the necessary for the occurrence of the specific programme outcome in a situation, where it must be present for the occurrence of the given outcome – the effect occurs only if the specific condition or a combination of conditions occur (y entails x). In the case in question, such the necessary condition would also be the internship as everytime the internship was offered the programme succeeded. Therefore, it may be stated that the analysis of five programmes leads to a conclusion that the condition sufficient and necessary for the intervention to end with introducing the inactive persons into the labour market was the internship offered to the participants in the programme.

Of course, this example is a significant simplification showing the general mechanisms of the QCA, which actually give much greater opportunities to infer about causality of the analysed phenomena⁸². Using Boolean algebra with respect to the analysed sets, we may obtain a number of configurations of conditions leading to a particular output. Having regard to the asymmetry of analysis as part of the QCA we may also introduce negations of the analysed conditions or even the output itself, which is a great advantage when compared to, for example, regression modelling⁸³. In the discussed case, we analysed only so-called crisp sets and the corresponding binary factors, where the affiliation to the sets of conditions was determined based on the simple dichotomic classification – whether the specific object has a particular characteristic or not. In the extended form of the QCA, called fsQCA (derived from *fuzzy set* QCA), the membership in sets of conditions is determined gradually and may assume many levels, though the sequence of three, four or five elements is applied. In our case, such a fuzzy set of the condition “period of economic inactivity” could take, for example, three forms: programme addressed to persons inactive in the labour market less than a year, from one to two years, and more than two years.

For the practical application of the QCA, in the causal inference we can use numerous computer programmes developed for this purpose. Most of them are free, for example, we may mention here fsQCA developed by Ragin himself⁸⁴. Its developmental version applied for multivariate result sets, that is, those allowing the inferred result to take

⁸² For detailed explanations of various kinds of relationships that may exist between the conditions and the output – including so-called INUS and SUIN conditions, please refer to: C. C. Ragin, *Redesigning Social Inquiry Fuzzy Sets and Beyond*, University of Chicago Press, Chicago and London 2008, C. Q. Schneider, C. Wagemann, *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis*, Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Mexico City 2007.

⁸³ Although, regression also may include combinations of independent variables as interactive factor, but in this combinations with even several variables, the interpretation of such interactive factor is very difficult.

⁸⁴ C. C. Ragin, S. Davey, *fs/QCA: Fuzzy-Set/Qualitative Comparative Analysis Version 2.5* [Computer Program], Department of Sociology, University of Arizona, Tucson 2009, <http://www.u.arizona.edu/~cragin/fsQCA/software.shtml> (access on 25.05.2017)

more than two forms, is *Tosmana* developed by Cronqvist⁸⁵. In addition to this, there are appropriate packages for the R environment and Stata⁸⁶.

QCA Procedure

In the further part, we are presenting the QCA procedure, step by step, while paying attention to key moments of the carried out analyses and then we are presenting the results that can be obtained using this technique.

Table 3. Steps in the QCA

Step 1: Construction of the model on a basis of relevant cases and causal systems
Step 2: Preparing the database along with the calibration of the sets of conditions (encoding)
Step 3: Developing the truth table and checking the results for consistency and contradictions
Step 4: Analysis of the truth table using algorithms minimising the causal conditions
Step 5: Evaluation of the results obtained
Step 6: Return to step 3 and repeating the analyses for the negation of the outcome set

Source: own study.

Before starting the analysis, as in case of each modelling, it is, naturally, required to embed the entire procedure in a specific theoretical context i.e. to find a justification for the inference carried out. Most commonly, we start the analysis with building the model, but in contrast to classical regression modelling, much more flexible, qualitative approach is required here. It is about the proper selection of cases to be analysed, so it is possible, on their basis, to explain the specific outcomes by means of causal conditions. The big advantage of the QCA, as a mixed approach, is the possibility of the simultaneous use of qualitative and quantitative data. For example, in order to explain the determinants of the innovation of enterprises using the QCA, we may use quantitative data collected during structured interviews and complete them with information obtained in the course of in-depth interviews with representatives of these firms. Into this analysis, we may also include the data from available public statistics informing about the condition of the firms in question. The inference should be based on the

⁸⁵ L. Cronqvist, *Tosmana: Tool for small-n analysis, Version 1.522* [Computer Program], University of Trier, Trier, Germany 2016, <https://www.tosmana.net/> (access on 25.05.2017).

⁸⁶ A review of various solutions may be found in: A. Thiem, A. Duşa, "Boolean minimization in social science research: A review of current software for qualitative comparative analysis (QCA)", *Social Science Computer Review* 2013, 31, pp. 505-521.

properly selected cases from all available. Nothing stands in the way to analyse all existing cases, however, it should be remembered that they must be adequately diversified. In the situation of the above-mentioned analysis of the causes of the innovation of companies, if the majority of the cases included the firms that are innovative, this absence of diversifying the set of the cases selected in this way would make it difficult or even impossible to carry out the analysis. This is a situation identical to regression modelling, in which it is undesirable for the dependent variable to be characterised by the low variance (diversity of values). In this first step, in addition to selecting appropriate cases to be analysed, we should also build a suitable causal model, which will be a basis for selecting the conditions likely to affect the tested outcome. The type of these conditions will enable their appropriate calibration in the next step. Building of this analytical model may be carried out in two ways. Firstly, we can do this using the aprioric method, starting with theory and deriving from it the relationships to be tested later on. Secondly, we may also use the exploration method and only at the stage of the analysis itself, build the causal model. In practice, as a rule, these two ways are combined – we make certain assumptions arising from the nature of the examined phenomena, and in the course of the analysis we check what conditions best explain the causes of these phenomena.

The calibration of the conditions determining the outcome, and of the outcome itself, is simply their encoding according to the specific rules. This step will look different when we want to carry out the QCA inference based on crisp sets (csQCA) and different in case of fuzzy sets (i.e. fsQCA). We should recall that in this first case, the set-membership is dichotomic – i.e. the conditions in question may or may not have the feature determining the membership in this set. For example, one of the conditions which may affect innovation of firms is to have a research and development department. Then, this factor shall assume two values: 1 – having such a department, or 0 – no such department. In case of fuzzy sets, the membership in the set is gradable – the given object may belong to this set to a greater or smaller extent or not to belong to it at all. When carrying out the inference based on fuzzy sets, we usually apply three, four or five levels of gradation of the membership in the set. In case of three levels, they are encoded as 1 – full membership, 0.5 – partial membership and 0 – no membership. However, when we want to apply four level, the values which may be conventionally defined as follows, are possible: 1 – full membership, 0.67 – medium membership, 0.33 - small membership and 0 – no membership. Applying three or four levels depends on the data that we have and the nature of the factor itself. For example, such a fuzzy set, which can be a condition of innovation of enterprises, is the size of the firm. We may then encode this information according to the classification of size of business into micro-enterprises (up to 10 employees) as “0”, small companies (10-49 employees) as “0.33”, medium (50-249 employees) as “0.67” and finally large companies (more than 250 employees) as “1”. For some reasons, most often resulting from the analyses carried out, we could decide to leave only three levels, for example, as a

result of combining micro and small companies and encoding them as “0” and leaving medium companies as “0.5” and large companies as “1”. In this case, the calibration of this condition is relatively simple, to some extent it naturally results from the nature of the data themselves. Slightly different is the case of quantitative (continuous) characteristics. Then, we may use the distribution of such data to determine the span of its values, and on a basis of certain criteria the required thresholds. The designation of these thresholds should not be automatic, but based on some particular theoretical model – e.g. the breakdown of age into cohorts according to the applicable standard. If there is no such conceptual reference point, we may apply the population distribution of characteristics based on reliable public statistics.

It is much more difficult to calibrate the fuzzy set based on qualitative data. In this situation, we cannot in fact use empirical distributions of characteristics and this encoding should be carried out based on a well-thought out conceptual scheme. A good procedure has been proposed in this case by Basutro and Speer (Table 3)⁸⁷. It is based on six steps shown below. This procedure has been developed for the purposes of the qualitative data collection procedure using in-depth interviews, but it may be equally well applied to encode qualitative secondary data.

Table 4. Procedure for calibration of qualitative data

Step 1: describing the operationalisation of the conditions and the outcome used in the model

Step 2: developing preliminary anchor points for the created sets

Step 3: encoding the data

Step 4: checking the consistency of the codes and their adjustment to the sets

Step 5: assessing the precision of the sets and defined values of thresholds

Step 6: final assigning the values to the encoded data

Source: X. Basurto, & J. Speer, “Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA)”, *Field Methods* 2012, 24, pp. 155-174

When summarising the recommendations regarding the data calibration in the QCA, it should be stressed that this process requires consideration and should be based on the accepted theoretical or conceptual model and to correspond to the needs of the model being built. The good model significantly facilitates select-

⁸⁷ X. Basurto, & J. Speer, “Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA)”, *Field Methods* 2012, 24, pp. 155-174.

ing the conditions themselves, but also assigning the appropriate levels to them. When calibrating the conditions, we should avoid too much automaticity, because the way in which we will prepare the data for analysis to a large extent determines the results obtained. The more that when analysing the small number of cases, any change in the data may significantly change the entire model. It is worth trying out various configurations of conditions and their levels, and seeing how they affect the results of the inference. In any case, however, it is important to act in a transparent manner and well describe the adopted calibration policies, so as to enable the assessment of the entire procedure and checking its relevance and reliability.

After preparing the data, we may start the analysis. The first activity is to prepare the truth table . As we wrote above, the table contains all possible causal conditions and corresponding outcomes. The number of the rows in this table depends on the number of conditions used to build the model. Before starting to minimise the causal conditions, we should check this table for so-called logical remainders and contradictory rows. The logical remainders are simply such configurations of conditions (i.e., the rows in the truth table), which do not exist in the analysed cases – they are empirically empty. The contradictory rows appear in the situations when the layout of the conditions in the row of the table leads to the result that may be simultaneously true or false (1 or 0), giving inconsistent results. To diagnose such situations, we use the measure of consistency, which in brief is an estimate of how the given configuration of conditions deviates from the ideal solution represented by the result equal to 0 or 1. It is also assumed that the contradictory rows are those whose consistency is lower than 0.75⁸⁸. It is recommended to treat the issues of the logical remainders and contradictory rows individually and before making a decision about their exclusion from the table to verify them based on an analysis of individual cases – whether such situations are possible and what is behind them. There are also three general strategies which can help in case of encountering such issues. It is about: 1) adding the cases for analysis – which may feed the data into the logical remainders and remove inconsistencies, 2) adding an additional condition or dividing one condition into several ones, 3) changing the calibration of the conditions or outcome. These recommendations are similar to those encountered in the problems of the statistical analysis of data, e.g., in the data missing or small frequencies.

⁸⁸ Detailed explanation of consistencies and contradictory rows may be found in: C. Q. Schneider, C. Wagemann, *Set-Theoretic Methods...* 2007.

When the truth table is already cleared, we may start the proper analysis, i.e. minimising this table by using the algorithms of Boolean algebra in case of the csQCA or Quine-McCluskey algorithm⁸⁹ when we carry out the fsQCA. Minimisation reduces all causal conditions to a simplified form, combining and excluding the similar or contradictory conditions. To make the functioning of the QCA more familiar, it is enough to say that as a result of these analyses, we obtain three solutions (causal recipes): complex, parsimonious and intermediate. In the complex solution, the final causal recipes are presented, for the calculation of which all causal conditions are taken into account, including those conditions that are not covered by empirical data (that is, taking into account the logical remainders). In the parsimonious solution, just the opposite, the logical remainders are excluded from the analysis, which leads to simpler causal recipes. And finally, the intermediate solution which is usually recommended as the best one, takes into account the logical remainders but in accordance with the intention of the user (he decides which of these conditions should be included).

Finally, we get the causal recipes, on a basis of the conditions included in the model, which indicate the causal conditions leading to the occurrence of the analysed outcomes. For example, in case of the parsimonious solution for the causes underlying the development of enterprises, we may obtain the following recipe:

$$\begin{aligned}
 & \text{R\&D} * \sim \text{PUBLIC SUPPORT} \\
 & + \text{SIZE} * \text{STRATEGIC MANAGEMENT} \\
 & + \text{R\&D} * \text{RISK TAKING} \\
 & \rightarrow \text{DEVELOPMENT}
 \end{aligned}$$

It should be understood in this way that to the development of companies, three alternative causal conditions may lead: 1) having a research and development department (R&D) by companies and not using public support (PUBLIC SUPPORT, where “~” means the negation of the condition), 2) in case of large companies (SIZE) having strategic management (STRATEGIC MANAGEMENT – thinking in terms of future) or 3) having a research and development department (R&D) and the willingness to take risks (RISK TAKING). The results, in particular where there are many causal recipes, may also be presented in a form of tables containing the information about the contribution of the condition to the individual recipes. An example of such the result table, obtained in the analysis of the causes for the development of companies has been shown below (Table 5).

⁸⁹ The algorithms developed by Quine and McCluskey algorithm are considered to be the most efficient ways to make calculations on a basis of the principles of Boolean algebra. They have been developed in order to effectively use computer techniques for minimising the truth function (cf. W., V., Quine, “A way to simplify truth functions”. *American Mathematical Monthly*, 1955, 62, 627–631; E., J., McCluskey, “Minimization of Boolean functions”. *Bell Systems Technical Journal*, 1956, 35, 1417–1444).

Table 5. Exemplary table of the causal recipes of the intermediate solution explaining the causes for the development of companies.

Factors	Solution 1	Solution 2
Company size (<i>SIZE</i>)	●	○
Strategic management (<i>STRATEGIC MANAGEMENT</i>)	●	○
Human capital management (<i>HC MANAGEMENT</i>)	●	●
Networking	●	
R&D (<i>R&D</i>)	●	○
Risk taking (<i>RISK TAKING</i>)	○	●

Note: ● = core causal condition (present); ○ = core causal condition (absent); ● = contributing causal condition (present); ○ = contributing causal condition (absent).

The presented exemplary explanation of the causes for the development of companies included two intermediate solutions. They consist of the number of conditions that affect, to a varying degree, the result of the inference. Based on the inference logic, we may conclude that some of them will occur both in the parsimonious and intermediate solution (more specifically, all conditions from the parsimonious solution must be included into the intermediate solution). Such conditions can be called core conditions and are marked in the table with the larger circles. On the other hand, the other conditions are added at the stage of obtaining the intermediate solution and are contributing to the causal recipes created in this way, therefore, were marked with the smaller circles⁹⁰. On the basis of two obtained solutions, it may be concluded that the first applies to the situation of large companies (the presence of the condition), the development of which is contributed to, first of all, by having a human capital management strategy and building networks of relationships with customers and other stakeholders. The contribution of the vision of strategic management and having an R&D departments is lower. Importantly, in this solution in the development of large companies the tendency to risk taking is not advisable. In turn, this element occurs in the second solution, applicable to small firms (the absence of condition), which, in order to develop, do not need to have an R&D department or to demonstrate strategic thinking, however, they manage human capital. It can therefore be said that a major difference in the causes for the development of companies would be their size, which determines two paths (recipes) – both based on human capital management, but in case of small companies, the recipe is based on the willingness to risk taking and in case of large ones – on having more formal action strategies.

⁹⁰ A detailed description of this method to present the QCA results may be found in: C. C. Ragin, P. C. Fiss, "Net effects analysis versus configurational analysis: An empirical demonstration", In C. C. Ragin (Ed.), *Redesigning social inquiry: Fuzzy sets and beyond*, University of Chicago Press, Chicago 2008, pp. 190–212.

Summary

The qualitative comparative analysis is one of mixed methods i.e. those combining the qualitative and quantitative approach. It may be applied in causal analyses which are so important in the impact evaluation. Therefore, undoubtedly it is worth being recommended to evaluate the public interventions, particularly, when we carry out the impact evaluation and we want to check how the theory of the programme or theory of the implementation work in specific conditions. Putting an emphasis on the theory – either the theory of the programme or theory of the implementation – is a common point of the TBIE and QCA, and the application of both approaches requires the evaluator to develop a model linking the causal factors with their effects. The QCA may be applied anywhere where we want to verify the accuracy of the adopted theory of the programme, to verify its adequacy to the given context, to test how the effectiveness of the intervention is affected by the way of its implementation. Unlike the classical statistical techniques allowing to determine the net effects of the intervention, the QCA allows to reconstruct the model of complex causal pattern leading to specific results. It is also important that as a result of applying the QCA we may identify various causal paths leading to the occurrence of the same effects – for example, different may be the paths leading to the development of companies operating in different sectors of the economy or in different countries.

It is worth stressing that the analyses of the causes of successes but also failures of public activities, carried out using the QCA, do not have to involve large financial or time inputs. Analyses are carried out, in fact, on small and medium-sized samples, from 5 to 50 cases. This does not require carrying out the extensive and costly studies on large, representative samples. For the analysis, we may also use the data of various nature – they may be the information derived from public statistics, qualitative data obtained during in-depth interviews or data collected by means of survey techniques. Also computer software used for the QCA modelling is commonly available and relatively easy to use. All of this makes this method a very useful tool for the evaluators who evaluate the impact of the given intervention, even if they do not have a large budget or a specialised research team.

The qualitative comparative analysis, despite its numerous advantages, also puts before the researchers the challenges that should be kept in mind when we decide to use this analysis. First of all, despite the relative simplicity of using this tool, it requires quite a lot of experience. This is particularly true for the selection of cases for analysis, building the causal model itself and encoding (calibrating) the factors included into the model. In addition, despite the precise inference rules based on the principles of logic and mathematical algorithms, the results of this analysis are not universal. As opposed to the classical statistical models, they cannot be generalised to broader populations. However, in many cases, getting familiar with

the system of causes leading to the specific result may be more valuable from the perspective of assessing the intervention logic than the estimation of the net effects of specific factors.

About the authors

Barbara Worek

Assistant Professor in the Department of Sociology of Economy, Education, and Social Research Methods at the Institute of Sociology of the Jagiellonian University in Cracow and an expert of the Jagiellonian University's Centre for Evaluation and Analysis of Public Policies. She has many years of experience both in conducting evaluation studies and training on evaluation. For several years, at the Jagiellonian University she has been conducting classes dedicated to evaluation and a seminar on the use of qualitative methods in evaluation and analysis of public policies. When conducting evaluation studies, she worked with many public administration institutions (at the central and regional level). She gained research experience by participating in Polish and international projects on social exclusion, labour market and lifelong learning. She is an author and co-author of publications on the labour market, human capital and evaluation, and of numerous research reports. Her scientific interests focus on the qualitative research methodology, human capital development and analysis of public policies. Since 2009, she has been involved in research within the framework of the PARP project "Human Capital Balance" where she is responsible for the issues related to adult learning, activity of the training and development sector and employers investing in human resources development.

Marcin Kocór

Assistant Professor in the Department of Sociology of Economy, Education, and Social Research Methods at the Institute of Sociology of the Jagiellonian University in Cracow and an expert of the Jagiellonian University's Centre for Evaluation and Analysis of Public Policies. For 18 years, he has been involved in marketing and social research. Over that period, he gained extensive research experience by carrying out projects, *inter alia*, for the Polish Agency for Enterprise Development (expert for examination of employers in the PARP project "Human Capital Balance"), Voivodeship Labour Office in Cracow (pilot studies of the Integrated Qualification System). He shares his experience by conducting, for 16 years, teaching classes covering the social and marketing research methods, research design and data analysis techniques. His scientific interests focus on the research methodology of social sciences, statistical data analysis techniques and issues of labour market and human capital.

Bibliography:

1. Babbie, E., *The Practice of Social Research*. PWN, Warsaw 2004.
2. Basurto, X., Speer, J., "Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA)", *Field Methods* 2012, 24, pp. 155-174.
3. Cook T. D., Scriven M., Coryn C. L., Evergreen, S. D. „Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven". *American Journal of Evaluation* 2010, 31, pp. 105-117.
4. Cronqvist, L., *Tosmana: Tool for small-n analysis, Version 1.522* [Computer Program], Trier, Germany: University of Trier 2016, <https://www.tosmana.net/> (access on 25.05.2017).
5. Donaldson S. I., Christie C. A., Mark, M. M. (eds.). *What counts as credible evidence in applied research and evaluation practice?*, Sage, Thousand Oaks, CA 2009.
6. European Commission, *Evalsed, Evalsed Sourcebook: Method and Techniques* 2013, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf (access on 24.02.2017).
7. Gates E., Dyson L., „Implications of the Changing Conversation About Causality for Evaluators", *American Journal of Evaluation*, 2016, Volume: 38 issue: 1, page(s): 29-46.
8. Gertler P. J., Martinez S., Premand P., Rawlings L. B., Vermeersch Ch. M. J., *Impact Evaluation in Practice*. The International Bank for Reconstruction and Development / The World Bank, Washington 2011.
9. Kwiotkowska A., „Jakościowa analiza porównawcza jako koncepcja metodologiczna w naukach o zarządzaniu", in: Lichtarski J., Nowisielski S., Osbert-Pociecha G., Tabaszewska-Zajbert E. (ed.) *Nowe kierunki w zarządzaniu przedsiębiorstwem - wiodące tendencje*, Scientific work of the Wrocław University of Economics 2014, No 340, pp. 65-77.
10. Leeuw F. „Reconstructing Program Theories: Methods Available and Problems to be Solved", *American Journal of Evaluation* 2003, 24(1), pp. 5-20.
11. Mayne J., „Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly", *Canadian Journal of Program Evaluation* 2001, 16(1), pp. 1-24.
12. McCluskey, E., J., "Minimization of Boolean functions", *Bell Systems Technical Journal*, 1956, 35, pp. 1417-1444.
13. Quine, W., V., "A way to simplify truth functions", *American Mathematical Monthly*, 1955, 62, pp. 627-631.
14. Pawson R., Tilley N. *Realist evaluation* 2004, http://www.communitymatters.com.au/RE_chapter.pdf (access on 25.02.2017).
15. Pawson R., Tilley N. *Realistic evaluation*. Sage, London, England 1997.
16. Ragin Ch. C., Becker H. S., (eds.) *What is a case. Exploring the foundation of social inquiry*. Cambridge University Press. Cambridge, New York, 1992.

17. Ragin, C. C., *Redesigning Social Inquiry Fuzzy Sets and Beyond*, University of Chicago Press, Chicago and London 2008.
18. Ragin, Ch. C., *The comparative method. Moving beyond qualitative and quantitative strategies*. University of California Press. Berkeley, Los Angeles, London 1987.
19. Ragin C. C., Fiss, P. C., "Net effects analysis versus configurational analysis: An empirical demonstration" in: C. C. Ragin (Ed.), *Redesigning social inquiry: Fuzzy sets and beyond*, University of Chicago Press, Chicago 2008.
20. Ragin, C. C., & Davey, S., *fs/QCA: Fuzzy-Set/Qualitative Comparative Analysis Version 2.5* [Computer Program], Department of Sociology, University of Arizona, Tucson 2009, <http://www.u.arizona.edu/~cragin/fsQCA/software.shtml> (access on 25.05.2017)
21. Rihoux B., Ragin Ch. C., (Eds.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks, CA: Sage 2009.
22. Schneider, C. Q., Wagemann, C., *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis*, Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Mexico City 2007.
23. Scriven M., „A summative evaluation of RCT methodology: an alternative approach to causal research". *Journal of Multidisciplinary Evaluation*, 2008, 5, pp. 11-24.
24. Stern E., Andersen O. W., Hansen H., "Editorial: Special issue: What can case studies do?", *Evaluation*, 2013, 19 (3), 213-216.
25. Stern E., Stame N., Mayne J., Forss K., Davies R., Befani B. *Broadening the range of designs and methods for impact evaluations (Report of a study commissioned by the Department for International Development, Working paper 38)* 2012. <https://www.oecd.org/derec/50399683.pdf> (access on 25.02.2017).
26. Szklarczyk D., Krupnik S., Strycharz J., Antosz P., Drożdżak Z., Łukasiewicz K., Szczucka A. *Configurational Analysis in the Evaluation of Complex Public Programs: Application in the Area of Knowledge Transfer*. (in:) Berger, E. S. C., Kuckertz A. (ed.) *Complexity in Entrepreneurship, Innovation and Technology Research*. Springer International Publishing, 2016. pp. 371-395.
27. Thiem, A., Duşa, A., "Boolean minimization in social science research: A review of current software for qualitative comparative analysis (QCA)", *Social Science Computer Review* 2013, 31, pp. 505-521.

Testing intervention theory using Bayesian Network Analysis: evidence from a pilot exercise on SMEs support⁹¹

Introduction

This paper deals with a pilot exercise carried out in the context of *ex post* evaluation of the European Regional Development Fund (ERDF) 2007-2013 programmes. It focuses on assessment of one of policy instruments under the Regional ERDF 2007-2013 Operational Programme Apulia. The instrument targeted a high number of micro and small enterprises, pursuing broad objectives and financing many different kinds of actions. We utilized a Theory-Based Impact Evaluation approach and we applied the Bayesian Network Analysis to investigate changes resulting from the said policy instrument, which are usually more challenging to identify. Due to the large variety of objectives, we argue that this instrument is not particularly productive in triggering long-term structural changes, while it has been crucial for supporting the survival of many micro and small enterprises at the time of economic crisis. Thanks to the mechanisms detected through the Bayesian Network Analysis it has been possible to identify particular patterns of change at the enterprise level. We argue that the effectiveness of this instrument could have been enhanced by placing more restrictions on the selection of sectors and investments, and by narrowing the targeted enterprises to those with the highest growth potential.

We discuss the application of a Theory-Based Impact Evaluation approach to assess *ex post* the effectiveness of support provided to SMEs under the Cohesion Policy programmes in 2007-2013. In particular, we test the use of an innovative methodology, namely the Bayesian Network Analysis (BNA), to assess the validity of intervention

⁹¹ This paper draws extensively from the study *Ex post evaluation of Cohesion Policy programmes 2007-2013 financed by the European Regional Development Fund (ERDF), and Cohesion Fund (CF) - Work Package 2: Support to SMEs - Increasing research and innovation in SMEs and SME development*. (Contract Number: 2014CE16BAT002) carried out by CSIL, selected by the Evaluation Unit, DG Regional and Urban Policy, European Commission, through a call for tenders by open procedure. This is one of sixteen studies dedicated to *ex post* evaluation of the ERDF and Cohesion Fund, and to investigating the key outcomes of Cohesion Policy in 2007-2013; it is also one of three studies dealing with enterprise support (the remaining two being on Financial Instruments -WP3- and Large Enterprises -WP4).

The authors are grateful for the very helpful insights from the EC staff and from other members of the Steering Group. The authors are responsible for any remaining errors or omissions. Intermediate and final reports carried out as a part of this study are available at the following website: http://ec.europa.eu/regional_policy/en/policy/evaluations/ec/2007-2013/#3 (access on 25.05.2017)

logic and to uncover the mechanisms of change and the outcomes generated by the policy instrument examined.

A vast strand of literature that investigates the effects of public interventions aimed at supporting enterprises relies on the econometric analysis or counterfactual techniques⁹². The logic underlying these methods is to determine the net impact that can be attributed to a given policy action. However, such methods do not explicate why a certain instrument has or has not worked, and which causal mechanisms can explain how the observed impact has been actually achieved. Such mechanisms are generally related to behavioural responses stimulated by the policy impulse and to external context characteristics.

For policy makers willing to learn from past experience and to improve policy design, gaining knowledge about the mechanisms that explain why an instrument achieved its goals is as important as finding quantitative evidence of its effectiveness. To this end, there is a need to look inside the 'black box' of the SMEs, examining if and how a given policy instrument succeeds in addressing the elements constraining SMEs' capacity to innovate and grow, and in stimulating a behavioural change⁹³.

A theory-based impact evaluation (TBIE) was developed in response to some of the limitations of counterfactual methods. TBIE is a well-established methodology⁹⁴ that makes explicit the underlying logic (or theory) of the intervention under assessment, and explores the assumptions and causal relations that determine the generation of certain effects, whether desired or undesired, expected or unexpected. It offers a valuable opportunity to explore why and how a given intervention has or has not generated a given effect. It consists, firstly, in reconstructing the intervention logic behind a policy instrument as designed by policy makers, and subsequently in testing the theory in order to reject or confirm the expectations about the mechanisms of change and the outcomes to be generated by the policy instrument.

We illustrate the application of a TBIE approach to an *ex post* evaluation of a policy instrument under the Regional ERDF 2007-2013 Operational Programme Apulia, namely

⁹² D. Mouqué, "What are counterfactual impact evaluations teaching us about enterprise and innovation support?", *Regional Focus* 2012, vol. 2.

⁹³ S. Massa, S. Testa, "Innovation and SMEs: Misaligned perspectives and goals among entrepreneurs, academics, and policy makers", *Technovation* 2008, vol. 28, no.7, pp. 393-407.

⁹⁴ B. Astbury, F.L. Leeuw, "Unpacking black boxes: mechanisms and theory building in evaluation", *American Journal of Evaluation* 2010, vol. 31, no.3, pp. 363-381

C. H. Weiss, "How can theory-based evaluation make greater headway?", *Evaluation Review* 1997, vol. 21, no. 4, pp. 501-524.

S. Carvalho, H. White, "Theory-based evaluation: the case of social funds", *American Journal of Evaluation* 2004, vol. 25, no. 2, pp. 141-160.

L. Blackman, S. Reich, "Randomized control trials: a gold standard with feet of clay?", in: S. Donaldson, C. Christie, M. Mark (eds.), *What counts as credible evidence in applied research and evaluation practice?*, Sage, Thousand Oaks, CA 2009.

the so-called TITLE II, which provided investment aid for micro and small enterprises, active from 2009 to 2014. The instrument was a combination of an interest subsidy and a grant for micro and small enterprises that had incurred a bank debt to start investment projects of different sorts. This study is dedicated to one of three pilot exercises on individual policy instruments, the other two being the Technological Credit implemented under Polish Innovative Economy Operational Programme 2007-2013 and the Support for industrial R&D and Innovation granted within the framework of regional ERDF Operational Programme for Castile and Leon in Spain⁹⁵. While the Polish and Spanish instruments were more ambitious in terms of objectives pursued and, to some extent, more innovative in the way they were conceived and implemented, the Apulian instrument is a more traditional support scheme and is representative of a number of policy instruments financed in the 2007-2013 programming period and aimed to help SMEs overcome the negative effects of economic crisis.

The novelty of our approach consists in the use of statistical analysis and Bayesian Networks to test the theory of intervention and uncover the mechanisms of change as well as the outcomes generated by the policy instrument in question. Bayesian Network Analysis (BNA) has been found particularly useful for analysing the complexity of the object of our study. BNA was used in combination with regression models to check the statistical significance of correlations among variables. It was crucial for testing the theory properly and for finding the hidden or unexpected mechanisms of change. In combination with other analytical methodologies, the Bayesian Network Analysis brought about robust results, leading to a valid conclusion if and in what way the said policy instrument is effective⁹⁶.

Our paper is structured in five sections. In section 2, we describe the policy instrument TITLE II. In section 3, we present the Bayesian Network methodology and its usefulness for addressing the selected case study. In section 4, we discuss the results of evaluation of the policy instrument, dividing them into results in terms of economic performance, enterprises behavioural change and mechanisms underlying changes detected through the Bayesian Network Analysis. In section 5, we draw conclusions.

⁹⁵ The pilot exercise on three selected policy instruments was specifically requested within the remit of the evaluation, which included many other activities (in particular: a literature review, a documentary review of 50 selected Operational Programmes and their Annual Implementation Reports, 8 in-depth case studies with extensive field work on selected Operational Programmes, and a seminar with stakeholders, experts and representatives of Managing Authorities). The three policy instruments were selected on the basis of case studies reports for representativeness (pursuing different and typical policy objectives in the field of SMEs support) and feasibility reasons (the number of beneficiaries and data availability, especially the availability of final beneficiaries contact details).

⁹⁶ D. Heckerman, D. Geiger, D. M. Chickering, "Learning Bayesian networks: the combination of knowledge and statistical data", *Machine Learning* 1995, vol. 20, pp. 197–243.

Context

The region of Apulia in the South of Italy is classified as Convergence region, and it is characterised by a prevalence of micro enterprises operating in traditional sectors, most of which are individual firms in the 0-2 persons employed class. The financial and economic crisis caused a significant fall in the level of growth and employment. At the end of 2013, the number of unemployed in the region was more than 70% higher than in 2007⁹⁷.

As a consequence of the global recession and in order to restrict the national public deficit, national budget co-funding Cohesion Policy, and also national policy instruments targeting enterprises' support, were severely cut, especially in the Southern regions. Under such circumstances, regional funds were called upon to offset the trend. The goal of the regional Operational Programme ERDF 2007-2013 was twofold: to promote R&D and innovation with more selective instruments targeting excellence in innovation and supporting more ambitious investment plans and structural change, on one hand, and, on the other hand, to support the competitiveness of the productive system and employment, typically addressing more generic and small scale investment projects.

Among the latter group there is the policy instrument object of our study, the so-called TITLE II, which provided aid to investment for micro and small enterprises (below 50 employees, as per the EC definition). The instrument consisted of a combination of an interest subsidy and a grant to micro and small enterprises that had incurred a bank debt to start investment projects of different sorts⁹⁸. The TITLE II received the highest volume of committed public funds among the set of SME-related instruments of the OP (almost EUR 120 million). It was launched in April 2009 and remained open until June 2014 with a one-stop-shop approach; the eligible sectors have been widened over the years⁹⁹. The instrument could finance six types of expenses: purchase of land; cost of construction and renovation works; purchase of infrastructure, purchase of machinery, equipment and vehicles; purchase of computer programs; purchase of patents and license rights. In general terms, TITLE II was supposed to promote business modernisation activities, including both generic types of expense (e.g. renovation works or purchase of new equipment, including furniture for commercial or adminis-

⁹⁷ European Commission, *Support to SMEs – Increasing Research and Innovation in SMEs and SME Development. Third Intermediate Report - Work Package 2*, submitted by CSIL in partnership with CSES and ZEW, Brussels 2015.

⁹⁸ *Ibid.*

⁹⁹ The call indicated at first the following sectors as eligible: craft enterprises, commerce enterprises, including retailers and food service enterprises (e.g. bar and restaurants); other firms in the manufacturing, construction and information and communication sectors. During the subsequent months of the same year new categories of activity within the commercial sector, and some health and social work activities (e.g. 'services for nursery schools') were added. The perimeter of the eligible businesses was progressively enlarged in the following years, ultimately also including activities in the green sector (e.g. plastic recycling) (European Commission 2015).

trative spaces), and others more geared to innovation (e.g. purchase of patents for the introduction of innovative production processes). The criteria defining the aid thresholds were revised on different occasions during the programming period with the aim to make the instrument more attractive for enterprises. The aid consisted of a grant intended to pay back part of the interest on the debt incurred to finance the investment. The amount of aid was computed as a percentage of the interest due on the incurred credit. Micro enterprises could benefit from an additional grant, computed as a share in the costs of machinery and equipment. Seven months after the launch of the instrument, this contribution was extended also to small enterprises. The maximum aid intensity (share of the total public contribution in the investment volume) increased over the years, going from 40% and 30% for micro and small enterprises respectively in April 2009, to 45% for both the types of enterprises at the beginning of 2011.

A local agency acted as the intermediate body managing the distribution of public incentives to enterprises. It was in charge of screening the applications received, mainly checking coherence of the financial plan with the expenses. Applicant firms had to include indicators on the employment (distinguishing between male and female employees) at the time of applying for TITLE II and expected to be generated thanks to investment by the end of investment project. However, these indicators were not binding and were not used to rank the projects. Investment projects were selected or rejected on a one-by-one basis, with no competition between different projects.

Bayesian Network Analysis

In this section we introduce the tool called Bayesian Network Analysis (BNA) and explain the reasons for its usefulness in the context of this study. BNA is an approach that combines a graphical map analysis with statistical analysis in order to show the connexions linking variables. A Bayesian Network illustrates the probabilistic relationship among a set of variables and their conditional independences. It also provides a compact representation of a joint probability distribution¹⁰⁰. Bayesian Networks are defined by a network structure, that is to say the directed acyclic graph (DAG), and a set of conditional probability distributions associated with the variables entering the DAG. The DAG is a set of random variables represented by nodes. The role of the network is to express the conditional independence relationships among the variables in the model through graphical separation. The hierarchical positioning of variables (that implies that A is linked to B, which is linked to C, which is linked to D) can be interpreted as a conjecture of causality between these variables. Causal relations, however, have to be validated by the analyst on the basis of prior knowledge on the variables. If the causal relation is not known, the nodes of the network are connected with each other without a specific direction.

¹⁰⁰ K. Murphy, *A Brief Introduction to Graphical Models and Bayesian Networks*, 1998. M. Horný, "Bayesian Networks", *Technical Report No.5*, Boston University 2014. <http://www.bu.edu/sph/files/2014/05/bayesian-networks-final.pdf> (access on 25.05.2017)

We highlight several advantages of the use of BNA in the context of this study¹⁰¹. First of all, BNA is helpful when the goal of the study is to identify the multiple and interconnected determinants of an outcome, for example a particular behaviour, and to understand in which way they are linked one to the other. Given the complex nature of SMEs, it is difficult to know a priori the causal mechanisms that link all the possible explicatory and dependent variables under investigation. BNA helps to untangle the complex relationships and to reveal the underlying and, at least partially, unknown causal system. It is possible to verify, based on available data, the existence of confounders, that is variables that are both dependent and independent variables in the statistical model, and to estimate the conditional probabilities for all the variables of the model. In this regard, it is important to highlight that the Bayesian Networks do not necessarily show the same correlations that emerge from regression models. Regression analysis identifies those variables that more significantly influence a previously defined dependent variable. Differently, the BNA shows which variables are independent from each other, based on their conditional probabilities of occurring. For example, whereas a regression shows that economic performance of SMEs depends on the amount of public support received, the BNA might show instead that economic performance depends on the type of change that occurred in the enterprise production function (for instance the acquisition of new production technologies), which in turn is associated with the public support. Hence, economic performance and public support are independent of each other in the network, once it is controlled for other variables.

Second, the BNA can be used for both predictive and diagnostic inference support¹⁰². The predictive support ('top-down reasoning') is based on extracting evidence from the parent variable and using it to predict the pattern of child variables, namely its posterior conditional probability distribution. The diagnostic support, or 'bottom-up reasoning', works in the opposite direction: based on the evidence on child variables, the BNA can be used to analyse the distribution of a parent variable. For example, investigating a possible relationship between the acquisition of new production technologies (parent variable) and the economic performance (child variable) can be done either by predictive reasoning (i.e. we observe a change in production technologies and update our knowledge about the influence of such change on economic performance) or diagnostic reasoning (i.e. we observe a change in economic performance and update our knowledge about the possible cause, by analysing the distribution of the production technology variable, and of any other parent variable).

¹⁰¹ R. Daly, Q. Shen, S. Aitken, "Learning Bayesian networks: approaches and issues", *The Knowledge Engineering Review* 2011, vol. 26, no. 2, pp. 99-157, European Commission, *Support to SMEs...*, *op. cit.*

¹⁰² R. Kenett, S. Salini, *Modern analysis of customer surveys: with applications using R.*, vol. 117, John Wiley & Sons 2012.

The third advantage of using BNA lies in its ability to work as potential predictors for intervention. In fact, if the parent-child relationships in the network are a stable and autonomous mechanism of change. Organizing the knowledge about a given policy instrument in such modular configurations improves the possibility to reasonably predict the effect of external interventions. Thanks to this modularity, it is possible to perform simulations and assess scenarios, by simply changing or setting the value of particular variables and verifying how its respective parent or child variables change in result. For example, knowing that the level of exports is linked to the size distribution of the enterprise, one could simulate what would happen to the distribution of the outcome variable when the distribution of the size variable changes, for instance by increasing the number of medium enterprises and decreasing the number of small ones.

For all these reasons, BNA is an innovative tool to support strategic decisions¹⁰³ and to carry out *ex post* evaluations.

Application and results

The logic strategy of TITLE II policy instrument has been tested empirically. The goal was to verify whether the expected outcomes were achieved as predicted by the policy maker, and also what mechanisms or drivers of change within beneficiary enterprises contributed to such outcomes. The analysis was guided by the following research questions¹⁰⁴:

- Did the policy instrument succeed in maintaining the employment in beneficiary micro and small enterprises? What changes in the firms' basic activities or what other factors can explain the observed achievements?
- Did the policy instrument succeed in stimulating an increase in the enterprises' propensity to invest? What changes in firms' basic activities or what other factors can explain the observed achievements?
- Are the changes observed compliant with the intervention theory? If deviations from the theory are recorded, why have they occurred?
- Did the policy instrument produce any other types of behavioural change in supported micro and small enterprises? What factors can explain the observed changes?

In order to answer these questions, a survey was sent to 2,441 small and micro enterprises which had completed their investment projects and had already received the regional public contribution. A total of 399 questionnaires have been filled in by a sample of firms, which represents 25% of enterprises we were actually in contact with. In pre-

¹⁰³ J. Pearl, "Causal inference without counterfactuals: comment", *Journal of the American Statistical Association* 2000, pp. 428-431.

P. Spirtes, "An anytime algorithm for causal inference.", *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics* 2001.

C. Glymour, G. Cooper, *Causation, computation and discovery*, MIT/AAAI Press, Cambridge, MA 1999.

¹⁰⁴ According to tender's specifications.

senting the results of this survey, we separate the domain of economic performance, the domain of behavioural change in enterprises, and the domain of mechanisms underlying the outcomes, which have been detected through the BNA.

Economic performance

The conducted analysis showed that the policy instrument was generally effective in generating positive economic results, particularly in terms of increased sales, the capacity to resist the effects of the crisis, and the reduced risk of unemployment. In their responses concerning the economic results achieved thanks to the supported investment project, 80% of enterprises stated that the project helped them to better resist the effects of economic crisis at least to a limited or moderate extent. As far as employment level is concerned, 12% of enterprises believe that the most relevant changes observed after the investment in their production function include maintaining the existing employment level, while 22% of enterprises declared that their level of employment has increased. Moreover, it can be observed that a great proportion of enterprises (85%) have recorded an increase in sales, although this increase was generally moderate.

In order to understand the mechanisms explaining such economic results, one should consider the type of investment carried out, since TITLE II could finance a highly diversified spectrum of investments. The data obtained indicate that the majority of supported projects were aimed at general expansion of business activity, rather than at investing into product or process innovation. The survey attempted to go beyond this broad classification and to specify more precisely the type and aim of expenses incurred. The surveyed enterprises have used the TITLE II contribution primarily to purchase new assets for production purposes or for commercial activities. The support was used to a lesser extent to finance restructuring works or construction of new buildings and offices. 29% of enterprises purchased more added-value goods, like information systems, equipment or robots for automation, patents and license rights. Regarding changes in their internal activities, a total of 204 enterprises have improved their existing products or services thanks to the investment.

The policy instrument left the beneficiary relatively free to choose its own path for investment plans. Therefore, not surprisingly, the statistical analysis found that the type and level of achieved economic effects largely depend on the type of changes produced within a given SME as part of the investment project. More specifically, the data analysis identified the following correlations:

- Firstly, an increase in the enterprise resilience to the crisis is positively and significantly correlated with investments, which increase the popularity of the enterprise and the number of employees, and reduce energy consumption.
- Secondly, an increase in sales is positively correlated with expanding the range of products and services offered, as well as with productivity gains and the hiring of

new employees. However, when controlling for other variables, such as the sector in which a given enterprise operates, and the enterprise size and category (craft, commerce or other), a negative association has been found with extending or improving of workspace.

- Thirdly, entrepreneurs who have started a new activity or have improved the existing products or services and those who have extended the range or offered new products, have developed new sales channels and made their enterprise more popular could benefit from the increase in their income more than others.
- Fourthly, sharper decreases in total costs are achieved by those enterprises that either reduced energy consumption or improved work organisation, or improved their efficiency in some other ways.
- Fifthly, enterprises which either have increased or maintained their employment levels, have extended the range of offered products and increased their popularity. This result is also positively and significantly correlated with the implementation of other investment projects simultaneously.

The BNA confirms these findings and adds some findings on the mechanisms generating economic results. In particular, it highlights that the enterprise's perception that its capacity to resist economic crisis has improved depends on gaining a higher number of clients, which in turn is positively associated with an increase in sales and with simultaneous implementation of other investment projects, different from the one for which the TITLE II contribution was received. Moreover, it appears that an increase in sales translates directly into an increase in the income of entrepreneurs, and is strongly linked with an increase in the number and types of clients. Finally, the employment of new workers, and therefore the growth of the business size, is the most relevant determinant of sales; the increase in the types of clients, on the other hand, is caused by the increase in popularity. Hence, the BNA allows not only to empirically test the validity of the theory of change underlying the designed interventions, but it also clarifies and defines more precisely some of the proposed mechanisms.

In addition to the types of change produced by the investment project, the features of the beneficiary enterprise are crucial explanatory variables of the economic results. More specifically, while average economic results seem positive overall if the full sample of respondents is considered, a disaggregated analysis, which distinguishes among different types of enterprise, allows new and interesting findings to come to light. First, manufacturing non-craft enterprises implemented more costly investment on average, which has enabled them to attain substantial changes in the production process. These enterprises, while not representing the majority of beneficiaries, have achieved important benefits that result in their higher resilience to the crisis and stronger generation of employment. Consequently, it seems that this category of enterprises is actually going through a specific development path, as also suggested by their higher propensity to continue to invest. Second, commerce enterprises have more

often used the TITLE II to purchase furniture and new assets for commercial activity and a large share of beneficiaries believe that these changes have brought positive increase in turnover. However, as mentioned above, improving the working spaces is not as effective as widening the range of products and increasing productivity. Third, craft enterprises, which represent the majority of supported enterprises (almost all of which being single entrepreneurs or micro enterprises), have achieved mixed results, but overall lower than those recorded by other categories. Crafters operating in the construction sectors considerably suffered from the crisis and help provided by the TITLE II was limited. Those operating in the manufacturing sectors obtained more positive benefits, but not as high as those attained by small-size enterprises.

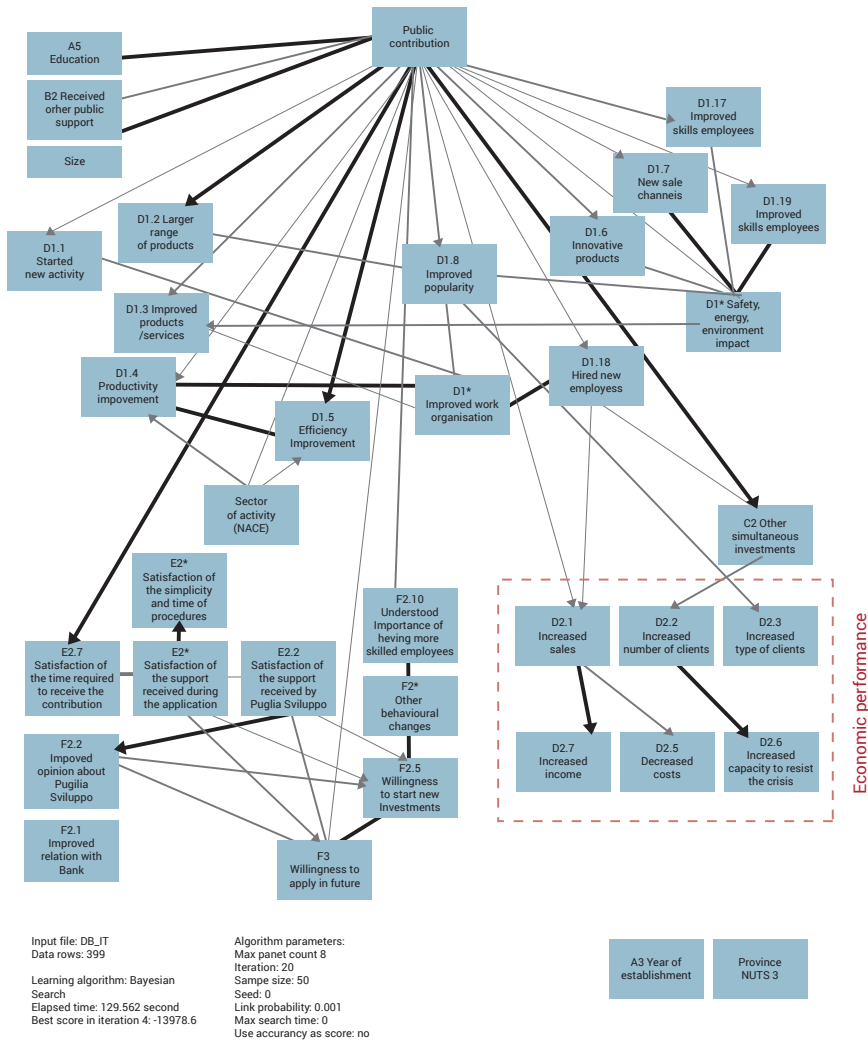
Behavioural change with respect to the propensity to invest

The second main goal of the policy instrument was to stimulate a behavioural change in the enterprises, consisting of an increased propensity to realise investment projects. 73% of the total number of enterprises stated that after benefitting from TITLE II, they started to consider the idea of implementing new investment projects never considered before. Responses to this question are strongly correlated with the enterprises' willingness to apply again for a public contribution in the future. It is important to highlight that out of the number of enterprises which had never benefitted from public support before receiving TITLE II, 70% have declared that they will probably apply again for public contribution in the future to support their potential investment plans. While this may open the door to opportunistic behaviour in the search of public support, it may also be a good opportunity for SMEs facing severe credit restrictions for genuine investment plans. In turn, the statistical analysis points out that intentions to make application in the future are stronger if the enterprises have positively assessed the support received by their consultants and the local intermediary body during the whole procedure of application to TITLE II. Positive opinions on these aspects offset the generally lower satisfaction with the administrative burden, difficulties of procedures and time required to submit the application or obtaining the public contribution after the project completion. In the Bayesian Network representation it is possible to see that the enterprises' inclination to think about other possible investment projects is strongly associated with their opinion on the overall delivery process of the policy instrument. It is also evident that behavioural change aspects are not strongly linked with the economic results achieved in the investment project, as would be expected.

Considering the features of beneficiary enterprises, available evidence shows that the pure additionality effect (the investment that would not have started without the aid), estimated on the basis of the enterprises responses, seems stronger for craft (micro) companies as compared with enterprises of the commerce sector. At the same time, craft and commerce enterprises are less likely to be thinking about new possible investment projects than enterprises of other types, and are less likely to apply again for other forms of public support in the future. While it is true that all three types of enter-

prise declare that they are planning to undertake other investment projects, non-artisans and non-commerce enterprises are relatively more oriented towards investment. Also, even in the same years when the investment project financed by TITLE II was carried out, half of them implemented other investment projects, a higher share than for the craft and commerce sectors. The differences are pronounced even when comparing micro-sized and small-sized enterprises, with the latter being characterised by a generally higher propensity to investment. The analysis also reveals other behavioural changes among certain beneficiary enterprises, for example in their preferences for particular types of employees (e.g. more skilled or younger). However, such changes concern only half of enterprises. Figure 1 presents the Bayesian Network related to TITLE II. It shows the links connecting various activities implemented by beneficiary enterprises, as well as changes caused by TITLE II on their production model, the characteristics of enterprises (e.g. size, sector, whether they have implemented other investment projects in the same years, etc.), their economic performance (variables from D2.1 to D2.7), and their behavioural change (from F2.1 to F2.10).

Figure 1. Changes triggered by TITLE II, according to the Bayesian Network Analysis



Note: Directed arrows represent a causal relation; simple links between variables indicate correlation, without any certain causal direction. The thicker the arrow, the stronger the correlation between the variables, as estimated by GeNIe. The graph includes some variables (bottom right) that, in spite of having been controlled for during the construction of the model, do not result in being strongly linked to any other particular variable. Asterisks indicate principal component variables.

Source: European Commission, "Support to SMEs – Increasing Research and Innovation in SMEs and SME Development", *Third Intermediate Report - Work Package 2*, submitted by CSIL in partnership with CSES and ZEW, Brussels 2015.

Mechanisms and conditions behind the generation of outcomes detected through the BNA

The policy instrument strategy does not describe clearly the expected mechanisms behind the generation of the desired effects, mainly due to a large variety of eligible expenses and types of investment supported. The investment could in fact determine different kinds of changes in the SMEs and in the economic effects produced, depending on whether it consisted, for instance, of the renovation of commercial premises and furniture, product and process innovation, the start of a new business activity or others. The policy makers present the logic of the instrument in a rather generic way, without describing the way in which different results are expected to take place.

The Bayesian Network Analysis confirms the existence of diversified effects and it helps to specify more explicitly the mechanisms and conditions behind the generation of outcomes, including those more implicit or, even, unexpected. Thanks to the use of BNA we are able to draw a number of considerations.

First, in spite of the variety of expenses made by the enterprises that were eligible for the TITLE II contribution, only a subset of them brought direct and positive economic effects. In particular, after controlling for different features of enterprises, we have found that higher effects on turnover are achieved by investments which enable the enterprise to widen the range of offered products, improve productivity and hire new employees. Investments which boost the enterprise's popularity increase its resilience to the crisis.

Second, the type and intensity of effects varied significantly according to the characteristics of beneficiaries. Small enterprises operating in the manufacturing sector carried out on average more costly investments aimed at acquiring new production assets. Thanks to the investment, such enterprises achieved productivity gains and improved the quality of products offered, which enabled them to enjoy large economic benefits. The commerce sector, particularly enterprises engaged in food service activities, has generally used TITLE II to renovate the furniture and spaces of the business premises. This type of expenses, while being associated with minor effects on turnover as compared with other types of investments, enabled the beneficiary enterprises to increase their sales. To some extent, the positive performance of this sector could be also explained by the positive trend of the tourism sector in the region. TITLE II could not bring any significant benefit to enterprises operating in the construction sector, which suffered more than other sectors from the bad macroeconomic trend and unemployment.

Third, whether enterprises achieve positive effects also depends on simultaneous implementation of other investment projects. These have been more often initiated by small manufacturing (non-craft) enterprises, which proved to be more dynamic than other types of enterprises in applying for public support and more willing to start other

investment projects in the future. Nevertheless, this class of enterprises represents a minor share of the instrument's beneficiaries.

The only causal association that the BNA disproves is the expected link between the enterprises' willingness to apply for other public support and their satisfaction with the simplicity of the application and delivery process of the instrument and with the economic results obtained. It is rather the support received by the local intermediate body and by private business consultants throughout the entire delivery process that made the application and fund disbursement easier, and consequently made the enterprises more willing to apply again.

Conclusions

The Apulian policy instrument TITLE II targeted a very high number of micro and small enterprises, pursuing broad objectives and financing many different kinds of actions. Although conventional evaluation techniques detected a limited improvement in terms of pure economic performance, the evaluation study demonstrated that the instrument played an important role in supporting enterprises to cope with the severe economic crisis. The evidence indicates that TITLE II managed to accomplish its expected outcomes at least in general terms, but when considering the types of projects implemented and enterprises supported, there can be many different interpretations of conditions explaining individual success stories. Given the limited selectivity of this policy instrument in terms of investment supported and changes triggered, it is even more challenging to assess it.

BNA was used to examine changes caused by the policy instrument, which are usually more difficult to identify. In fact, changes occur over time in a dynamic process and thus it is hard to detect and quantify them. BNA revealed some interesting behavioural changes at the enterprises level, in terms of propensity to invest, hiring strategies and capacity to attract new customers. This innovative technique allowed us to identify mechanisms and conditions behind the generation of results. In general, the study conducted on the Apulian TITLE II instrument shows that when the initial theory is poorly specified and the policy instrument does not point to a precise mechanism to achieve the desired effects, the path along which this policy instrument can achieve its goals is not straightforward. Due to the broadness of objectives, we can say that this instrument is not particularly effective in triggering long-term structural changes, but it has been appropriate for supporting the survival of many micro and small enterprises in a critical period from an economic and social point of view. However, thinking about the logic behind the portfolio of instruments, it can be argued that other policy instruments may have a more important role in bringing about more profound behavioural changes. An instrument like TITLE II may create room for enterprises that are more likely to be reactive to other policies with more ambitious focus on structural changes. Thanks to the mechanisms detected through the Bayesian Network Anal-

ysis, it may be assumed that if more restrictions on the sectors addressed and the type of investment to be supported were introduced, so as to narrow the scope of the policy instrument to enterprises and activities with the highest potential to grow and contribute to regional economic competitiveness, the effectiveness of TITLE II would have been higher.

About the authors

Emanuela Sirtori

Emanuela Sirtori earned a degree in Economics from the University of Milan and is currently a PhD fellow in Economics of Technology, Innovation and Development at the UNU-MERIT School of Governance (Maastricht). In 2009 she attended a Summer School on Quantitative methods for evaluating the impact of the EU funds through counterfactual impact evaluation methods. She is currently a partner and a project manager at CSIL in the field of regional development, industrial policy and innovation, small business economics. Over the years, she has been involved in different evaluation studies on behalf of the European Parliament and the European Commission. She had a leading role in the ex post evaluation of Cohesion Policy programmes supporting SME innovation and development during the 2007-2013 period. In the frame of this evaluation study, she tested the Bayesian Network Analysis as an innovative tool to process firm level survey data.

Elena Vallino

Elena Vallino holds a degree in Political Sciences at the University of Torino, a master's degree in International and Development Economics at the University of Applied Sciences in Berlin, Germany. She completed a PhD Program in Economics of Complex Systems at the University of Torino, spending some research periods in Belgium and in Arizona (USA). She is currently a Senior Researcher at CSIL specializing in the field of evaluation of public policies and the EU Cohesion Policy. In the past, she worked for Italian NGOs and for the German Technical Cooperation Agency in Indonesia. She has had several teaching assignments and research fellowships and she has published scientific articles in the fields of economics of developing countries, environmental economics, transport economics, and agent-based modelling.

Silvia Vignetti

Silvia Vignetti holds a degree in Economics from the Bocconi University. In the past she had teaching appointments at the University of Milan in Macroeconomics, Economics of European Integration and Cost-benefit Analysis. She is currently a partner of CSIL and the Director of the Evaluation Unit. She is an expert in regional development, the EU Cohesion Policy and evaluation of projects and programmes. She has gained relevant experience in this field and has been involved in several work packages of the ex post evaluation of the EU Cohesion Policy programmes of the periods 2007-2013 and 2000-2006 in the fields of business support, infrastructure development, culture and tourism, demographic change. She has been involved in several studies on behalf of the European Parliament on the effectiveness and delivery mechanisms of the Cohesion Policy and is currently involved in a research study feeding into the debate on the post 2020 reform of the EU Cohesion Policy.

Bibliography:

1. Astbury B., Leeuw F. L., "Unpacking black boxes: mechanisms and theory building in evaluation", *American Journal of Evaluation* 2010, vol. 31, no. 3.
2. Blackman L., Reich S., "Randomized control trials: a gold standard with feet of clay?", in: S. Donaldson, C. Christie, M. Mark (eds.), *What counts as credible evidence in applied research and evaluation practice?*, SAGE, Thousand Oaks, CA 2009.
3. Carvalho S., White H., "Theory-based evaluation: the case of social funds", *American journal of evaluation* 2004, vol. 25, no. 2.
4. Daly R., Shen Q., Aitken S., "Learning Bayesian networks: approaches and issues", *The Knowledge Engineering Review* 2011, vol. 26, no. 02.
5. European Commission, "Support to SMEs – Increasing Research and Innovation in SMEs and SME Development", *Third Intermediate Report - Work Package 2*, submitted by CSIL in partnership with CSES and ZEW, Brussels 2015.
6. Glymour C., Cooper G., *Causation, computation and discovery*, MIT/AAAI Press, Cambridge, MA 1999.
7. Heckerman D., Geiger D., Chickering D. M., "Learning Bayesian networks: the combination of knowledge and statistical data". *Machine Learning* 1995, vol. 20.
8. Horný M., "Bayesian Networks", *Technical Report No.5*, Boston University 2014 <http://www.bu.edu/sph/files/2014/05/bayesian-networks-final.pdf> (access on 25.05.2017).
9. Kenett R., Salini S., *Modern analysis of customer surveys: with applications using R*, vol. 117, John Wiley & Sons 2011.
10. Massa S., Testa S., "Innovation and SMEs: Misaligned perspectives and goals among entrepreneurs, academics, and policy makers", *Technovation* 2008, vol. 28, no. 7.
11. Mouqué, D., "What are counterfactual impact evaluations teaching us about enterprise and innovation support?" *Regional Focus* 2012, vol. 2.
12. Murphy K., *A Brief Introduction to Graphical Models and Bayesian Networks*, 1998.
13. Pearl J., "Causal inference without counterfactuals: comment", *Journal of the American Statistical Association* 2000.
14. Spirtes P., "An anytime algorithm for causal inference", *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics* 2001.
15. Weiss C. H., "How can theory-based evaluation make greater headway?", *Evaluation Review* 1997, vol. 21, no. 4.

The use of contribution analysis as a theory-based evaluation approach in the ex post evaluation of support to large enterprises under Cohesion Policy programmes 2007-2013

Introduction

Theory-based evaluation (TBE) is gaining significant popularity in the field of evaluation of European structural and investment funds (ESIF). Although it is not a new approach – it appeared in the evaluation literature in the early 70’s¹⁰⁵ and is widely used in development evaluation for instance – it is still a challenging task to apply it in a practice in the context of evaluating European funding instruments.

This chapter aims to present an example of the use of TBE, incorporated in the ex post evaluation of Cohesion Policy programmes 2007–2013, within Work Package 4 (WP4) – Support to large enterprises.¹⁰⁶ This ex post evaluation studied eight EU Member States, which represented 75% of total spending on enterprise support under Cohesion Policy. Seven of these accounted for the greatest absolute spending on large enterprise support of all EU Member States (Poland, Germany, Portugal, Spain, Hungary, Italy and the Czech Republic), while Austria had the highest proportional spending in this support area.

The ex post evaluation followed the concept of theory-based evaluation, emphasising the reconstruction and testing of the identified Theories of Change. More specifically, it applied Contribution Analysis¹⁰⁷, a variant of theory-based evaluation, the design of which is particularly useful for evaluating complex causal questions.

¹⁰⁵ F. Leeuw, *Theory-Based Evaluation - Guidance produced for DG Regional Policy*, European Commission 2012.

¹⁰⁶ KPMG/Prognos, *Support to large enterprises. Final Report – Work Package 4, Ex post evaluation of Cohesion Policy programmes 2007-2013*, Brussels 2016

¹⁰⁷ J. Mayne, “Contribution analysis: addressing cause and effect”, in: R. Schwartz, K. Forss and M. Marra (eds.) *Evaluating the Complex*, Transaction Publishers, New Brunswick, NJ 2011, pp. 53–96.

Prior to this evaluation, relevant counterfactual impact evaluations on large firm support had often found the support to have no impact or an insignificant impact.¹⁰⁸ These results had important implications for the effectiveness of the support; however, they could not open the 'black box' explaining large firms' behavioural change. Thus, the goal of the ex post WP4 evaluation was to assess the rationale for, implementation of and evidence of effectiveness of Cohesion Policy. Policy rationales and literature were examined to develop typical Theories of Change for why and how such support works. These narratives were critically assessed using evidence from case studies. Such evidence was confronted with counterfactual impact evaluations (where available).¹⁰⁹

The remainder of the chapter is organised as follows: section 2 and 3 present the rationale behind the TBE approach. Section 4 describes the process of the reconstruction of Theories of Change. Section 5 presents the process of testing the Theories of Change (with a Polish example). Section 6 provides selected policy conclusions of the WP4 evaluation that help demonstrate the policy-relevant outcomes of the testing of the Theories of Change. The last section summarises our conclusions on the use of TBE and Contribution Analysis in evaluation practice.

The “black box” trap in evaluation

Counterfactual impact evaluation (CIE) allows to answer the question whether 'a given intervention works'.¹¹⁰ Put differently, it allows to assess whether an idea for solving a given socio-economic problem has brought the expected results or not. Although such knowledge is often sufficient to decide whether or not an intervention should be continued in its current form, in most cases it is not enough to learn 'why the intervention works', 'how it works' and finally 'how to make it better'. This is the problem of the so called 'black box': based on CIE results it is possible to track the inputs and outputs of a programme, but the relationship between the two remains unknown. To find answers to these questions, evaluations, apart from defining the intervention impact, must focus on cause-and-effect mechanisms that make the expected effects materialise. This is important because in most cases interventions (a programme, a project, etc.) are neither totally rejected nor continued unchanged after an evaluation. The socio-economic challenges that public interventions try to address (unemployment, poverty, social exclusion, low innovation and competitiveness of enterprises,

¹⁰⁸ D. Mouqué, *What are counterfactual impact evaluations teaching us about enterprise and innovation support?*, DG for Regional and Urban Policy, Brussels 2012 http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_02_counterfactual.pdf

¹⁰⁹ Ex post evaluation of Cohesion Policy programmes 2007-2013, financed by the European Regional Development Fund (ERDF) and Cohesion Fund (CF) Work Package 4: Support to large enterprises - Tender Specifications - European Commission, Directorate-General Regional And Urban Policy, Policy Evaluation and European Semester. 2014CE16BAT033.

¹¹⁰ A. Martini, Wprowadzenie do podstawowych pytań stawianych w ewaluacji oddziaływania interwencji: Czy interwencja coś zmieniła? A jeśli tak, to dla kogo?, in: A. Haber., R. Trzciński (eds.), *Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych*, Polish Agency for Enterprise Development, Warsaw 2011, pp. 23-31

etc.) are usually of lasting nature, therefore, subsequent programmes usually represent an evolution rather than revolution of the previously implemented interventions. Evaluation is an important source of evidence when it comes to the design or redesign of an intervention, therefore, it needs to provide answers to all of the complex questions raised above.

To fully use the CIE outcomes and understand what is behind the success or failure of public interventions, it is necessary to conduct simultaneously an analysis of processes and mechanisms making the defined changes occur. An important advocate of such an approach is e.g. Howard White¹¹¹, who promoted the use of theory-based impact evaluations. Most recently, the approach suggested by White has also been recommended by the European Commission with reference to evaluations of operational programmes implemented under the 2014-2020 perspective.¹¹²

Theory-based evaluation

Over the last 40 years, the concepts for conducting research using a theory-based approach have been developed and several specific methodological approaches have been elicited, such as theory-driven evaluation¹¹³, theory-based evaluation¹¹⁴ and realistic evaluation¹¹⁵. These concepts have a major aspect in common: they all treat public interventions as a 'theory' of how a defined socio-economic problem should be solved¹¹⁶.

While designing public policies and programmes, it is a natural assumption that a given intervention is necessary (in certain cases, sufficient) to address the socio-economic challenge. Many of these policies are targeted at individuals (e.g. through training) or entities (e.g. support to SMEs, NGOs etc.). It is assumed that changes at this micro level (i.e. programme beneficiaries) will, at some point, reach the 'critical mass' and will bring measurable and visible effects also at the macro level. These macro changes are usually expressed as specific objectives or intended results of the intervention.

¹¹¹ H. White, *Theory-Based Impact Evaluation: Principles And Practice*, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi 2009

¹¹² European Commission, *Guidance document on monitoring and evaluation. European Cohesion Fund European Regional Development Fund, Concepts and Recommendations*, Belgium 2014

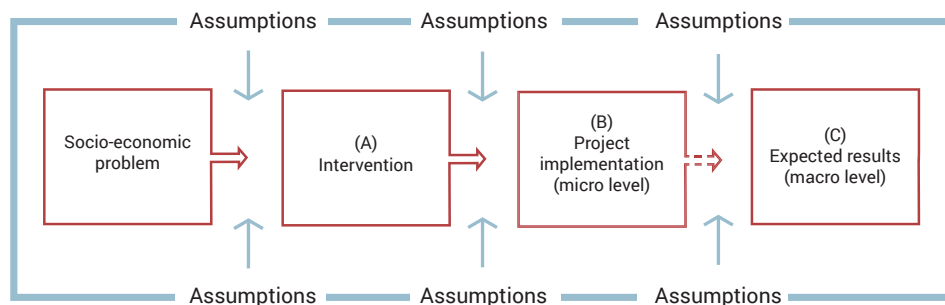
¹¹³ H. T. Chen, P. H. Rossi, The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science, *Social Forces* 1980, 59, pp. 106-122

¹¹⁴ C. H. Weiss, "Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families", in: J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (eds.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts*, New York, NY: Aspen Institute 1995, pp. 65-92

¹¹⁵ R. Pawson, N. Tilley, *Realistic Evaluation*, London 1997

¹¹⁶ J. Górniak, "Ewaluacja w cyklu polityk publicznych", in: S. Mazur (ed.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*. Cracow: Cracow University of Economics, Malopolska School of Public Administration 2007

Scheme 1. Overall logic of programme implementation



Taking the example of a programme supporting R&D activities in large enterprises, a theory-based approach interprets the intervention as a certain impulse (action A) aimed at encouraging companies to be more active in the field of R&D. Projects implemented by the companies (R&D investments) should increase the level of their innovation capacity, but should also result in innovative products or services which, after being introduced to the market, should have a positive impact on their competitiveness, their overall financial performance and further development in economic activity (action B). At the same time, providing such public aid is expected to not only improve the condition of the companies themselves, but also improve the regional or national economy, induce spill-over effects on local SMEs or improve international trade (action C).

The simplified general logic presented above can be further specified. On the one hand, the achievement of these results depends on many other (often correlated) external factors and involves further assumptions, without which the intervention would not work out in practice. Most of these assumptions are not expressed explicitly in the intervention, which makes the evaluation of their role in achieving the intended results difficult.

On the other hand, public interventions are also much more complex, as they include many intermediate “micro steps”, which are in causal relationships with each other. Starting with the public funding, the chain of these micro steps leads to achievement of the intended results. The interventions often also result in indirect effects, intended or unintended, which further complicate causal chains.

The goal of TBE is to reconstruct the detailed causal ‘Theory of Change’ of an intervention and to test it against the available evidence. Leeuw’s paper on ‘Theory-Based Evaluation’¹¹⁷ borrows Carol Weiss’ definition of Theory of Change, described as:

¹¹⁷ DG REGIO, *EVALSED Sourcebook: Method and Techniques, Regional and Urban policies*, Brussels 2013

A way to describe the set of assumptions that explain both the mini-steps that lead to the long-term goal and the connections between policy or programme activities and outcomes that occur at each step of the way.

Among the many variants of theory-based evaluation, in the WP4 evaluation, 'Contribution Analysis' was chosen to evaluate large firm support, because of its ability to de-compose complex causal chains into micro steps, internalise assumptions and external factors in a single framework ('Theory of Change') and test them against various sources of information in a rigorous way. In technical terms, Contribution Analysis can:

- clarify which 'causal packages'¹¹⁸ the programmes were a part of,
- identify the interactions between the components of causal packages,
- establish how necessary and/or sufficient the programmes were and
- judge the likely contribution of the programmes.

Contribution Analysis, as a specific variant of theory-based evaluation, focuses on 'drawing causal links and explanatory conclusions between observed changes and specific interventions'¹¹⁹. This approach relies on assumptions that should be made visible as both requirements and limitations to the evaluation. As Leeuw¹²⁰ writes, 'a reasonable contribution claim can be made if:

- There is a reasoned Theory of Change for the intervention: the key assumptions behind why the intervention is expected to work make sense, are plausible, may be supported by evidence and are agreed by at least some of the key players.
- The activities of the intervention were implemented as set out in the Theory of Change.
- The Theory of Change – or key elements thereof – is supported and confirmed by evidence on observed results and underlying assumptions, both of experts and of facts: the chain of expected results has occurred and the Theory of Change has not been disproven.
- Other influencing factors have been assessed and either shown not to have made a significant contribution, or their relative role in contributing to the desired result has been recognised.¹²¹

¹¹⁸ Many interventions do not act alone and the desired outcomes will be the result of a combination of causal factors, including other related interventions, and events and conditions external to the intervention. Indeed, many interventions are designed to be part of such a 'causal package', and even if not so designed, their evaluation needs to take these other factors into account (J. Mayne, Contribution analysis: Coming of age?, *Evaluation* 2012 18: 270).

¹¹⁹ *Ibid.*, pp. 55

¹²⁰ F. L. Leeuw, Linking theory-based evaluation and contribution analysis: Three problems and a few solutions, *Evaluation*, 18(3), 2012, pp. 348-363

¹²¹ J. Mayne, Contribution analysis: addressing cause and effect, in: R. Schwartz, K. Forss and M. Marra (eds.) *Evaluating the Complex*, Transaction Publishers, New Brunswick, NJ 2011, pp. 53–96.

Methodological steps as suggested by Mayne do not necessarily follow a strict step-by-step logic, but constitute a highly iterative process with multiple rounds of revisions of previous exercises.

These are general steps, which were followed during the WP4 evaluation:

1. Set out the cause-effect issue to be addressed
Acknowledge the attribution problem, determine the specific cause-effect question being addressed, determine the level of confidence required, explore the type of contribution expected, determine the other key influencing factors, and assess the plausibility of the expected contribution in relation to the size of the programme.
2. Develop a Theory of Change
Build a Theory of Change and a results chain, determine the level of detail, determine the expected contribution of the programme, list the assumptions underlying the Theory of Change, include consideration of other factors that may influence outcomes, determine how much the Theory of Change is contested, assess the contribution story.
3. Gather existing evidence on the Theory of Change
Gather evidence (e.g. Counterfactual Impact Evaluations, other impact evaluations, studies).
4. Assess the resulting contribution story
Assess the logic of the links in the Theory of Change, the credibility of the overall story, gaps in the theory, and stakeholders' agreement on the story.
5. Seek out additional empirical evidence
Identify what new data is needed, gather evidence (e.g. through case studies, stakeholder interviews).
6. Revise and strengthen the contribution story
Revise the contribution story, go back to Step 4 if necessary.

Reconstruction of programme theory

A good programme theory must fulfil certain criteria: it must be plausible, achievable and testable. The process of reconstructing the programme theory is mainly of conceptual nature. This process is usually iterative, the Theory of Change being reconstructed and refined during the whole duration of an evaluation. The reconstruction of programme theory needs to include both the:

1. description of the cause-effect chain of an intervention, including indirect effects;
2. identification of key assumptions (and external factors) of an intervention.

Creation of the cause-effect chain of an intervention

The programme theory should identify the socio-economic problem to be solved (the intended change) and the chain of steps through which the intervention is supposed to address the problem. The theory should also identify the indirect effects or wider benefits that are induced by the intervention. At each stage, the cause-effect chain should identify all key factors and causes of the expected changes, paying great attention to how the change takes place at the supported individuals or entities. According to Carol Weiss, a key matter is to distinguish all key mini-steps that create a characteristic cause-effect chain where some actions are the consequences of other actions resulting from a given intervention.¹²²

The reconstruction of theories of changes in the WP4 evaluation was an intensive work phase with multiple rounds of refinement. Each information source added different layers to the reconstructed theories. Roles of the key information sources were typically the following:

- Programme document (OPs, calls for applications) contained information on general, high-level objectives, although often they were lacking specific information on large enterprises (rationale, causal chain, assumptions, etc.). They were key sources for identifying inputs, activities and first-level outcomes of the activities.
- Programme planners and other Managing Authority representatives in Member States were found to be key to understanding the question: “why have large enterprises been supported from the European Regional Development Fund (ERDF)?”. Generally, they could recall crucial information (often implicit) on the “intended change” that was driving the planning and modification process of programmes.
- Literature review was helpful in reassessing those intended changes. Its most important role was, however, to make assumptions, external factors, indirect effects and wider benefits explicit, which could then be incorporated into Theories of Change, where applicable.

Altogether, 27 programme-level Theories of Change have been reconstructed in the evaluation process. As a result of a multi-round classification process, four generalised theories of large firm support have been reconstructed in the eight countries concerned. The theories include the most frequent elements of the country-level theories, and also build on the outcomes of the literature review, especially those related to key assumptions, influencing factors, indirect and wider effects. Below, these four generalised theories of change are presented.

¹²² C. H. Weiss, “Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families”, in: J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (eds.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts*, New York, NY: Aspen Institute 1995, pp. 65-92

Table 1. Overview of the four generalised Theories of Change

Theory of Change	Description
'LE1: Large-scale business investment'	Financial support to investments of large enterprises (including foreign-based ones) with the primary aim of increasing employment in the programme area in the long term.
'LE2: Technological upgrading'	Financial support to large enterprises to assist them to implement upgrades in technology with the aim of strengthening their competitiveness and thereby the growth potential of the regional economy.
'LE3: Innovation support'	Financial support to large enterprises to assist them to implement innovative investment projects for new products or processes, with the aim of strengthening the potential for regional innovation and the long-term growth of GDP and employment (high-quality jobs).
'LE4: Investment in R&D capacity'	Financial support to large enterprises to help them set up, expand or improve R&D facilities and/or carry out R&D activities, with the aim of expanding the regional knowledge base and the long-term R&D and innovation capacity of the region (including the creation of research jobs).

Source: Support to large enterprises. Final Report – Work Package 4, Ex post evaluation of Cohesion Policy programmes 2007-2013, Consortium of KPMG and Prognos, Brussels 2016

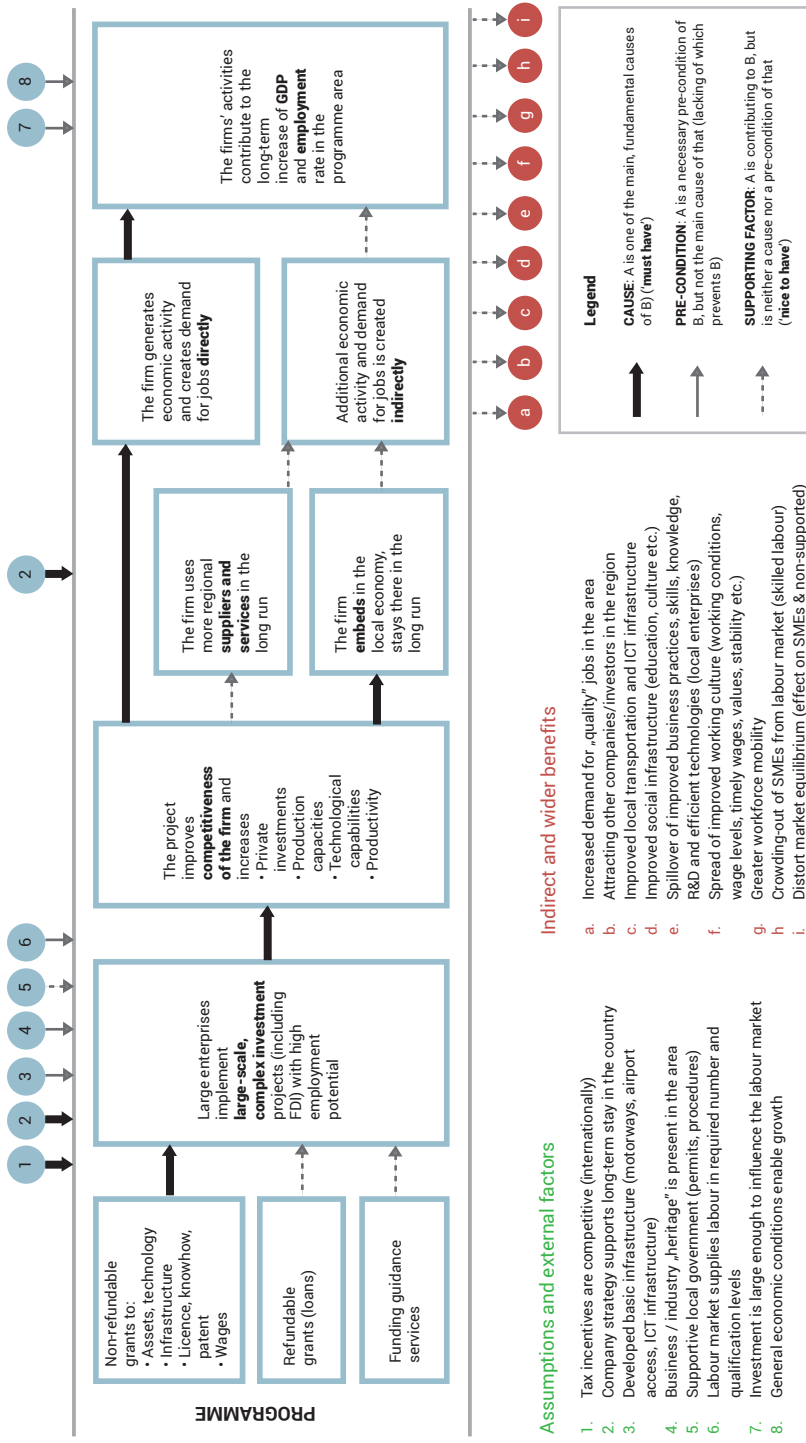
Each of 27 detailed, programme-level Theories of Change and 4 generalised ones were described with the use of the flow-chart diagram approach. Below we present an example of such a diagram for one of the reconstructed theories – 'LE1: Large-scale business investment'. This type of generalised Theory of Change was identified in five countries – Hungary, Italy, Poland, Germany and Spain. In the first three countries, the theory refers to country level programmes. In the case of Germany and Spain, regional level interventions constituted the framework for the analysis. In the case of Poland, this theory refers to investments made under Priority Axis 4 (e.g. Measure 4.5) of the Innovative Economy Operational Programme 2007-2013. The diagram covers the following key elements of the Theory of Change:

1. Intended change: the last box at the end of the results chain.
2. Intermediary steps to attain the change: description of inputs and intermediary steps in-between.

3. Causality of steps: represented by 3 types of arrows:
 - a. cause: *A* is one of the main, fundamental causes of *B* ('must have');
 - b. pre-condition: *A* is a necessary pre-condition of *B*, but not the main cause of that (lack of which prevents *B*);
 - c. supporting factor: *A* is contributing to *B*, but is neither a cause nor a pre-condition of that ('nice to have').
4. Key assumptions and external factors: represented by green circles explained below and channelled in to the logical chain of the chart from outside the programme "border".
5. Indirect and wider effects: changes that are supported by the intended change, but are not directly attributable to the intervention (represented by red circles in the chart).

Taking into account typical programme provisions oriented at achieving this goal, it is possible to state that the support to the enterprises in the form of subsidies and financial instruments is to cause development investments in these companies, followed by a visible employment growth. The large scale investments are made by purchasing fixed assets, including machines and equipment for production. Such activities are meant to increase the overall production level and capacities, technological capability and productivity. An improvement of the enterprise competitiveness should translate in turn into an improvement of its financial results, which will allow it to constantly develop and increase employment. Programmes supporting the 'LE1' Theory of Change were expected to have an impact on the country-level employment indicators.

Figure 1. Theory of Change: 'LEI: Large-scale business investment'



Source: KPMG/Prognos, Support to large enterprises. Final Report – Work Package 4, Ex post evaluation of Cohesion Policy programmes 2007-2013, Brussels 2016

Identification of key assumptions of an intervention

The model shown in Figure 1 represents an ideal situation where the micro steps of the Theory of Change and the expected results occur consecutively and form direct causal links. The theory presented above will, however, turn into practice only under particular conditions, when additional assumptions are met. Therefore, the Theory of Change must include the assumptions whose fulfilment is crucial to let the theory materialise. In the example presented above, these assumptions are indicated in the diagram with a green colour. As an example, the following key assumptions have been distinguished in the process of reconstruction of 'LE1: Large-scale business investment' Theory of Change in the WP4 evaluation:

1. Tax incentives are competitive (internationally).
2. Company strategy supports long-term stay in the country.
3. Developed basic infrastructure (motorways, airport access, ICT infrastructure).
4. Business/ industry "heritage" is present in the area.
5. Supportive local government (permits, procedures).
6. Labour market supplies labour in required number and qualification levels.
7. Investment is large enough to influence the labour market.
8. General economic conditions enable growth.

As shown in Figure 1, the assumptions can play different roles in the causal chain. Some of them were found to be important pre-conditions, while others just supporting factors. For example, the eighth assumption (*General economic conditions enable growth*) is referring to the external economic environment, which can have a decisive influence on the materialisation of the Theory of Change and thus should be treated as an important pre-condition. Should this assumption not be met, for instance in the case of an economic downturn, even the best investments are not likely to result in the expected outcomes. These assumptions and external factors are often mutually interrelated, therefore, the nature of these interrelations must be verified in the course of the evaluation.

Verification of the Theories of Change

The conceptual research phase described above is followed by an empirical phase that involves testing the reconstructed Theories of Change. According to Górnjak, evaluation is to check whether, in the light of the impact obtained, the programme theory can be maintained or should be rejected.¹²³ Methodological steps for analysing Theories of Change using a contribution analysis approach, as suggested by Mayne¹²⁴, do not necessarily follow a strict logic, but rather describe a highly iterative process with

¹²³ J. Górnjak, "Ewaluacja w cyklu polityk publicznych", in: S. Mazur (ed.), *Ewaluacja funduszy strukturalnych - perspektywa regionalna*. Cracow: Cracow University of Economics, Malopolska School of Public Administration 2007

¹²⁴ J. Mayne, "Contribution analysis: addressing cause and effect", in: R. Schwartz, K. Forss and M. Marra (eds.) *Evaluating the Complex*, Transaction Publishers, New Brunswick, NJ 2011, pp. 53 -96.

multiple rounds of revisions of previous exercises. These steps have been the basis for both reconstructing and testing the Theories of Change.

The verification of the Theories of Change in the course of TBE needs to be enriched basically with two kinds of evidence.

1. The first group of evidence could be referred to as factual¹²⁵. It informs 'how the intervention is/was going' in terms of the type and functioning of the project selection system, an actual level of innovation capacity of the implemented projects or the scope of the investments made, etc. To a great extent, factual evidence is to provide data that will allow to verify whether the adopted assumptions of an intervention have been actually fulfilled. Such data was partially provided by monitoring systems, previously conducted studies and additional research, based on the mixed method approach.
2. The second type of evidence (sometimes described as counterfactual) refers to the problem of capturing the impact of intervention. Since rigorous quantifiable evidence on the effectiveness of large enterprise support was not available in the majority of the cases, the identified Theories of Change were tested on multiple qualitative dimensions, and by confronting these information sources (data triangulation). The testing was carried out in a standardised framework with main sections on the (A) Background of the enterprise (context), (B) Inputs and activities, (C) Direct effects (and their sustainability), (D) Indirect effects and wider benefits (and their sustainability) and (E) Summary and conclusions (including alternative explanations). Evidence was collected for each significant micro step, outcome, key assumption and influencing factor in the Theory of Change, followed by assessment of whether these steps (e.g. effects) materialised, whether they were in causal relationship with the previous step in the chain and whether the evidence was reliable.

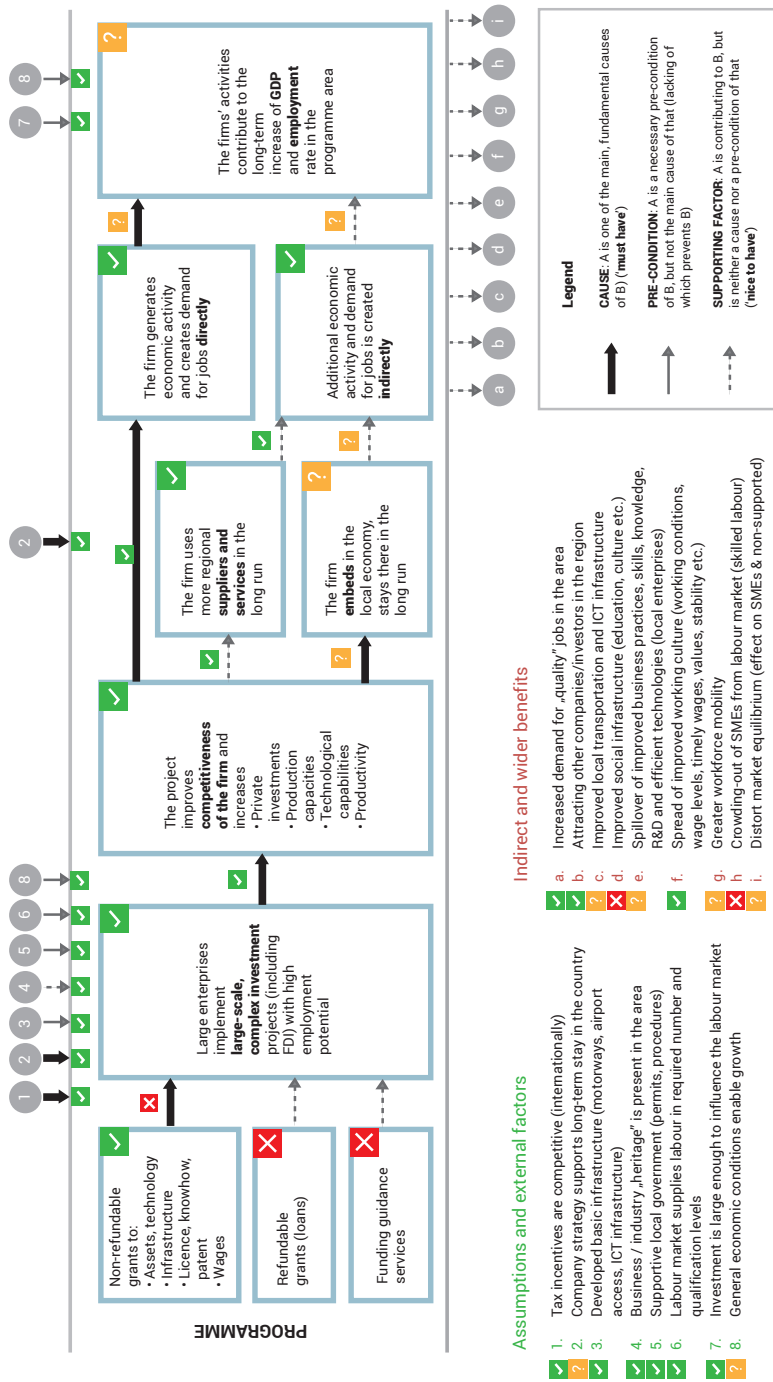
The Contribution Analysis was implemented through a comparative case study design. This involved eight case studies at the level of individual Operational Programmes in the eight Member States. At the heart of this structure were 45 company case studies that followed a multi-respondent design, with over 130 interviews. The company case studies utilised a systematic framework to assess the contribution of large enterprise support to project implementation and observed outcomes. Moreover, the monitoring data analysis and the outcomes of a comprehensive literature review were incorporated into the analysis, covering more than 100 scientific articles and evaluations on enterprise support. This empirical basis was used at different stages of the triangulation. While first interviews with managing authorities and intermediate bodies helped

¹²⁵ H. White, *Theory-Based Impact Evaluation: Principles And Practice*, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi 2009

uncover funding rationales and strategies for large enterprise support, interviews with corporate officials enabled the evaluation to explore the causal relationships and the resulting changes in the large firms' behaviour, as well as the extent of their contribution to the socio-economic development of the region. To reflect on these findings, additional interviews were performed with academic experts to counterbalance the potential bias in the data sets. This combined analytical approach enabled us to formulate our conclusions on a solid foundation of evidence.

The process of testing the theories of change was documented with the use of the flowchart diagrams. The figure below presents 'LE1: Large-scale business investment' Theory of Change – in the case of the Polish OP Innovative Economy (Measure 4.5).

Figure 2. Assessment of Theory of Change: 'LE1: Large-scale business investment', the case of Poland



Source: KPMG/Prognos, Support to large enterprises. 2nd Interim Report – Work Package 4, Ex post evaluation of Cohesion Policy programmes 2007-2013, Brussels 2016

*Question mark refers to the situation in which there was not enough evidence to either reject or confirm the existence of a causal link or impact

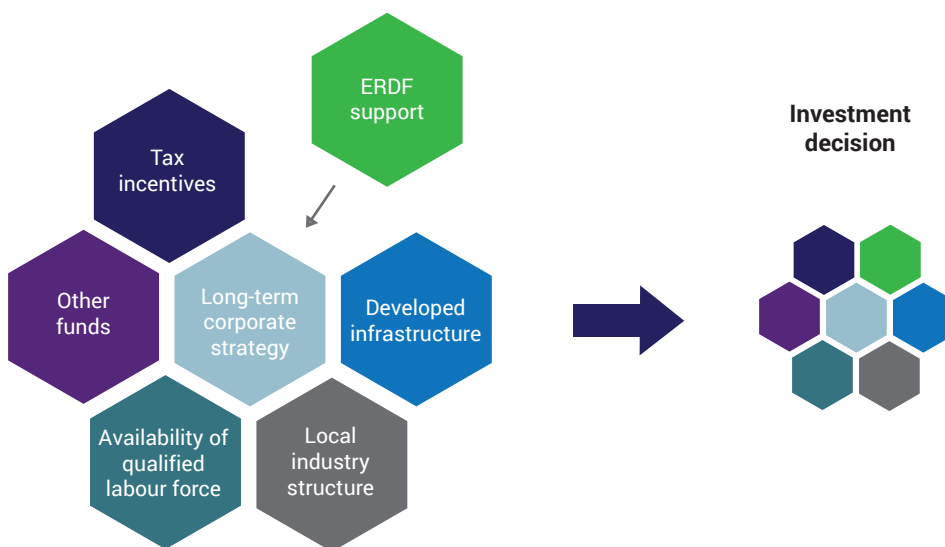
Policy conclusions

The use of theory-based evaluation in the WP4 project, combined with the available counterfactual impact evaluations, allowed to make the following conclusions regarding the identified theories of change in the field of large enterprise support from ERDF:

1. An array of factors determined the investment decisions of large enterprises in the context of Cohesion Policy support.

Analysis of the causal package and the roles of other influencing factors revealed that in the case of large enterprises, ERDF support typically acted as a pre-condition for these investments, i.e. as a necessary but not a sufficient condition for project implementation. As shown in Figure 3, the EU support was often only one of many influencing factors of investment decisions, the most important of which were the long-term corporate strategy, the availability of transport infrastructure and the local industry structure.

Figure 3. Array of factors that determined investment decisions of large enterprises



Source: KPMG/Prognos, Based on empirical evidence from the 45 company case studies, 2016

2. In 20% of cases, the desired behavioural change took place and the ERDF support was among the main causes of project implementation.

The induced behavioural change in these cases constituted the changes in the scope and timing of the implementation of strategic projects that would not have been implemented otherwise. This pattern was observed most frequently among large enterprises facing financial problems in the aftermath of the financial crisis.

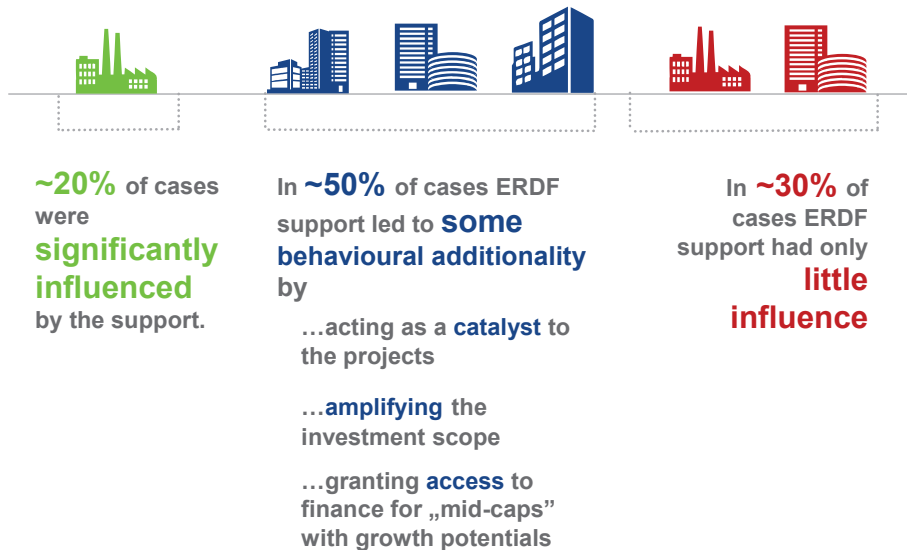
3. In 50% of cases, ERDF support was successful in inducing changes in corporate behaviour, particularly influencing the timing and scope of the implemented projects.

In 50% of cases, ERDF support was a necessary condition of project implementation and at least some degree of behavioural additionality was induced. Typically, the support often acted as a catalyst to project implementation, leading large firms to invest earlier than planned. Support also affected the project scope, notably by influencing firms to add new project activities (e.g. co-operation with other economic actors) or undertakings (e.g. hiring more people, using environmentally friendly technologies).

4. In 30% of cases, ERDF support had little influence on the behaviour of large enterprises.

The evaluation found that in the remaining 30% of cases, the EU funding played only a supporting role in project implementation and was not a necessary part of the causal package that led to the investment. In such cases, the influence of ERDF support on large firms' behaviour was small, often because projects were grounded in longer-term corporate plans. Particularly low behavioural additionality was achieved for projects focusing on basic technological upgrades, but it was much higher for large-scale business investments, innovation support and R&D support.

Figure 4. Influence of ERDF support on large enterprise behaviour



Source: KPMG/Prognos, Based on empirical evidence from the 45 company case studies, 2016

5. Ninety per cent of supported projects had positive direct effects, but – as described above – not all observed outcomes were fully explained by the EU support.

For the majority of projects, both the production capacity and the productivity of large enterprises have risen and the creation of jobs was widely achieved at the project level. For most firms, there were strong causal linkages between the project and the observed outcomes. Nevertheless, due to only moderate ability of ERDF support to influence large firms' behaviour, the impact of support was less straightforward.

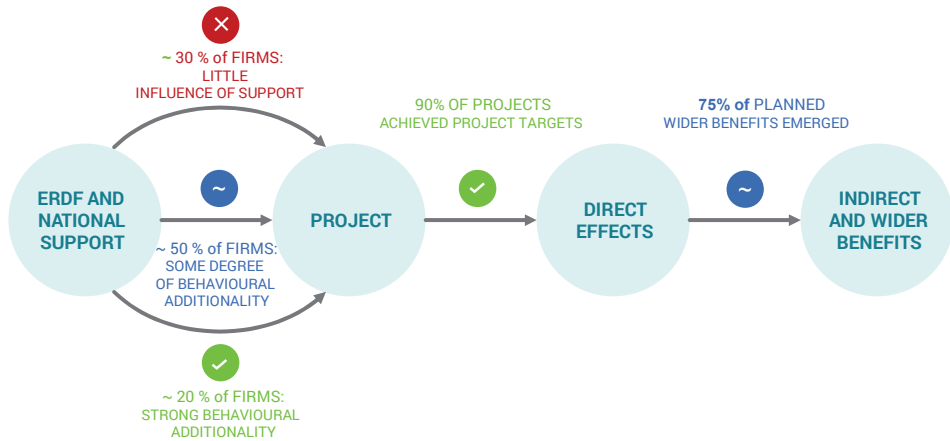
6. Where planned, indirect and wider benefits have emerged in 75% of cases; however, in many cases, the lack of planning for indirect and wider benefits and a focus on inputs and direct results prevented the occurrence of such impacts.

Wherever such effects were foreseen, 75% of desired indirect and wider benefits were at least partly induced. Where spillovers to SMEs were expected to take place, the majority of the projects managed to achieve positive effects. There were similar results for improvements in workforce mobility, working culture, business and social infrastructure, as well as the attraction of other firms to the region. For these wider benefits to emerge, additional conditions, such as sufficient absorption capacity for new knowl-

edge among the benefiting organisations (e.g. SMEs) and pre-existing collaborative ties, had to be in place.

The above conclusions are illustrated by the following figure.

Figure 5. Simplified causal chain and effectiveness of support



Source: KPMG/Prognos, Based on empirical evidence from 45 company case studies, 2016

Conclusion

The requirements of the European Commission, articulated in the guidance for 2014-2020 evaluation plans, will lead to a growth of research that collects plausible evidence on the impact of the implemented public policies.

Experiences from recent years indicate, however, that policy decisions can hardly rely solely on the results of counterfactual impact evaluations. While counterfactual studies can rigorously answer the question whether or not a given policy has had an effect, they have limited capability to identify the underlying reasons behind the success or failure of public interventions. Theory-based evaluations can overcome this limitation of quasi-experimental designs by thinking in terms of 'Theories of Change' that explain how an intervention is supposed to produce its results. Theory-based approaches can in particular support policy decisions by unveiling the complex causal packages that enable interventions to work, and can prove especially powerful if used in combination with counterfactual impact evaluations.

Prior to the WP4 evaluation, counterfactual impact evaluations of large firm support had often found the support to have no impact or an insignificant impact. These

studies could not open the 'black box', i.e. explain the contribution of financial support to the large firms' behavioural change.

The WP4 evaluation used Contribution Analysis, a variant of theory-based evaluations, to assess this contribution. Firstly, the evaluation reconstructed the different Theories of Change of large enterprise support, which were often not explicitly described in programme documents. The Theories of Change described the intended outcomes of interventions, the micro steps that were supposed to lead to these outcomes, key assumptions, external factors, indirect effects and wider benefits of interventions, as well as the nature of causal linkages between these elements. Secondly, the evaluation applied a systematic qualitative testing approach with a multi-respondent design to assess the identified Theories of Change. The empirical work relied heavily on semi-structured interviews with corporate officials (C-level officials, EU fund managers, project managers, employees working with the purchased machinery) and implementing institutions (managing authorities and intermediate bodies). Therefore, triangulation of evidence and critical reflection on the findings, including inputs from independent academic experts and mayors of the localities where the investments took place, were important for the internal validity of this evaluation.

Finally, the evaluation concluded whether the expected results were achieved, whether the Theory of Change as a whole provided a reasonable explanation for the occurrence of these results, and whether there were any other alternative explanations of these results that were not included in the original Theory of Change. Eventually, the Contribution Analysis could make reasonable claims about the contribution of the EU support to the achieved results.

This theory-based evaluation has shown that a comprehensive case study approach, rooted in qualitative empirical research but also building on available counterfactual impact evaluations, was able to create a coherent and credible narrative on the contribution of the EU support to the observed changes, both in the large firms' behavior and in the socio-economic environment of the supported regions. In-depth understanding of the causal chain of large firm support could subsequently form the basis for policy-relevant recommendations that may assist policy makers in designing similar interventions in the future.

About the authors

András Kaszap

András Kaszap leads the EU consulting and evaluation group of KPMG in Hungary as a Senior Manager. He is experienced in planning, delivering and managing evaluations (*ex ante*, interim, *ex post*, thematic and meta evaluations) and business impact assessments. As an evaluation practitioner, he has applied the most state-of-the-art quantitative and qualitative tools and approaches, including theory-based and counterfactual impact evaluations, to create value for clients in both the public and the private sector. He has managed and delivered multiple evaluations in the international context, covering 13 European countries. He is specialized in evaluating the EU Cohesion Policy and the EU Common Agricultural Policy.

Jan-Philip Kramer

PhD in economic geography from the Leibniz University of Hannover and the UC Berkeley. Senior Project Manager in the Business Unit Economy/ Innovation / Region and Head of EU-Office in Brussels working on the full policy cycle from foresight studies, policy strategies and programming, monitoring and evaluation (*ex ante*, *ex post*/ theory-based impact evaluations) to impact assessments. Thematic focus is on ESI Funds (esp. ERDF), Smart Specialisation Strategies (S3) and Clusters as well as scenario-based strategy development for industries and regions. With more than 10 years of experience, Jan-Philip Kramer has worked for several international and German clients, including the European Commission (DG REGIO, DG RTD), the EUREKA Secretariat, the German Federal Ministry of Education and Research, as well as numerous regional clients.

Rafał Trzciński

Graduate of the Institute of Sociology at the University of Warsaw. Since March 2017, he has been the Director of the Centre for Research and Statistical Education of the Central Statistical Office. In the years 2016-2017, he was the Director for research and quantitative analyses in IDEA of Development Foundation. He has many years of experience in research and evaluation of public programs gained during his work in the public sector (Poland Agency for Enterprise Development in the years 2005-2014; Institute for Educational Research in the years 2014-2015), non-governmental organisations (IDEA of Development Foundation in 2014-2017; ISEE Foundation in the years 2013-2016) and cooperation with private operators (EGO S.C., KPMG and others). As part of the conducted research projects, he also cooperated with the European Commission and the World Bank. He specialises in the methodology of statistical causal research. He is an author and co-editor of publications and articles in this area. He also coordinated and, personally, as an analyst, conducted several evaluation studies,

where the counterfactual approach had been applied. He gives training and lectures on the methodology of evaluation studies using the counterfactual approach. He is a member of the Polish Evaluation Society.

Bibliography:

1. Bienias S., Herman-Pawłowska K. (eds.), *Ocena wpływu regulacji. Poradnik dla administracji publicznej*, Warsaw 2015.
2. Chen, H. T., Rossi, P. H., "The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science", *Social Forces* 1980.
3. Górniak J., „Ewaluacja w cyklu polityk publicznych”, in: Mazur S. (ed.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Cracow Univeristy of Economics, Malopolska School of Public Administration, Cracow 2007.
4. Leeuw F., "Reconstructing Program Theories: Methods Available and Problems to be Solved", *American Journal of Evaluation* 2003, no. 24(1).
5. Leeuw F., *Theory-based evaluation*, EVALSED 2012.
6. Leeuw F., "Linking theory-based evaluation and contribution analysis: Three problems and a few solutions", *Evaluation* 2012, 18(3).
7. Martini A., „Wprowadzenie do podstawowych pytań stawianych w ewaluacji oddziaływania interwencji: Czy interwencja coś zmieniła? A jeśli tak, to dla kogo?”, in: Haber A., Trzciński R. (eds.), *Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych*, Polish Agency for Enterprise Development, Warsaw 2011.
8. Mayne J., "Contribution analysis: Coming of age?", *Evaluation* 2012 18: 27.
9. Mouqué D., *What are counterfactual impact evaluations teaching us about enterprise and innovation support?*, DG for Regional and Urban Policy, Brussels 2012.
10. Olejniczak K., *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji śródkresowych INTERREG III*, Centre for European Regional and Local Studies, University of Warsaw, SCHOLAR Publishing House, Warsaw 2008.
11. Pawson, R., Tilley, N., *Realistic Evaluation*, London 1997.
12. Weiss, C. H., "Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families", in: Connell J., Kubisch A., Schorr L. B., Weiss C. H. (eds.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts*, Aspen Institute, New York, NY 1995.
13. White H., "Theory-Based Impact Evaluation: Principles And Practice", *Working Paper n. 3*, International Initiative for Impact Evaluation, New Delhi 2009.

Other:

1. CSO, *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobrych grup kontrolnych*, Warsaw 2015.

2. ISR, *Ewaluacja zmian w systemie wyboru projektów w ramach wybranych działań PO IG 2007-2013*, report commissioned by the Ministry of Infrastructure and Development, Warsaw 2013.
3. KPMG/Prognos, *Support to large enterprises. Final Report – Work Package 4, Ex post evaluation of Cohesion Policy programmes 2007-2013*, Brussels 2016.
4. European Commission, *The Programming Period 2014-2020. Guidance Document on Evaluation Plans*, 2015.
5. PAG Uniconsult, *Wpływ polityki spójności na poziom i jakość zatrudnienia w Polsce*, report commissioned by the Ministry of Infrastructure and Development, Warsaw 2010.

How to better understand the results of counterfactual analyses? Selection of beneficiaries for qualitative research using statistical matching, illustrated by an example of the impact evaluation of the Operational Programme Innovative Economy 2007-2013

Introduction

The aim of this Chapter is to present a technique of selecting beneficiaries for qualitative research, based on statistical matching (hereinafter “technique”), as well as to show the benefits resulting from its application. The technique will be described against the background of counterfactual analyses and based on experience from the application of this technique during the *ex post* evaluation of selected measures of the Operational Programme Innovative Economy, OPIE (hereinafter “study” or “PARP study”). The main purpose of the study was to estimate what level of observed changes in enterprises which implemented projects under OPIE is the direct effect of the Programme.¹²⁶ As part of the study, the effects of Programme Measures 4.1, 4.2, 4.4 and 8.2 were analysed. The estimation covered both the net effects (comparing the beneficiary’s situation after the project completion to a hypothetical scenario in which no support would be granted to that beneficiary) and the relative effects (comparing the beneficiary’s situation after the project completion to a hypothetical scenario in which the beneficiary would have used support under a different measure). The qualitative component was supplemented by qualitative analyses, and its aim was to analyse the success and failure factors of the companies supported under the Measure 4.4.

This Chapter shows how counterfactual analyses may be combined with the theory-based evaluation approach. Additionally, it illustrates certain limitations to the application of counterfactual analyses and presents the way to overcome these limitations.

¹²⁶ *Analiza efektów netto wybranych działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013, z wykorzystaniem podejścia counterfactual impact evaluation, w ramach projektu ewaluacyjnego PARP “Barometr Innowacyjności”, the consortium of the Center for Evaluation and Analysis of Public Policies of the Jagiellonian University and EGO – Evaluation for Government Organizations s.c., commissioned by the Polish Agency for Enterprise Development, Cracow 2015 http://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf, (access on: 24.04.2017)*

The Chapter will further present the context for using the technique. The description refers both to the theoretical background and to the practice of *ex post* evaluation. This is followed by a presentation of the technique itself along with an example of its application in a study. Further, the key results of conducted analyses as well as the conclusions of the study are described. The summary, in turn, presents the conclusions on the use of this technique.

The context of technique application

Counterfactual analyses have been used globally for several decades now, though their first applications in *ex post* evaluations in Poland have occurred relatively recently.¹²⁷ In a situation, where possibilities to conduct a randomized experiment in practice do not exist (or are significantly limited)¹²⁸, counterfactual analyses seem to be a sufficiently good approximation of the assessment of intervention effects.¹²⁹ David Storey described counterfactual analyses as “the evaluator’s paradise” we all should aim for in each *ex post* evaluation.¹³⁰

The process of granting the largest so far amounts of structural assistance to Poland, realised under the EU cohesion policy for the period 2007-2013, having been completed, the authorities commissioning *ex post* evaluation increasingly want to use counterfactual analyses. The growing popularity of counterfactual methods stems, among others, from the recommendations of the European Commission, which has concluded that counterfactual analyses and theory-based evaluation should be the key approaches used in the *ex post* evaluation.¹³¹ In Europe, there is already quite a substantial number of studies using counterfactual methods in the evaluation.¹³² Likewise, one can already talk about a considerable tradition of research in this field with regard to interventions implemented in Poland.¹³³

¹²⁷ The first examples of this type of studies in Poland appeared in 2006 in PARP evaluations concerning the effects of pre-accession programmes Phare 2002 Economic & Social Cohesion (support components for Small and Medium-Sized Enterprises and Human Resource Development).

¹²⁸ E.g. for ethical (entities not selected for support may feel disadvantaged), substantive (where the aim of an intervention is to grant support to specific entities) or financial reasons.

¹²⁹ P. Rosenbaum, *Observational studies*, Springer, New York 2002.

¹³⁰ D. Storey, “Six steps to heaven: Evaluating the impact of public policies to support small businesses in developed economies”, in: D. Sexton, H. Landstrom (ed.), *The Blackwell Handbook of Entrepreneurship*, Blackwell, Oxford 2000, pp. 176-193.

¹³¹ European Commission, *Guidance document on monitoring and evaluation. Concepts and Recommendations*, 2014.

¹³² D. Mouque, *What are counterfactual impact evaluations teaching us about enterprise and innovation support?*, DG for Regional and Urban Policy 2012.

¹³³ See, *inter alia*, PARP evaluations (e.g. the *ex post* evaluation of support for Small and Medium-Sized Enterprises and Human Resource Development under Phare 2002 and 2003 ESC programmes, studies on the net effects of the projects implemented under the Sectoral Operational Programme Improvement of the Competitiveness of Enterprises (SOP ICE) 2004-2006 supporting the increase of the competitiveness of enterprises, analyses of net effects of selected measures under the OP IE 2007-2013 directly addressed to enterprises, the assessment of the impact of co-financing of post-graduate studies and education in the age group 50+ under the Operational Programme Human Capital (OP HC) 2007-2013 or the assessment of the impact of the OP DEP 2007-2013 on the development of entrepreneurship in Eastern Poland).

At the same time, the debate between theoreticians and methodologists on the appropriate strategies for planning and conducting *ex post* evaluations continues. They often underline the need for adopting a broader view on the research approach used in evaluation, as well as tailoring research design each time to match research questions and available resources.¹³⁴ Also, the search continues for conceptual framework, new approaches and techniques, which would allow for an optimal combination of different evaluation paradigms.¹³⁵ For instance, Howard White suggested theory-based impact evaluation, which combines the advantages of counterfactual analyses and theory-based evaluation.¹³⁶

The postulate to combine counterfactual analyses with theory-based evaluation stems from the fact that the former merely enable the estimation of averaged intervention effects. It do not allow, however, for tracing the mechanism that has led to given effects (it do not answer questions about the causes). During the theory-based evaluation, interventions are broken down into a number of hypotheses on how programme resources are transformed through measures into intervention effects.

The theory-based evaluation may complement counterfactual analyses, allowing to better understand the estimated effects of the intervention. Net effects values rarely correspond with the observed (gross) changes in the area covered by a given intervention, and also their directions may be different from the initial expectations (positive or negative), which in turn causes many difficulties as to the interpretation. It may also turn out that even if an average result of an intervention is zero, a certain (even relatively small) group of beneficiaries has achieved outstanding effects. What is more, these effects may constitute a sufficient argument for the overall effectiveness of a given intervention. Complementing the results of counterfactual analyses with conclusions from a theory-based evaluation will also help in effectively communicating the intervention evaluation outcomes. Proper interpretation of analyses results (from results to conclusions and recommendations) frequently raises a number of objections and is not always clear to the stakeholders. On the other hand, evaluation results which are not consistent with the expectations of recipients of the study (also with the common sense) may result in them being inclined to negate. However, the formulation of accurate hypotheses concerning the expected direction of the results as well as contextual interpretation of the effects recorded is useful in building a narrative which is easily comprehensible for recipients of the study.

¹³⁴ E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, B. Befani, *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, Department for International Development 2012

¹³⁵ Different theory-based evaluation approaches were described in detail in this publication under the heading "A comprehensive review of the theory-based evaluation concept" by Katarzyna Hermann-Pawłowska and Paulina Skórska.

¹³⁶ H. White, "Theory-based impact evaluation: principles and practice", *Journal of Development Effectiveness* 2009, vol. 3, pp. 271–284.

To conclude, as it stems from theory and existing evaluation practices, counterfactual analyses should continue to be taken into account as one of the best approaches used for *ex post* evaluation. At the same time, it is worth to consider using other techniques, where their application is possible and appropriate. This will enrich the understanding of the raw intervention effects. The final design of a research should, however, depend on the informational needs (expressed in terms of specific research questions) and resources available for a project implementation. An example of such an approach in evaluation, i.e. the approach that improves the understanding of intervention effects, as estimated using counterfactual methods, and consistent with the research questions and resources, is described in further parts of this Chapter.

A vital criterion for the usefulness of the technique presented in this Chapter is its contribution towards a better understanding of counterfactual analyses results. This is why our description of the technique presented in this Chapter is deeply embedded in the context of the study in which it was applied. In addition to the complementarity of the said technique to the counterfactual analyses applied in evaluation, also using the specific experiences drawn from the entire study allows for illustrating the challenges described earlier, related in general to the application of counterfactual analyses in evaluation.

The technique of interest is illustrated with an example of its use for interpretation of the effects of the OPIE Measure 4.4 – “New investments of high innovative potential”. The aim of this Measure was to support enterprises making new investments (and implementing consultancy and training projects necessary for their realisation) involving the purchase of innovative technological solutions. The key expected outcomes include improved innovativeness (including a higher potential to engage in research and development) of enterprises and their higher competitiveness. The Measure was selected for this type of analyses due to, above all, its relatively large budget (and high expectations with regard to the effects produced), compared to other measures implemented under the OPIE.

The technique applied against methodological background of the study

The procedure for estimating net effects

Due to legal limitations, statistical offices as a general rule do not make any microdata available to external entities (the Polish Central Statistical Office does so as well¹³⁷). In the course of previous counterfactual evaluations of programmes offering subsidies for Polish firms, two procedures for using the microdata gathered and stored by the

¹³⁷ Cf. the Act of 29 June 1995 on Official Statistics (Dz. U [Journal of Laws] of 1995, No 88, item 439, as amended).

CSO have been developed so far.¹³⁸ The first one was used in the study discussed in this Chapter. In simplified terms, the procedure was as follows:

1. At the beginning, unsuccessful applicants were statistically matched to beneficiaries, based on the data collected in the process of application for subsidy. Each beneficiary – as far as it was possible – was to be matched with at least two similar unsuccessful applicants, which increased the chance of later obtaining a complete data for each matched pair (the beneficiary and the control entity).
2. The list of beneficiaries and their “statistically twin companies”, selected among unsuccessful applicants, was submitted to the CSO with a request to calculate the averaged effect indicators of interest in both groups for selected years (that is, year of application, years following the completion of financed projects, and the year immediately preceding the completion; the same calendar years were assumed for control entities as for their matched counterparts).
3. The CSO identified companies from the submitted list based on REGON and NIP numbers, performed appropriate calculations, and then provided the relevant averaged values for identified groups of enterprises (i.e. the group of beneficiaries and the control group), while maintaining statistical confidentiality (i.e. requirements on the anonymity and safety of microdata).
4. Finally, based on data obtained from the CSO, calculation of intervention effects and their interpretations were made, using the difference in differences (DID) technique.¹³⁹

Given the small size of the sample obtained from the CSO, based on which it was possible to estimate the effects, and the limited possibilities for carrying out a balance check (i.e. control of similarity) across compared groups using the procedure described above, this Chapter presents the estimations of net effects for Measure 4.4 OPIE obtained in the course of a different study, in which one of the authors of this Chapter has participated.¹⁴⁰

The net effects for Measure 4.4 OPIE presented here were obtained according to the second (alternative) procedure for estimating net effects, using CSO’s microdata. The procedure was as follows:

1. At the beginning, a list of intervention beneficiaries was submitted to the CSO.
2. Based on the beneficiaries’ REGON and NIP numbers, the CSO identified these enterprises and marked them in its databases.

¹³⁸ Data for the purpose of counterfactual analyses may be also obtained directly from enterprises through surveys. This solution is, however, more costly and causes additional burden on companies.

¹³⁹ As a result of the analysis described, for 99 submitted pairs of beneficiaries of Measure 4.4 and ineffective applicants, the CSO was able to identify only 31 pairs for which it had a full data package in the SP reports (“Annual survey of the enterprises”) database.

¹⁴⁰ Central Statistical Office of Poland, *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych*, Warszawa 2015.

3. Next, the evaluator developed appropriate codes for data analysis (script) and handed them to the CSO.
4. The CSO, using the same statistical analysis software, ran the code which, in step one, matched the pairs of “statistical twins” to the beneficiaries (from the population of remaining companies whose data are at CSO’s disposal), and then, in the second step, calculated specific indicators of intervention effects.
5. Finally, the CSO submitted the results of analyses to the evaluator who assessed the quality of the match obtained and the stability of estimated effects (the procedure may be repeated until a satisfactory match and stable estimated effects are obtained).

In this case, analyses were conducted directly on the CSO’s microdata. Statistical confidentiality is still maintained, i.e. the external entity has no direct access to microdata, and only designs and remotely supports the process of analyses, conducted by the CSO’s employees.

The procedure for estimating relative effects

In addition to net effects, estimations of relative effects may be useful in the evaluation of interventions. For instance, such estimations allow for determining the effectiveness of a given intervention, compared to another one with similar objectives and addressed to a similar target group. Within the scope of the study in question, Measure 4.2 OPIE had relatively similar objectives as Measure 4.4.¹⁴¹ Enterprises from the target group (beneficiaries of Measure 4.4) were paired up (via statistical matching) with similar companies from the reference group (beneficiaries of Measure 4.2). In the process of matching the companies being compared, control variables from the same sources as in the estimation of net effects were used, that is, application data. The effects were calculated using the DID technique for the year when the funded project was completed (Z) and for the subsequent year (Z+1). Information about the values of effect indicators presented in this paper was obtained from the OPIE evaluation study – “Innovation Barometer”.¹⁴²

¹⁴¹ The objective of Measure 4.2 was to support R&D activity of enterprises and to support developing industrial and utility designs as well as their implementation. “Both the scope of possible financing (based primarily on the General Block Exemption Regulation [...] No 800/2008 – GBER), the target group which could benefit from the support (high growth enterprises – SMEs as well as large ones) and common objectives at the priority axis level coincided [...]” between Measures 4.4 and 4.2 OP IE (Cf. J. Pokorski, “Building the innovation potential of enterprises in Poland – the outcomes of the Innovative Economy Operational Programme”, in: P. Zadura-Lichota (ed.), *Innovative entrepreneurship in Poland. Revealed and hidden potential for innovation in Poland*, PARP, Warszawa 2015).

¹⁴² It was an on-going evaluation conducted by PARP in 2011-2015 among the beneficiaries (entrepreneurs) of selected OP IE measures (Cf. 2. „Barometr Innowacyjności. Ewaluacja *on-going* działań Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013, skierowanych do przedsiębiorstw”, Consortium of ARC Rynek and Opinia sp. z o.o. and Exacto sp. z. o.o., commissioned by PARP, Warszawa 2015 www.badania.parp.gov.pl/barometr-innowacyjnosci (access on: 24.04.2017).

The procedure of selecting beneficiaries for qualitative interviews

Quantitative estimations of (net and relative) effects allow for assessing whether a given intervention was successful or not. However, qualitative research is necessary to understand the reasons for the success or failure. This Chapter presents the procedure of selecting beneficiaries for qualitative interviews, using statistical matching. In this study, in-depth interviews were conducted with the contrast pairs of beneficiaries of Measure 4.4 OPIE – one of beneficiaries transformed the support to a market success, and the other did not. This procedure was meant to identify key success and failure factors of the intervention in question, by referring to a specific situation of a given beneficiary. The procedure of matching contrasting pairs of beneficiaries was as follows:

1. At the beginning, a criterion of success was defined. It was a modified¹⁴³ return on sales (ROS), calculated as the ratio of earnings before taxes (EBT) and net revenue from sales. Return on sales is one of the primary indicators used in assessing the financial situation of an enterprise and its competitiveness, hence the decision to choose it as the criterion of intervention success.
2. A regression model was specified, in which the value of ROS one year after the completion of the supported investment (Z+1) was predicted, using the ROS value in the year in which the investment begun (R) as well as a set of variables describing the enterprise and a number of detailed information on the funding received, including: the start and completion date of the investment, the type of business activity (limited liability company, public limited company, other), the size of the company (up to 49 employees, 50-249, 250 and more), the sector or main activity (production, other), the year when the firm was established, the percentage of own contribution to the investment, the amount of support received, and the voivodeship in which the enterprise has its registered office. In addition to substantive criteria, the selection of control variables in the model was dictated by pragmatic (the availability of data) and statistical (the significance of variables in the model) criteria. The outcomes of model diagnostics turned out to be satisfactory, thus ensuring correctness of residuals estimation. The application of the regression method allowed for controlling the potential of a given enterprise in the model, which resulted in obtaining a more objective value of success indicator. Regression model residuals constitute an individual indicator of success for each company, more accurate than, for instance, comparing the situation of the company to an average situation of other companies.

¹⁴³ The modification of the indicator involved substituting the net profit, which is usually found in the numerator of the indicator, with gross profit (that is, profit before the income tax). This was due to the availability of data. For the "Innovation Barometer" studied gross profit and not net profit, which was the result of the desire to identify the impact that the intervention had on the profit, without taking into account changes concerning taxation.

3. In the next step, residuals from the regression model, i.e. the differences between the observed and predicted ROS level one year after the completion of the funded investment (Z+1), were saved. In this case, residuals may be interpreted as the measure of success – the enterprise has transformed the funding received into a market success (obtained a higher ROS than it was predicted in the model based on the potential of the enterprise “at the start”) or has failed to do so (obtained a lower ROS than expected). This logic for interpreting regression residuals is taken from the Educational Value Added (EVA) modelling used for measuring effectiveness of schools and teachers.¹⁴⁴ In Poland, the EVA method for schools was developed in 2005-2015, as part of a project implemented by the Educational Research Institute.¹⁴⁵
4. Beneficiaries of Measure 4.4 were ranked according to residual values. This way, two groups were formed: successful companies (above the third quartile) and companies that failed to turn the funding received into an increase of the ROS value (below the median). Enterprises falling above the third quartile and below the median were kept for further analyses (as it was intended to retain the proportion of 1:2 in the prepared sample for qualitative research, i.e. one successful company for two unsuccessful companies), while the rest were deleted from the draft database.
5. As the next step, the statistical matching procedure was applied, which consisted in finding (matching) similar companies (“statistical twins”) differing only by the fact of being classified into the group of successful companies or to the group of unsuccessful companies. Statistical matching may generally use the same set of covariants as in the regression model which served for obtaining residuals. In this case, the Mahalanobis distance matching (MDM) was applied between companies, according to ROS values and to the amount of the received co-financing (continuous variables were recoded into 5 categories, based on quintiles) together with the simultaneous exact matching on the following variables: the start and

¹⁴⁴ R. Chetty, J. N. Friedman, J. E. Rockoff, “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates”, *American Economic Review* 2014, vol. 104, pp. 2593–2632.

C. T. Clotterfeld, H. F. Ladd, “Recognizing and Rewarding Success in Public Schools”, in: H. F. Ladd (ed.), *Holding Schools Accountable: Performance-Based Reform in Education*, Brookings Institution, Washington, DC 1996.

M. Ehlert, C. Koedel, E. Parsons, M. J. Podgursky, “The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence From School- and Teacher-Level Models in Missouri”, *Statistics and Public Policy* 2014, vol. 1, pp. 19–27.

T. J. Kane, J. E. Rockoff, D. O. Staiger, “What does certification tell us about teacher effectiveness? Evidence from New York City” *Economics of Education Review* 2008, vol. 27, pp. 615–631.

H. F. Ladd, R. P. Walsh, “Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right”, *Economics of Education Review* 2002, vol. 21, pp. 1–17.

W. J. Webster, G. H. Olson, “A Quantitative Procedure for the Identification of Effective Schools”, *The Journal of Experimental Education* 1988, vol. 56, pp. 213–219.

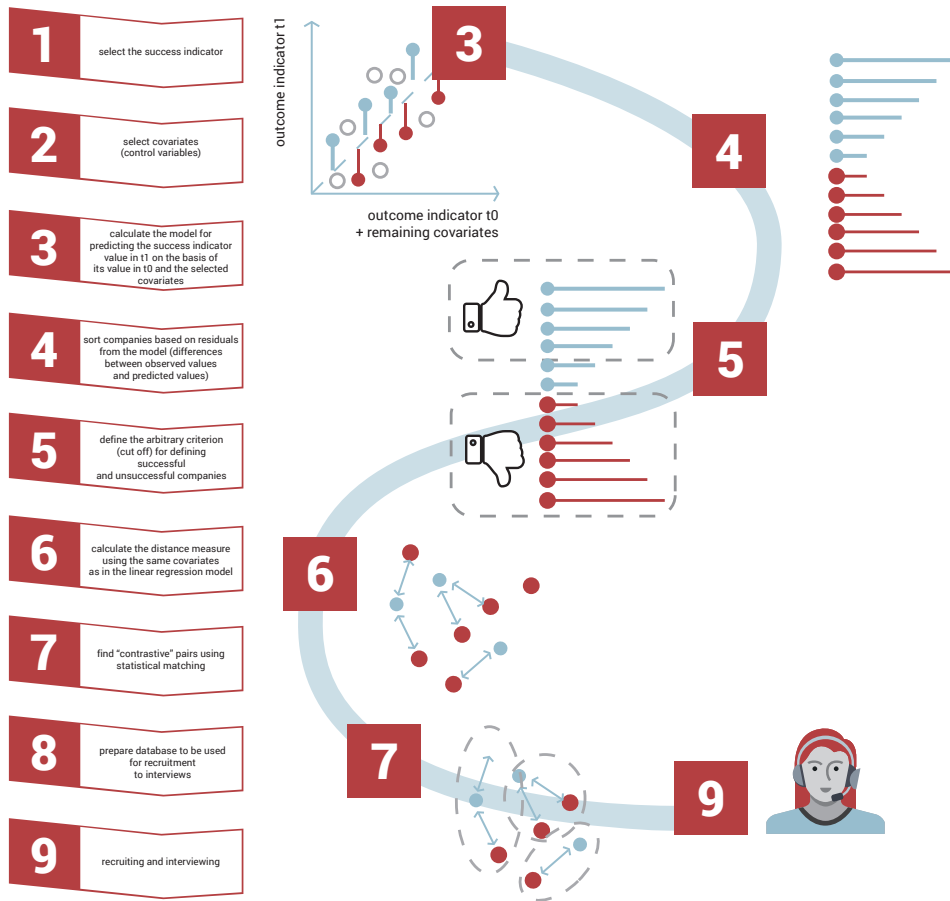
¹⁴⁵ Cf. The Educational Research Institute – project titled “Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych” www.ewd.edu.pl, (access on 24.04.2017).

completion year of the investment, the size of the company, and the sector of main activity.

6. Finally, for every “successful company”, where possible, two companies were selected (twin companies in terms of defined set of characteristics) which were unsuccessful. In result of this matching, not only pairs of companies were obtained, but also triplets. Consequently, if the first “unsuccessful company” matched with the “successful company” refused to give an interview, it was possible to survey the second company.

Selecting beneficiaries of Measure 4.4 of OPIE for the in-depth interviews according to the abovementioned procedure enabled highlighting (i.e. sharpen, considering the proper contrast) the success and failure factors by carrying out relatively few interviews, which lowered the costs of the study significantly (the scale of the in-depth study on enterprises was limited, which is hard to achieve in other cases, that is of research projects which do not use this kind of precise selection). Finally, 6 individual in-depth interviews (IDI) were conducted: 3 with successful companies and 3 with unsuccessful ones. The interviews were conducted with representatives of companies who were best informed on the possible impact of received co-financing on the competitiveness of their companies. These were most frequently members of the managing body or employees directly responsible for project implementation. Figure 1 presents a graphic representation of this procedure.

Figure 1. Scheme of beneficiaries selecting procedure for qualitative research with the use of statistical matching



Source: Own elaboration.

Statistical matching

In the discussed study, the statistical matching was used to select the sample for estimating net effects and relative effects, as well as for selecting beneficiaries for qualitative interviews. As the statistical matching was used in each of the three key elements of the study herein discussed, this method, especially its most popular technique, i.e. propensity score matching (PSM), is presented briefly.

Statistical matching is a group of analytical techniques which may be considered as good practice for the evaluation research. "It is based on the intuitively attractive idea of contrasting the outcomes of programme participants with the outcomes of »com-

parable« nonparticipants. Differences in the outcomes between the two groups are attributed to the programme.¹⁴⁶ Statistical matching, by means of statistical control of the observed variables, attempts at imitating the randomised experiment which is considered as the “golden standard” in proving causation.¹⁴⁷

PSM is the most frequently used statistical matching technique.¹⁴⁸ Readers interested in the topic may refer to the publications introducing different techniques based on propensity score (PS), as well as to discussion on differences between those techniques and the use of multiple regression for estimating intervention outcomes¹⁴⁹, and also to publications on the theoretical basis for first elaborations on statistical matching techniques and examples of their application in research practice in various fields¹⁵⁰. Practical guidelines on how to apply PSM are also available¹⁵¹. Studies in the Polish language are also worth recommending¹⁵². Critical opinions regarding techniques based on the PS vector should also be noted, making one alert while applying the technique and interpreting its results. For example, using data from a large educational experiment, Elizabeth Wilde and Robinson Hollister proved, that experimental conditions are not reproduced effectively enough by using PS based on PS techniques¹⁵³. Moreover, some authors (e.g. Gary King and Richard Nielsen¹⁵⁴) warn that incautious application of PSM may cause stronger bias and model dependence. They suggest careful selection of control variables and exact matching on all of them, or at least on those most important. These suggestions were used in the analyses conducted under the discussed study.

¹⁴⁶ J. J. Heckman, H. Ichimura, P. Todd, “Matching As An Econometric Evaluation Estimator”, *Review of Economic Studies* 1998, vol. 65, p. 261.

¹⁴⁷ T. Dunning, “Improving Causal Inference: Strengths and Limitations of Natural Experiments”, *Political Research Quarterly* 2008, vol. 61, pp. 282–293.

¹⁴⁸ P. Rosenbaum, D. Rubin, “The central role of the propensity score in observational studies for causal effects”, *Biometrika* 1983, vol. 70, pp 41–55.

S. L. Morgan, C. Winship, *Counterfactuals and causal inference*, Cambridge University Press 2014

¹⁴⁹ P. C. Austin, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies”, *Multivariate Behavioral Research* 2011, vol. 46, pp. 399–424.

¹⁵⁰ E. A. Stuart, “Matching methods for causal inference: A review and a look forward”, *Statistical Science: A Review journal of the Institute of Mathematical Statistics* 2010, vol. 25, pp. 1–21.

¹⁵¹ M. Caliendo, S. Kopeinig, “Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys* 2008, vol. 22, pp. 31–72.

¹⁵² P. Strawiński, *Propensity Score Matching. Własności małopróbkowe*, Wydawnictwo Uniwersytetu Warszawskiego, Warsaw 2014; R. Trzciński, *Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2009.

¹⁵³ E. T. Wilde, R. Hollister, “How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment”, *Journal of Policy Analysis and Management* 2007, vol. 26, pp. 455–477.

¹⁵⁴ G. King, R. Nielsen, *Why Propensity Scores Should Not Be Used for Matching*, 2016.

Research outcomes

Counterfactual analyses

The outcomes for beneficiaries of Measure 4.4 of OPIE have turned out to be contradictory to the effects foreseen in the intervention logic, which were originally expected by the evaluation team. For most of indicators, beneficiaries under analysis did not achieved better results than the unsuccessful applicants who applied for support under this Measure (net effects) or the Measure 4.2 of OPIE beneficiaries¹⁵⁵ (relative effects). The impact of intervention in terms of increased innovativeness and export was an exception. In the case on unsuccessful applicants, the net effect for Measure 4.4 was positive for innovative activity indicators¹⁵⁶ and export activity indicators¹⁵⁷. The most surprising fact was that beneficiaries of Measure 4.4 reported a loss more frequently than those of Measure 4.2 (both, in the year of finalising the investment and in the following year). The opposite outcome should be expected for a significantly higher support value in the case of Measure 4.4.

Table 1. Statistically relevant net and relative outcomes for Measure 4.4 of OPIE

	CSO study		PARP study			
	Net effect 2008-13	Standard deviation	Relative effect for 4.2 (t0-t1 ^a)	Standard deviation	Relative outcome for 4.2 (t0-t2 ^b)	Standard deviation
Enterprises not making any loss (%)	not applicable	not applicable	-26**	13	-33**	13
Exporting enterprises (%)	11*	6	-1	3	0	0
Innovative enterprises (%)	20**	8	not applicable	not applicable	not applicable	not applicable

Note: * statistical significance at the level of $p < 0.01$; ** $p < 0.05$; ^a the investment completion year compared to the start year; ^b the year after completion year of the investment compared to the start year, the positive value means that the indicator value for the 4.4 beneficiaries is higher than for the unsuccessful applicants (net effect) or for beneficiaries of Measure 4.2 (relative effect); the negative value indicates the reverse situation.

¹⁵⁵ The support value under the 4.2 Measure was on average four times lower than under the 4.4 Measure. Therefore, lower outcomes could be expected in this case.

¹⁵⁶ For the purpose of designation of entities conducting innovative activity, it has been assumed, in line with the CSO study (2015), that an entity conducting innovative activity is every company for which the sum of costs of finalised development work and of the rights to inventions, patents and trademarks value is above zero.

¹⁵⁷ In the CSO study (2015), conducting export activity was identified by means of income from export sales.

Having such results, evaluator is faced with a difficult challenge. S/he should be able to interpret them and to provide on their basis some recommendations for the future interventions of this type. In the case of the discussed study, the qualitative interviews proved to be useful.

Conclusions from the interviews with beneficiaries

In the qualitative part of the study, a number of factors were identified, which could potentially influence the translation of results of investments supported under Measure 4.4 into changes in the profitability of beneficiary enterprises (ROS value). The identified factors are related to the characteristics of beneficiaries, the specificity of sector in which they operate, the way of investment implementation, as well as to the social and economic environment (including economic situation). The characteristics of beneficiaries depend on the competence of their employees, especially of the managerial staff, and on the strategy pursued by the enterprise. Higher competencies and a coherent strategy, focused on introducing innovative products on the market and providing for complementary development investments, were factors conducive for translating the OPIE grants into higher enterprise's profitability.

Also the specificity of the sector in which a given company operates, which had an impact on enterprise's strategy, was of significant importance for translating investment outcomes into company's success. In some sectors introducing new products on the market is more important for the competitiveness of the enterprise than in others. Some beneficiaries pointed out the dominating role of other factors: the volume of production capacities, as well as prices and quality of the offered goods and services. In specific sectors where innovativeness is less important for the competitiveness of the company, beneficiaries are more inclined to treat projects supported under Measure 4.4 simply as investments, not in terms of boosting innovativeness.

As regards the manner of implementing a new product on the market, attention should be paid to good management of the entire project supported under Measure 4.4, as well as to the thoughtful planning for introducing a product or a service on the market. The point is to carry out adequate public relations and marketing activities and to select adequate distribution ways and channels for new products or services. Some beneficiaries of Measure 4.4 are manufacturing companies operating in the business-to-business (B2B) segment. Introducing new product, especially for a new target group, requires different communication and distribution channels. On the basis of respondents' answers, it can be concluded that in several of the analysed enterprises, consistency of implementation of such measures was missing. When developing a new product, companies frequently failed to devise a new and adequate way of introducing it into the market. Finally, it is worth noting that changes in the social and economic environment had an impact on the demand for innovative products. Apart from the general deterioration in the economic situation at the time when projects financed

under Measure 4.4 were finalised, the situation in individual sectors was also important. Frequently, the deteriorating situation in a given sector had a significant impact on the economic situation of the enterprise and on the final outcomes of the project.

As one may see, several factors contributed to the fact that some of beneficiaries of OPIE Measure 4.4 recorded a relatively small positive impact of investments supported under this Measure on their enterprise profitability. The most important factors which could be influenced by beneficiaries included the lack of appropriate monitoring of the situation on the market in the context of demand for a new product and of planning appropriate ways to reach a client, as well as the lack of any coherent and consistently implemented strategy which would focus on introducing new products and services on the market.

The clear compliance in assigning beneficiaries to a given category (success or failure) with their subjective perspective on results of using the grant, was an important conclusion drew from the interviews. In the course of interviews, beneficiaries assigned to the success category based on the quantitative data, assessed relatively high the way in which they had used the obtained support for the enterprise development. Similarly, companies assigned into the failure category based on the quantitative data, when being interviewed stated that they could have used the grant in a more effective way. This compliance confirmed both, the proper selection of ROS as a success indicator and the accuracy of the approach (technique) itself.

Conclusion

The technique of selecting beneficiaries for qualitative study using statistical matching, as presented in this Chapter, proved to be highly useful for better understanding of the outcomes of intervention (net and relative effects). It allowed the authors to identify the possible causes for success or failure of the intervention under analysis. It served for deeper understanding of the intervention and as such it turned out useful for refining the theory behind a similar type of programmes in the future.

Some comments, however, should be made. The number of conducted interviews was effective, however relatively small. In future applications of this technique, it would be worth to consider increasing the number of interviews. This would allow for applying additionally the qualitative comparative analysis (QCA), as a technique for data analysis.¹⁵⁸ It should also be remembered that the result obtained from counterfactual analysis may also be a consequence of different factors, not included in the model, e.g. a

¹⁵⁸ The application of such a technique was planned under the evaluation of the Public Assistance Programme of the National Centre for Research and Development conducted by the consortium of IDEA of Development Foundation, IMAPP sp. z o.o. PAG Uniconsult sp. z o.o. and the Center for Evaluation and Analysis of Public Policies of the Jagiellonian University. The QCA technique has been described in this publication in the chapter entitled "Qualitative comparative analysis in theory-based impact evaluation" by Barbara Worek and Marcin Kocór.

relatively short time of measurement after receiving support (a possibility of deferred outcomes). The results obtained using the discussed technique do not bring unequivocal solutions as to the results of counterfactual analyses, however they provide an opportunity for better understanding those results and consequently for strengthening conclusions that were drawn from them. In this sense, the use of the analysed technique, and more precisely of its qualitative component, should be treated as complementary to the counterfactual analyses, and not as its alternative. An important characteristic of the technique discussed in this Chapter is its low cost. However, it is worth remembering, that to conduct interviews with the representatives of contrasting pairs of beneficiaries, the consent of both of them is required. This, in turn, involves relatively sophisticated motivational techniques that need to be elaborated (in this case such techniques included: referring to the scientific nature of the research project and to its social utility, as well as to the unique qualities of respondent in the selection process).¹⁵⁹ Moreover, it should be emphasised that the described technique should be applied only in the situation when we want to answer research questions that are adequate for this technique, i.e. related to the causes of differences in beneficiaries outcomes. For other research questions, different ways of combining counterfactual analysis with qualitative research or case studies should be considered.

The discussed technique, providing tools for better understanding of outcomes of counterfactual analyses, was also used, in a slightly modified form, in other *ex post* evaluations in Poland, e.g. to interpret the influence of Operational Programme Development of Eastern Poland 2007-2013 on entrepreneurship at the local level.¹⁶⁰ This proves not only the utility of the approach itself, but also its flexibility in terms of subject of the study and the research questions.

¹⁵⁹ The interview conducted with only one beneficiary is not sufficient to be used for further analyses.

¹⁶⁰ Cf. "Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej", the consortium of EVALU sp. z o.o., WYG PSDB sp. z o.o., and EGO – Evaluation for Government Organizations s.c., commissioned by PARP, Warsaw 2016.

About the authors

Maciej Koniewski

Specialist in research methodology and quantitative data analysis methods. He started his work in the market research industry by implementing more than 50 national and international projects (i.a., for John Deere and Deutsche Telekom). Since 2011, he has been cooperating with the Jagiellonian University's Centre for Evaluation and Analysis of Public Policies in Cracow (i.a., on evaluation projects of the Polish Education Evaluation System, support for building an education evaluation system in Saudi Arabia, counterfactual impact evaluation of the measures of the Innovative Economy Operational Programme 2007-2013 (OPIE) for the Polish Agency for Enterprise Development). He was also a member of a team consulting for the Central Statistical Office with regard to using counterfactual analysis to investigate the effectiveness of the OPIE and Regional Operational Programmes. In 2013-2015, he worked with a team of Educational Value Added at the Educational Research Institute, where he was involved in, i.a., analysis and reporting of results of current research on the relevance of the EVA indicators, training and popularisation of the method. He is an author and co-author of several publications on educational research and educational measurement.

Seweryn Krupnik

Director of the Jagiellonian University's Centre for Evaluation and Analysis of Public Policies in Cracow. He is PhD in Sociology (his doctoral thesis was devoted to the evaluation of investment grants for enterprises). He is a certified project manager and coach and a member of the Audit Committee of the Polish Evaluation Society. His scientific interests are related to public policies aimed at supporting enterprises (in particular, direct support, social enterprises, knowledge transfer from universities to enterprises), the use of the theory-based evaluation, analysis of cause-and-effect relationships and the use of institutional theories in understanding the reality. For several years, he has been combining the scientific activity with the implementation of projects on public policy analyses.

Bibliography:

1. Austin P. C., "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies", *Multivariate Behavioral Research* 2011, vol. 46.
2. Caliendo M., Kopeinig S., "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys* 2008, vol. 22.
3. Chetty R., Friedman J. N., Rockoff J. E., "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates", *American Economic Review* 2014, vol. 104.
4. Clotterfeld C. T., Ladd H. F., "Recognizing and Rewarding Success in Public Schools", in: Ladd H. F. (ed.), *Holding Schools Accountable: Performance-Based Reform in Education*, Brookings Institution, Washington, DC 1996.
5. Dunning T., "Improving Causal Inference: Strengths and Limitations of Natural Experiments", *Political Research Quarterly* 2008, vol. 61.
6. Ehlert M., Koedel C., Parsons, E., Podgursky M. J., "The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence From School- and Teacher-Level Models in Missouri", *Statistics and Public Policy* 2014, vol. 1.
7. Heckman J. J., Ichimura H., Todd P., "Matching As An Econometric Evaluation Estimator", *Review of Economic Studies* 1998, vol. 65.
8. Kane T. J., Rockoff J. E., Staiger D. O., "What does certification tell us about teacher effectiveness? Evidence from New York City", *Economics of Education Review* 2008, vol. 27.
9. King G., Nielsen R., *Why Propensity Scores Should Not Be Used for Matching*, 2016. www.polmeth.wustl.edu/files/polmeth/psnot4.pdf (access on 24.04.2017).
10. Ladd H. F., Walsh R. P., "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right", *Economics of Education Review* 2002, vol. 21.
11. Morgan S. L., Winship C., *Counterfactuals and causal inference*, Cambridge University Press 2014.
12. Mouque D., *What are counterfactual impact evaluations teaching us about enterprise and innovation support?*, DG for Regional and Urban Policy 2012.
13. Pokorski J., *Towards Innovative Economy. Effects of Grants to Enterprises in Poland*, Polish Agency for Enterprise Development, Warsaw 2011.
14. Pokorski J. "Building the innovation potential of enterprises in Poland – the outcomes of the Innovative Economy Operational Programme", in: P. Zadura-Lichota (ed.), *Innovative entrepreneurship in Poland. Revealed and hidden potential for innovation in Poland*, Polish Agency for Enterprise Development, Warsaw 2015).
15. Rosenbaum P., *Observational studies*, Springer, New York 2002.
16. Rosenbaum P., Rubin, D., "The central role of the propensity score in observational studies for causal effects", *Biometrika* 1983, vol. 70.

17. Stern E., Stame N., Mayne J., Forss K., Davies R., Befani B., *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*, DFID: Department for International Development 2012.
18. Storey D., "Six steps to heaven: Evaluating the impact of public policies to support small businesses in developed economies", in: Sexton D., Landstrom H. (ed.), *The Blackwell Handbook of Entrepreneurship*, Blackwell, Oxford 2000.
19. Strawiński P., *Propensity Score Matching. Własności małopróbkowe*, Wydawnictwo Uniwersytetu Warszawskiego, Warsaw 2014.
20. Stuart E. A., "Matching methods for causal inference: A review and a look forward", *Statistical Science: A Review journal of the Institute of Mathematical Statistics* 2010, vol. 25.
21. Trzciniński R., *Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych*, Polish Agency for Enterprise Development, Warsaw 2009.
22. Webster W.J., Olson G.H., "A Quantitative Procedure for the Identification of Effective Schools", *The Journal of Experimental Education* 1988, vol. 56.
23. White H., "Theory-based impact evaluation: principles and practice", *Journal of Development Effectiveness* 2009, vol. 3.
24. Wilde E. T., Hollister R., "How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment", *Journal of Policy Analysis and Management* 2007, vol. 26.

Other:

1. *Barometr Innowacyjności. Ewaluacja on-going działań Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013, skierowanych do przedsiębiorstw*, Consortium of ARC Rynek and Opinia sp. z o.o. and Exacto sp. z o.o., commissioned by PARP, Warsaw 2015 http://badania.parp.gov.pl/images/badania/Raport_Barometr_Innowacyjnosci.pdf (access on 24.04.2017).
2. *Analiza efektów netto wybranych działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013, z wykorzystaniem podejścia counterfactual impact evaluation, w ramach projektu ewaluacyjnego PARP "Barometr Innowacyjności"*, the consortium of the Center for Evaluation and Analysis of Public Policies of the Jagiellonian University and EGO – Evaluation for Government Organizations s.c., commissioned by Polish Agency for Enterprise Development, Cracow 2015 http://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf (access on 24.04.2017)
3. *Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej*, the consortium of EVALU sp. z o.o., WYG PSDB sp. z o.o., and EGO s.c., commissioned by Polish Agency for Enterprise Development, Warsaw 2016.
4. European Commission, *Guidance document on monitoring and evaluation. Concepts and Recommendations*, 2014.

5. Central Statistical Office, *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych*, Warsaw 2015 www.stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/5807/1/1/raport_rpo_poig.pdf (access on 24.04.2017).
6. The Act of 29 June 1995 on Official Statistics (Dz. U. [Journal of Laws] of 1995, No 88, item 439, as amended).

Stratified Propensity Score Matching – quasi-experimental research design for theory-based impact evaluation with lacking dependent variable

Introduction

Despite a multitude of studies using impact evaluation logic and analysing the efficiency of funds transferred to Poland, the majority of them fail to provide quantified evidence on the impact of the funds on economic development at the local level. Estimation of the net effect of absorbed funds at the NUTS 4 and NUTS 5 level is problematic for methodological reasons. Firstly, the possibility to apply macroeconomic models, which are often used for estimating the impact of the EU funds at the national or voivodeship level, is limited at the level of local economies. Secondly, the fact that in principle all Polish NUTS 5 used the funds in various forms does not allow to apply a quasi-experimental approach based on Propensity Score Matching (PSM). This is because the use of logistic regression on which PSM is based is impossible when each unit was, to a lesser or greater extent, subject to intervention, the effects of which we want to evaluate.

The aim of this article is to present assumptions and examples of using an innovative method called Stratified Propensity Score Matching (SPSM), which allows to calculate the net impact of the EU funds on economic development of NUTS 5 and NUTS 4. Contrary to estimation of relative causal effects (RCE), it still involves the real estimation of net effects. While RCE requires the existence of at least two groups of funds or interventions which are then compared with one another, SPSM offers a possibility to estimate the effects of individual funds or interventions.

The article consists of several parts. The first one presents research approaches used for impact evaluation which provide quantitative data on net effects of the projects supported from the EU Cohesion and Common Agriculture Policies funds. The limitations of those approaches in impact evaluation studies focused at the local level will be discussed. The next part presents the SPSM method which addresses the methodological challenges faced by evaluations of the EU funds' impact on development of local economies. The final part includes two examples of ex post evaluations where SPSM was successfully adapted to estimate the net impact of funds absorbed by Polish NUTS 5. In both of presented studies, conclusions from

counterfactual analyses were supplemented with in-depth qualitative research (partly using the identified pairs of “twin” territorial units), which allowed to explain the identified differences in the impact of the absorbed funds. The use of the SPSM method, combined with qualitative methods, allowed for a comprehensive analysis of the effectiveness of intervention at the local level.

Impact evaluation – research designs used for estimation of economic effects of the EU funds

The impact evaluation, which as a rule aims at assessing the impact of public intervention, allows to provide quantified evidence only when two research designs are applied, namely, statistical study (structural estimation) and experimental and quasi-experimental design.¹⁶¹ It has been confirmed in recommendations on analysing the impact of State aid, where the European Commission describes those study designs as “able to identify the causal impact of the scheme itself, undistorted by other variables that may have had an effect on the observed outcome, e.g. general macroeconomic conditions.”¹⁶²

The logic of the first design consists in evaluation of effects by applying advanced statistical methods to determine the relation (dependencies) between the intervention and the expected or actual outcomes. Under this approach, the most frequent used methods for estimating the effects of public intervention include statistical analysis methods - mainly econometric modelling and regression analyses. From the beginning of 1990s, several macroeconomic models (HERMIN, QUEST, EcoMod, GIMF, MaMoR, EUImpact or RHOMOLO) which allow to evaluate the impact of the EU funds on the changes in GDP, investment, employment and unemployment or value added of production were developed,. There is a multitude of studies using macroeconomic models, in particular for *ex ante* evaluation of operational programmes co-financed from the EU funds under the Cohesion Policy and the Common Agricultural Policy.¹⁶³ It is worth mentioning that structural estimation is virtually the only approach used in *ex ante* evaluations of the impact of the EU funds on development of Polish regions.¹⁶⁴

¹⁶¹ Cf. E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, B. Befani, *Broadening the Range of Designs and Methods for Impact Evaluations*, Department of International Development – Working Paper 38, Washington DC 2012 and M. Bamberger, J. Vaessen, E. Raimondo, *Dealing With Complexity in Development Evaluation: A Practical Approach*, 1st Edition, SAGE Publications, 2015.

¹⁶² European Commission, “Common methodology for State aid evaluation”, Commission Staff Working Document, p. 7, http://ec.europa.eu/competition/state_aid/modernisation/state_aid_evaluation_methodology_en.pdf (access on 25.05.2017).

¹⁶³ At first, *ex ante* evaluations of the impact of the EU funds on the economies of state beneficiaries of the cohesion policy and the Common Agricultural Policy used the HERMIN model created by J. Bradley from the Economic and Social Research Institute and implemented for the first time in Ireland in 1989. Subsequent *ex ante* evaluations began to use “national” models, such as EcoRET in Hungary (H. Schalk and A. Vargi), or LATFUN in Lithuania (Baltic International Centre for Economic Policy Studies).

¹⁶⁴ For example: *ex ante* evaluations of all operational programmes implemented in Poland in the years 2007-2013 included estimations of the impact of the EU funds on the Polish economy calculated with the HERMIN model.

The modelling usually covers extensive supply and demand effects of the funds absorption by the economies at the national or regional level. It includes comparing the results that the economy will achieve in a scenario assuming the lack of external funds with the outcomes of the scenario with external funds.¹⁶⁵ The models are usually based on the concept of equilibrium which reflects the functioning of the economy as a whole (general equilibrium models) or of its specific sector (partial equilibrium models). In terms of timeliness of the modelled processes, the models can be static or dynamic. Since 1980s, the Dynamic Stochastic General Equilibrium (DSGE) models, used for forecasting and analysing public policies in numerous countries, have enjoyed huge popularity and been widely used by, *inter alia*, central banks, international organisations and the European Commission. The unique feature of the DSGE models is the comprehensive analysis of economic events, taking into account their dynamics and stochastic nature.¹⁶⁶

However, these methods have certain serious limitations due to several reasons. Firstly, each model is based on numerous theoretical assumptions regarding the functioning of the economy and its entities and the relations between them. The models thus reflect the “simplified” reality and the credibility of modelling results depends on reliability and comprehensiveness of assumptions constituting the foundation of each model. Secondly, the models always assume a positive change. The “tautological” nature of assumptions of the applied models does not allow to capture the negative effects of the EU funds, which may occur in the long-term, when demand effects are over.¹⁶⁷ Thirdly, models used for evaluating the Cohesion Policy funds usually assume that structural funds are fully absorbed and effectively disbursed, but the reality may be different.¹⁶⁸ Fourthly and finally, modelling is possible only if absorbed funds accounts for a significant percentage of total GDP of the country or the region, otherwise the impact of support funds will be too small and virtually “undetectable” in the modelling results. In addition, due to complexity of interregional links and “leaks” related to the

¹⁶⁵ Despite a relatively common use of such models to predict the consequences of public policy instruments implementation, the critics point out that they reduce the extremely complex economic reality to mathematic equations which are to reflect this reality. In 2010, the discussion on limitations in using the DSGE models was initiated by leading American economists working at the Committee on Science and Technology. It is worth quoting R. Solow, a member of the Committee, who stated that “[DSGE models] take it for granted that the whole economy can be thought of as if it were a single, consistent person or dynasty carrying out a rationally designed, long-term plan, occasionally disturbed by unexpected shocks but adapting to them in a rational, consistent way” (cf. R. Solow, “Building a Science of Economics for the Real World. Prepared Statement for the House Committee on Science and Technology” 2010, <https://www.gpo.gov/fdsys/pkg/CHRG-111hhr57604/pdf/CHRG-111hhr57604.pdf> (access on 25.05.2017)).

¹⁶⁶ For more information on the DSGE models and their use in analysing the effects of the Common Agricultural Policy see: A. Bezat, S. Figiel, J. Kufel, *Model dynamicznego, stochastycznego stanu równowagi ogólnej jako narzędzie wspierania formułowania założeń wspólnej polityki rolnej*, Institute of Agricultural and Food Economics – National Research Institute, Warsaw 2009.

¹⁶⁷ G. Gorzelak, “Fakty i mity rozwoju regionalnego”, *Studia Regionalne i Lokalne*, No 2(36)/2009, pp. 5-27.

¹⁶⁸ L. Polverari, J. Bachtler et al., *Balance of Competences Cohesion Review: Literature Review on Cohesion Policy*, Final report to the Department for Business, Innovation and Skills 2014.

flow of production factors, such methods are justified only for analyses at the national level.¹⁶⁹ Therefore, modelling does not address the need to determine the impact of the EU funds on local economies.

It is worth mentioning another method used for evaluating the impact of the funds on economic development, i.e. the analysis of relationships between the changes of indicators depicting the socio-economic development of NUTS 3, NUTS 4 or NUTS 5, and the volume of the EU funds absorbed.¹⁷⁰ Contrary to econometric modelling, such research do not allow for estimation (forecasting the impact of funds on selected indicators), but use historical data to determine the potential direction and strength of the relationship between the observed changes and the value of funds used. This is a serious limitation for the use of this method for ex post evaluations. In addition, the results of such analysis only confirm the fact that some phenomena exist, but do not explain causality (statistically significant relationship between the change of dependent and independent variable may be accidental or may be due to a third variable). For example: positive correlations between the obtained funds and the economic growth in NUTS 5 and NUTS 4 usually result from their general wealth – units with higher budgetary proceeds are able to obtain more funds (e.g. under the Cohesion Policy and the Common Agricultural Policy of the EU). The correlation analysis cannot thus be used for impact evaluation which as a rule is based on causality between intervention and results.¹⁷¹

Another research approach that may be used for impact evaluations includes experiments and quasi-experiments. These study designs allow to determine the net effects of specific intervention by means of identification and then comparison (within the same population) of the units that differ in only one aspect, i.e. being or not being subject to intervention. As a rule, this approach is characterised by significant internal validity, which means that the differences observed in the experimental and the control group are with a high probability the result of the implemented intervention.¹⁷²

¹⁶⁹ It should be noted, however, that despite this limitation the analyses using econometric modelling are applied also at the level of regions. An example may be the use of the HERMIN model to estimate the impact of the EU funds on the economies of 16 Polish voivodeships.

¹⁷⁰ It is about determining the correlation ratio between the changes in selected indicators and the volumes of the EU funds in absolute or per capita terms.

¹⁷¹ In Poland, the examples of the use of such method include the following studies: D. Wojtowicz, T. Kupiec, "Fundusze unijne szansą na wzrost gospodarczy? Studium przypadku województwa lubelskiego", *Studia Regionalne i Lokalne*, No 3(61)/2015, J. Misiąg, W. Misiąg M. Tomalak, *Ocena efektywności wykorzystania pomocy finansowej Unii Europejskiej jako instrumentu polityki spójności społeczno-gospodarczej oraz poprawy warunków życia*, University of Information Technology and Management, Rzeszów 2013, and G. Gorzelak, "Wykorzystanie środków Unii Europejskiej dla rozwoju kraju – wstępne analizy", *Studia Regionalne i Lokalne*, 3(57)/2014.

¹⁷² E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, B. Befani, *Broadening the Range of Designs and Methods for Impact Evaluations*, Department of International Development – Working Paper 38, Washington 2012.

The quasi-experimental approach is not common in evaluations of the EU funds' effectiveness in strengthening the economic growth at the national, regional and local level.^{173,174} Relatively few studies aimed at estimating the net effect of the absorption of the Cohesion Policy funds on the economies and based a quasi-experimental research design, include the analyses by Hagen and Mohl¹⁷⁵, Beckera et al.¹⁷⁶, Pellegrini et al.¹⁷⁷, Percoco and Gagliardi¹⁷⁸ and – as the only study focused at the local level – by Crescenzi and Giuy¹⁷⁹. Such studies most often use the following methods to divide the units into groups covered by intervention and control groups: PSM, Regression Discontinuity Design (RDD) or Local Average Treatment Effect (LATE).

However, the implementation of quasi-experimental research design using the PSM for evaluating the impact of the EU funds on local economies in Poland has some serious limitations. Firstly, the PSM method is traditionally based on the logistic regression model, which means that a sine qua non condition for its use is to identify a binary dependent variable. In practice, PSM may be used in research with precisely defined groups of units covered and not covered by intervention (in this case: using and not using the EU funds). However, we deal with a different situation: all the units concerned were subject to intervention, but to a varying extent. Logistic regression cannot thus be used for evaluating the impact of the Cohesion and Common Agricultural Policies,

¹⁷³ Counterfactual approach is widely used in evaluations of projects and programmes co-financed from the EU funds, where individuals or enterprises are units "divided" into experimental and control groups. Examples of studies using counterfactual methods in Poland include, *inter alia*, PARP's evaluations (e.g. *Ex post evaluation of Phare 2002 and 2003 SSG programmes, Studies on the net effects of the projects implemented under the Sectoral Operational Programme Improvement of the Competitiveness of Enterprises (SOPICE) 2004-2006 supporting the increase of the competitiveness of enterprises, Analyses of net effects of selected measures under the OPIE 2007-2013 directly addressed to enterprises, evaluations of the impact of co-financing of post-graduate studies and education in the age group 50+ under the Operational Programme Human Capital (OPHC) 2007-2013*) and regional evaluations (*Evaluation of support provided under projects implemented by non-governmental organisations under the regional component of the OPHC for Pomorskie Voivodeship in the context of the future financial perspective, Evaluation of the Regional Operational Programme Warmia and Mazury in the context of competitiveness of companies, products and services*).

¹⁷⁴ Cf. J. Pierkowski, P. Berkowitz, "Econometric assessments of Cohesion Policy growth effects: How to make them more relevant for policy makers?", *Regional Working Paper*, European Commission, Directorate-General for Regional and Urban Policy, 2015.

¹⁷⁵ T. Hagen, P. Mohl, "Does EU structural funds promote regional growth? New evidence from various panel data approach", *Regional Science and Urban Economics* 2010, 40(5), pp. 353-365.

¹⁷⁶ O. Becker, P. Egger, M. von Ehrlich, R. Fenge, "Going NUTS – The effect of EU structural funds on regional performance", *Journal of Public Economics* 2010, 94 (9-10), pp. 578-590.

¹⁷⁷ G. Pellegrini et al., "Measuring the Impact of the European Regional Policy on Economic Growth: a Regression Discontinuity Approach", *Papers in Regional Science* 2013, vol. 92, no. 1.

¹⁷⁸ M. Percoco, L. Gagliardi, "The Impact of the European Cohesion Policy in Urban and Rural Regions: a Discontinuity Approach", *Bocconi Working Paper* 2014, <https://www.unibocconi.it/wps/wcm/connect/5651416a-469d-4d80-af74-a0eae9d8d4d/Paper+3.pdf?MOD=AJPERES&useDefaultText=0&useDefaultDesc%20=0> (access on 25.05.2017)

¹⁷⁹ R. Crescenzi, M. Giua, "Spatial discontinuity for the impact assessment of the EU Regional Policy. How does the net impact of the Policy differ across countries?", Paper presented at the Second EU Cohesion Policy Conference, Riga, 5-6 February, 2015.

in case of Polish NUTS4 and NUTS 5 as all of them implemented projects co-financed from the EU funds, although the scope and value of those projects varied.

SPSM model – theoretical assumptions

The evaluation practice in recent years shows that the net effect of intervention is usually determined using methods derived from PSM. PSM was proposed as a method in 1983 by Rosenbaum and Rubin.¹⁸⁰ Since then it has won a large number of supporters and critics. In recently years, PSM has become increasingly popular, primarily due to recommendation of the European Commission to use counterfactual methods for evaluating the impact of the EU cohesion policy funds.¹⁸¹ However, as stated in the introduction, the PSM method has not yet been used in the majority of evaluations of territorial and regional development, in particular of the impact of the EU Cohesion Policy. The lack of a binary dependent variable (support received or not received) could in some cases be circumvented by using multinomial logistic regression analysis, but in this case we would have to deal with a discrete and measurable support measures with a small number of levels (support for regions is usually a continuous variable). Therefore, more advanced versions of logistic regression can be used only occasionally. What should be done, if there is no dummy variable that can be used as a dependent variable?

The proposed solution is Stratified Propensity Score Matching (SPSM). Under this model, the treatment variable may be a continuous variable. Moreover, the treatment variable is not a dependent variable, since the SPSM model does not need a dependent variable to calculate propensity scores. Therefore, the treatment variable does not answer the question “was my unit subject to intervention or no?”, but it answers the question “to what extent has my unit been subject to intervention?”. Counterfactual situation is defined differently - it does not consist in intervention or the lack of intervention in a given unit, but in the extent of intervention. In this case, the traditional counterfactual definition becomes a special case of the counterfactual situation defined above.

In the SPSM model, propensity scores are calculated in a completely different way than in the PSM model. The statistical method used in this model is fuzzy hierarchical clustering. The clustering consists in selecting the groups of units that are as similar as possible. Contrary to simple clustering where each unit has a single cluster value

¹⁸⁰ D. B. Rubin, N. Thomas, “Combining propensity score matching with additional adjustments for prognostic covariates”, *Journal of the American Statistical Association* 2000; G. W. Imbens, “Nonparametric estimation of average treatment effects under exogeneity: A review”, *The Review of Economics and Statistics* 2004; J.L. Schafer, J. Kang, “Average causal effects from nonrandomized studies: A practical guide and simulated example”, *Psychological Methods* 2008; S. Setoguchi, S. Schneeweiss, M. A. Brookhart, R. J. Glynn, E. F. Cook, “Evaluating uses of data mining techniques in propensity score estimation: A simulation study”, *Pharmacoepidemiology and Drug Safety* 2008.

¹⁸¹ https://crie.jrc.ec.europa.eu/sites/default/files/documents/CIE_Guidance_EN.pdf (access on 28 March 2017).

assigned, in fuzzy hierarchical clustering a unit is assigned not only with a discrete cluster value (i.e. being a member of cluster 1. or 2.), but also the probability of being a member of the cluster. In this model, propensity values within a given segment are treated as propensity measures. Further procedure is similar to the classic PSM model, namely, units belonging to a cluster are paired based on probability of their being assigned to a given cluster. Examples of the practical use of the method are described further in the chapter and the technical description of statistical procedures is presented in the next subchapter.¹⁸²

The method is called stratified PSM, since it results in obtaining several clusters which may be treated as stratification layers. A “side effect” of this method is the possibility to perform the analysis not only on the entire paired set of units subject to intervention, but also on its subsets. The hierarchical methods allows to select i.a. small, but highly specialised clusters for which net effects may be different than for the entire population of units subject to intervention. One of disadvantages of PSM is that it is able to generate net effects only for the entire population. In order to analyse a lower level (e.g. evaluate net effects in other divisions or subgroups than the entire analysed sample), it is necessary to repeat the PSM procedure. Contrary to PSM, SPSM preliminarily divides the sample (or population) into clusters, thus enabling an analysis at various levels. Practical implications of this approach are presented in the examples provided further in the chapter.

SPSM – technical description

The proposed method is based on Ward’s hierarchical clustering algorithm¹⁸³, which is a agglomerative method. This means that each observation is analysed as a single cluster and in subsequent iterations the closest clusters are merged into one. The algorithm completes its operation after obtaining the number of clusters earlier assumed by the researcher. The estimation of the optimal number of clusters under hierarchical methods goes beyond the scope of this article, however an abundant literature on the subject is available.¹⁸⁴

¹⁸² It is also worth noting that the technical description in the next subchapter is presented only to ensure replicability of the method for other researchers. We do not want the SPSM method to be perceived as a “black box” method where propensity measures are generated in a mysterious and unclear way. The readers who are not interested in statistical complexities of the method may skip the technical part and go to examples of its use.

¹⁸³ J. H. Ward, “Hierarchical Grouping to Optimize an Objective Function”, *American Statistical Association Journal* 1963.

¹⁸⁴ D.G. Garson, *Cluster Analysis*, Statistical Associates Publishers 2014; C. C. Aggarwal, C. K. Reddy, *Data Clustering: Algorithms and Applications*, Chapman and Hall/CRC 2013; B. S. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster Analysis*, Wiley 2011.

Ward's algorithm analyses the error sum of squares at each step in the following way:

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

where x_i is the score of the i th unit and n is the sample size. The algorithm is repeated until a sufficient number of clusters (or "sets" according to Ward's nomenclature) is obtained.

Each iteration is defined as:

$$S(p_{n-1}, n - 1) = [S(p_{n-1}, n)] \cup [S(q_{n-1}, n)]$$

where p_{n-1} is the smaller of the two clusters (subsets), and q_{n-1} is the larger of the two clusters (sets).

Hierarchical clustering procedures (similarly to other clustering procedures, such as k-means or k-medians clustering) calculate only the classification into a cluster as a default. In the course of calculating subsequent steps, a distance table is also created - most frequently in the form of a dissimilarity matrix or a similarity matrix - which may be used to calculate the probabilities of allocation to individual clusters. In result of distance table transformations we obtain the values of fuzzy clustering. Given the information on cluster allocation and on the distances from the distance table, one may calculate the Fuzzy Hierarchical Clustering Scores (FHCS), which in the subsequent step shall be used as PSM scores¹⁸⁵. *FHCS is defined as follows:*

$$FHCS_i = |DS_i - \bar{x}(DS)|$$

where DS stands for Dissimilarity Score).

The procedure of calculating fuzzy clustering scores should be performed for each cluster separately. As stated above, FHCS in further procedure are treated as PSM scores and thus further procedures accounts for the differences between the values of each two units which are closest in terms of FHCS. Therefore, the maximum number of pairs given for each cluster is $n-1$, assuming that the number of units in a given cluster = n .

Method of assessing net effect

In a situation where SPMS sores have been calculated, one may proceed to estimate net effect. In the examples presented below, differences in the amount of support as well as differences for individual success indicators have been calculated (for the first study, there was just one success indicator, for the second study, the number of indi-

¹⁸⁵ Understood as measure of similarity between units, interpreted in the same manner as in the case of classical PSM method.

cators was more than 10)¹⁸⁶. Since SPSM scores demonstrate proximity between the units within a given cluster, it is necessary to calculate first the values of effects within clusters, and only subsequently net effects for the entire sample. This method does not allow for calculating net effects based on two or more success indicators at the same time. However, it allows for calculating total net effect, which is estimated based on the sum of differences in the value of support and the levels of success indicators, however not in clusters but for the entire population, paired using PSM scores.

Summing up the technical part, it should be noted that:

- the SPSM method works well in impact evaluation of policies that cover a majority of or all units in a given population, and it works well in particular in the evaluation of effects of the EU Cohesion Policy funds at the level of gymnasias or NUTS 4;
- SPSM allows for calculating net effects of interventions not only at the level of entire population, but also in individual sub-groups (e.g. net effects in urban NUTS 4 or net effects in territorially large NUTS 4 of predominantly agricultural character);
- as opposed to the PSM method, the operation of which is limited to situations where there are two groups of units (supported and not supported), SPSM operates in a situation where all or almost all units have received support;
- in order to allow for discussing a counterfactual situation in this case its definition has been extended. In the classical sense, a counterfactual situation is an answer to the following question: "what would have happened if the discussed units had not received any support at all?"; in SPSM, the following question is answered: "what would have happened if the discussed units had received a different amount of support?". However, this is still a counterfactual situation.

Examples of using SPSM in evaluation studies¹⁸⁷

Evaluation of Axis 3 and 4 of Rural Development Programme 2007-2013

The first example of SPSM method application is the impact assessment of the implementation of Rural Development Programme 2007-2013 (RDP 2007-2013) on changes in the functioning of economy, taking place in rural areas¹⁸⁸. The analyses were preceded by a classification of NUTS 5, based on their selected social, economic and spatial characteristics. Creating reasonably homogeneous groups of NUTS 5 was necessary in order to compare changes in own revenue per capita (input variable) in twin NUTS 5 that differed in terms of outlays under RDP 2007-2013, and thus to specify

¹⁸⁶ In each evaluation study, success indicators should be defined in advance. Success indicators are independent from the described method. Examples demonstrate that one may expect e.g. both, net effects for own incomes of gaminess as well as net effects of the influx of foreign capital in NUTS 4.

¹⁸⁷ Evaluations have been presented in which the authors of this chapter belonged to the research team.

¹⁸⁸ "Impact Assessment of RDP 2007-2013 on the quality of life in rural areas, taking into account the LEADER approach", EGO - Evaluation for Government Organizations s.c., report commissioned by the Ministry of Agriculture and Rural Development, Warsaw, November 2016.

the net inflow of such funds. This also gave grounds for answering a question on the impact of NUTS 5 characteristics on the achieved net effects of using the RDP 2007-2013 funds. The classification of NUTS 5 was based on selected social, economic and spatial characteristics, accounting for selected data from 2006 (year preceding intervention) to allow for distinguishing NUTS 5 which were very similar before the calls for applications for co-financing under the Programme began.

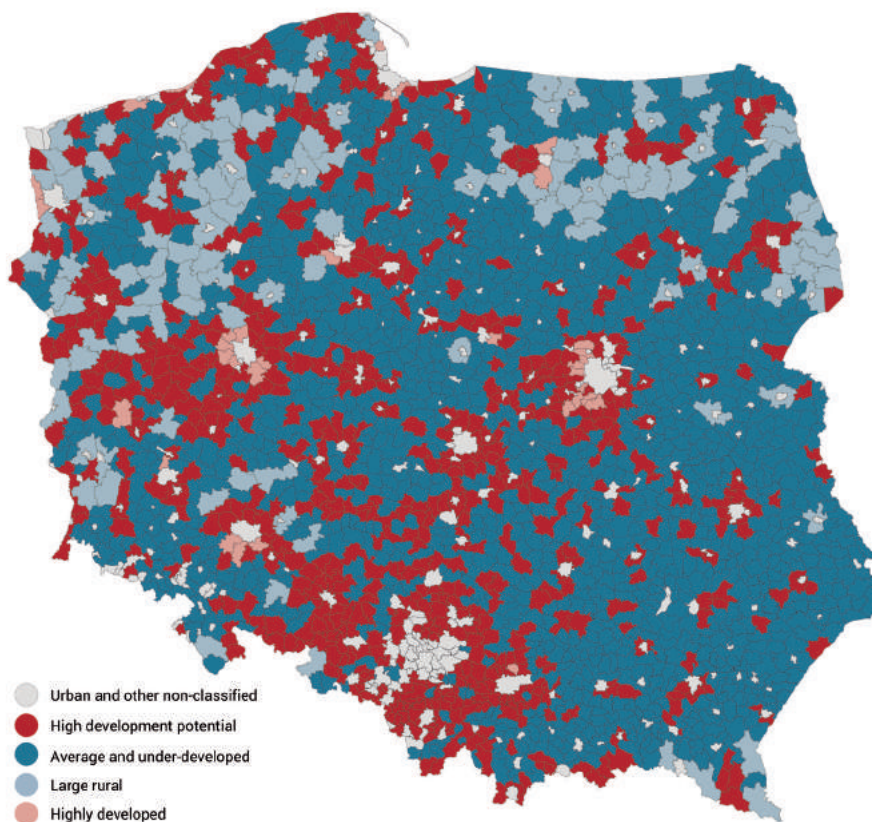
The classification of gamins was carried out in a number of stages. First, synthetic indicators were developed, used for size reduction. In result of this activity, more than ten indicators were obtained, related to the number of NUTS 5 population. In the second step, the obtained indicators were standardised to allow for creating a cluster of equal strength. The classic Z standardization was used (Z-scores)¹⁸⁹.

A subsequent stage of the analysis was an attempt to identify NUTS 5 that are similar in terms of the observed characteristics. Method based on hierarchical cluster analysis was used to this end¹⁹⁰. The proximity matrix created in the course of cluster analysis provided information on the distance of individual NUTS 5 from the center of a given cluster. Using the fuzzy clustering method, each NUTS 5 was assigned a probability of belonging to each of the four clusters that have emerged: (1) NUTS 5 high development potential, (2) average and under-developed NUTS 5, (3) large rural NUTS 5, (4) highly developed NUTS 5. In the last step, discriminant analysis was applied to obtain the final probabilities of belonging to a given cluster. The latter were treated in the analysis as PSM scores - measures of the proximity between NUTS 5 - and they allowed for selecting pairs of the so called twins. NUTS 5 were defined as "twins" by their similarities before the programme was launched, namely in 2006, regarding such characteristics as: area, income, population, unemployment, etc.

¹⁸⁹ The standardization of variables is to ensure their comparability and mitigate the effects of using different scales.

¹⁹⁰ More information about the method can be found in: D. Wojtowicz, Ł. Widła-Domaradzki, "Estimating the impact of the EU Cohesion Policy on local level", in: N.F. Dotti, (ed.), *Learning from implementation and evaluation of the EU Cohesion Policy: Lessons from a research-policy dialogue*, RSA Research Network on Cohesion Policy, Brussels 2016, <http://www.regionalstudies.org/uploads/documents/Lessons4CP.pdf> (access on 25.05.2017).

Figure 1. Spatial distribution of NUTS 5 according to adopted classification (2006, situation before the implementation of Axis 3 and 4 of RDP was launched)



Source: Impact Assessment of RDP 2007-2013 on the quality of life in rural areas

In modelling the net impact of projects co-financed from the EU funds, the authors used the value of funds financing projects implemented under RDP as well as programmes implemented under the EU Cohesion Policy. The assumption was made that the higher the amounts absorbed by a given NUTS 5, the bigger their observed net effect will be¹⁹¹. For each type of intervention (RDP and other programmes financed under the EU Cohesion Policy) an "ideal model" was created, in which all the absorbed funds were accountable for the total income increase. The "ideal model" was treated

¹⁹¹ The assumption had to be made, since only one variable was used in the study. In order to explain the impact of two different forms of project financing (RDP and the EU Cohesion Policy funds) without making such assumptions, would result in a situation where an intervention involving lower financial outlays will always have a higher net effect (because net effect is described as a relation between the level of expenditure and the change in NUTS 5 own income per capita).

as a point of reference and was tested against the actual situation. In the subsequent step, for each counterfactual pair, the amount of support and changes in own income of the NUTS 5 per capita were subtracted from each other. The sum of differences for all the counterfactual pairs gave grounds for estimating the net effect of funds, as presented in table 1. The net impact demonstrates to what extent a single euro absorbed by a given NUTS 5 contributed to the increase in local government's own income per capita. For example, the value of 0.049 euro indicates that each single euro that was used contributed to almost 5 eurocents increase in local government's own income.

Table 1. Net effect of funds absorbed by NUTS 5 under RDP and programmes implemented under the EU Cohesion Policy (in Euro)

	EU Cohesion Policy funds	RDP and the EU Cohesion Policy funds	RDP
All NUTS 5	0.049	0.083	0.085
NUTS 5 with high development potential	0.049	0.085	0.095
Average and under-developed NUTS 5	0.049	0.080	0.080
Large rural NUTS 5	0.056	0.094	0.081
Highly developed NUTS 5	0.054	0.096	0.090

Source: own elaboration.

Results of modelling have shown that in the case of rural NUTS 5, higher net effects in the form of increase in own income per capita were generated by projects co-financed under RDP rather than from the EU Cohesion Policy funds. Additionally, the highest net effects were observed in NUTS 5 classified as units with high development potential and highly developed. Simultaneously, the most positive impact in terms of increase in local government's own income was observed for investments financed under RDP and from Cohesion Policy funds in highly developed and large units.

Evaluation of Axis 1 of Operational Programme Development of Eastern Poland (2007-2013)

The "Evaluation of impact of Operational Programme Development of Eastern Poland 2007-2013 (OPDEP) on the development of entrepreneurship in Eastern Poland" is another example of a study in which SPSM was successfully introduced.¹⁹²

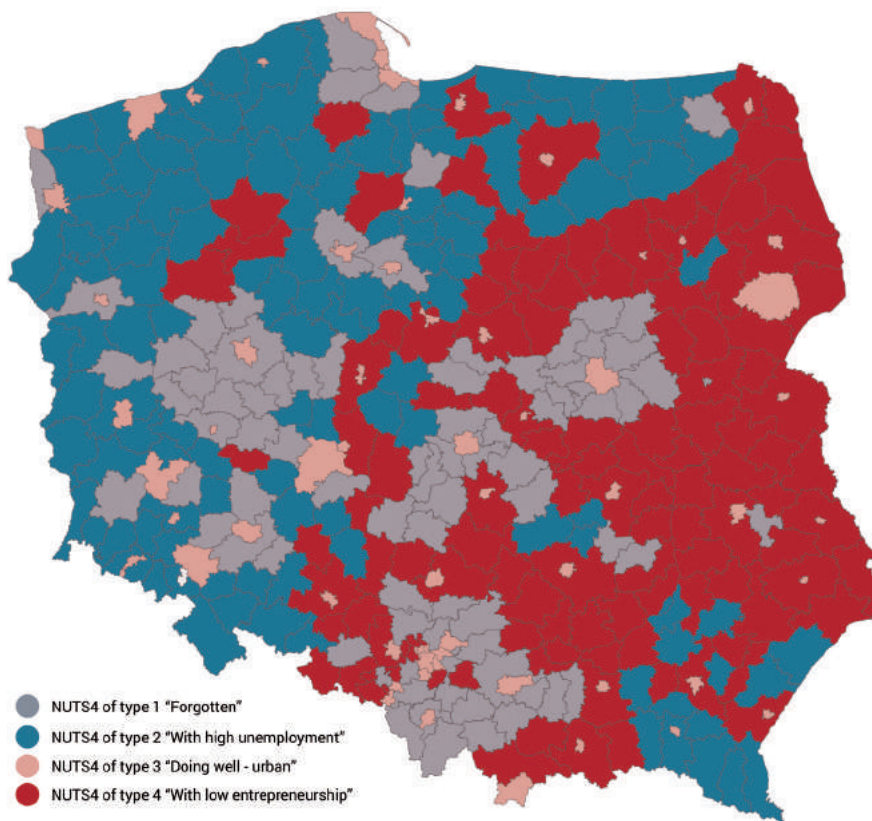
In the above study, NUTS 4 were analysed, not NUTS 5 as in the RDP study. The difference resulted from the aim of the study: changes taking place in the development of rural areas should be analysed at the highest possible disaggregation level whereas studies dedicated to entrepreneurship development should take into account at least the NUTS 4 level due to the possibility of occurrence of the so-called spill-over effects (a significant change in entrepreneurship indicators in one NUTS 5 will affect the situation on the adjacent NUTS 5 - treating them as separate, not connected entities will negatively affect the quality of statistical analysis). Also the availability of certain indicators at the NUTS 4 but not at the NUTS 5 level was important.

Like in the first of studies described, also in the case of OPDEP, the analyzed population was divided into strata. The map below presents the distribution of NUTS 4 types in Poland; however only the NUTS 4 from the macro region of Eastern Poland were used for SPSM modelling. The typology of pewits was made based on the BDL indicators of central Statistical Office (GUS) for 2006. Cluster variables were grouped into three blocks: entrepreneurship, infrastructure, complementary variables¹⁹³. Four clusters of NUTS 4 were distinguished: NUTS 4 that are doing well; NUTS 4 with low level of entrepreneurship; NUTS 4 with high unemployment; and NUTS 4 that may be called „forgotten”: units adjacent to big towns and agglomerations, characterised by area a little smaller than average and by population density a little lower than the average NUTS 4 in Poland (see Fig. 2).

¹⁹² "Evaluation of the impact of Operational Programme Development of Eastern 2007-2013 (OPDEP) on the development of entrepreneurship in Eastern Poland" prepared by consortium Agrotec Polska sp. z o.o., WYG PSDB sp. z o.o and EGO – Evaluation for Government Organizations s.c., report commissioned by PARP, Warsaw, December 2016.

¹⁹³ Detailed methodological assumptions (selection of variables used to obtain probability scores, detailed description of NUTS 4, etc) can be found in the methodological annex to the final report from the study.

Figure 2. Spatial distribution of NUTS 4 by clusters (2006, situation before the implementation of OPDEP 2007-2013 was launched)



Source: Evaluation of the impact of Operational Programme Development of Eastern 2007-2013 (OPDEP) on the development of entrepreneurship in Eastern Poland

Like in the first of studies described, also in this modelling the values of funds financing projects implemented under OPDEP were used as well as under other programmes promoting entrepreneurship development, implemented in Eastern Poland under the EU Cohesion Policy. Among the funds that had the same aim as the funds allocated under Axis 1 of OPDEP¹⁹⁴ were the following programmes: Innovative Economy Operational Programme, Infrastructure and Environment Operational Programme and 16 Regional Operational Programmes. In contrast to the RDP evaluation study, in OPDEP evaluation 16 success indicators were distinguished, not just one. In this evaluation study, success indicators were understood as different variables that may be under-

¹⁹⁴ The quoted evaluation studies focused on Axis 1 of OPDEP.

stood as indicators of improvement in economic competitiveness (e.g. the number of employed, the number of new companies, foreign capital, etc.). Therefore, the counterfactual modelling using the SPSM model had to be performed 16 times. As an example, the results of modelling for one success indicator is presented below. This is the difference between the number of newly established companies and the number of companies struck off the REGON register. A positive balance means that the number of newly established companies was higher than the number of companies struck off the register; a negative balance means that in given period more companies were struck off than entered in the register. All values were relativized to the expenditure of one million of PLN within the framework of OPDEP and other funds allocated for development of entrepreneurship. Indicator presented in the table also shows the average annual change. It means that every million of PLN spent under OPDEP resulted in average annual increase of 0.13 in the number of companies in the macroregion.

Table 2. Net effect of funds absorbed by NUTS 5 under OPDEP and programmes implemented within the framework of EU Cohesion Policy (in euro)

	Net effect for newly established / struck off
All NUTS 4	0.13
NUTS 4 of type 1 ("forgotten")	-0.52
NUTS 4 of type 2 (with high unemployment)	-0.02
NUTS 4 of type 3 (doing well - urban)	-0.08
NUTS 4 of type 4 (with low entrepreneurship)	0.33

Source: own elaboration.

As can be seen, net effect is positive. The above means that as regards newly established companies, funds disbursed under Axis 1 of OPDEP had better effect than funds allocated for the development of entrepreneurship. However, the analysis within clusters shows a relatively high variation between NUTS 4. Insignificant, nevertheless negative net effects for NUTS 4 with high unemployment and urban NUTS 4 doing well, imply that other funds (other than under OPDEP) had a more positive impact on the discussed balance. An extremely different situation is observed for NUTS 4 of other two types. In the case of "forgotten" NUTS 4 (including suburban NUTS 4, adjacent to agglomerations), the balance of OPDEP implementation is negative: every million of PLN spent under OPDEP functioned less effectively in these NUTS 4 than a million

spent from different funds. In this case, the effect equal to -0.52 means that in NUTS 4 under consideration, for every two million of PLN spent, more less one more company was struck off the register than newly established. The situation is very different in NUTS 4 with low entrepreneurship - here the OPDEP funds functioned more effectively than other funds dedicated for entrepreneurship development. Each three million of PLN spent under OPDEP resulted in the establishment of one more company than struck off the REGON register.

Conclusion

The method presented in this chapter responds to the challenges related to the limitations in the use of impact evaluation in estimating the net effects of interventions implemented at the local level. As has been shown, SPSM models have two major advantages, compared to the classic PSM method. Firstly, these models do not require a binary dependable variable to be defined in order to determine the propensity scores for the units under consideration. What's more, in practice it is not necessary to identify the dependable variable, since in SPSM, another clustering algorithm is used for calculating propensity scores. Secondly, SPSM offers a possibility to assess net effect on various subsets of units covered by a given intervention, and not on the entire population. Evaluation studies where SPSM was used, show that results obtained may constitute the key knowledge for decision makers - they provide information on the net effect generated by the projects, depending on the level of social and economic development in the units where they are implemented and on the kind of intervention (e.g. actions, priority axis, programmes) or funds that finance them (e.g., European Rural Development Fund, European Rural development Fund, local governments' own resources).

Summing up, it should be pointed out that SPSM models are either models that allow for assessment of relative effects, nor macroeconomic models. In contrast to relative effects analysis, in SPSM model, the effects of a single intervention may be analysed. It is not a macroeconomic model, since it does not explain whether the obtained effects are high or low - using macroeconometric models would be much more appropriate for this purpose.

A SPSM model used correctly should obviously be related to the nature of interventions. The classical counterfactual methods should continue to be used and the SPSM model should not be perceived as competitive, but complementary rather. In the authors' opinion, the SPSM model presented should be used in a situation when support is of continuous nature, that is that when it is impossible to find units which did not receive any financial support. The examples presented in the text concerned territorial evaluation studies, however this method may be implemented more broadly - it may be used for participants of training who receive different kinds of support, or unemployed who have been subject to any continuous intervention (e.g., coupons of different values).

About the authors

Dominika Wojtowicz

PhD in Management Science. She graduated from the Koźmiński University and International Relations at the University of Warsaw. She is a scholar of the Faculty of Political Science at the *Università di Siena* (Italy) and the Faculty of Management of the *Universidad Politecnica de Cartagena* (Spain). Her research interests focus on the broadly understood effectiveness of public intervention geared towards strengthening the regional and local development. She participated in many research projects focused on specific areas of public policies. She is an author of expert opinions in the field of impact of the EU funds on regional economies in Poland, on behalf of, *inter alia*, the European Parliament and the European Commission. She was a coordinator of studies on modern evaluation methods for specific areas of intervention and public regulation, studies on the use of evaluation in managing regional operational programmes in Poland and Spain, as well as the use of evaluation practise in the the Regulatory Impact Assessment. She is Assistant Professor at the Chair of Economy of the Koźmiński University. She gives lectures on the methodology of quantitative and qualitative research used in policy analyses during postgraduate studies for public servants (Evaluation Academy, Analyst Academy, Management in Public Administration).

Łukasz Widła–Domaradzki

Statistician, senior analyst and data scientist with nearly 20 years of experience. He worked for international research companies (MillwardBrown, Ipsos), European evaluation agencies (ICF International) as well as in the public sector (Polish Agency for Enterprise Development, National Information Processing Institute). He works in different software environments, such as SPSS, Stata, Statistica or SAS, specializing in structural modeling (in AMOS or SmartPLS) and counterfactual analysis (PSM, SPSM). He participated in various evaluation and scientific (medical, sociological, psychological and economic) research projects, in which he acted as a statistical analysis expert (e.g. in “MUS - Ministries as Learning Organizations” coordinated by the Chancellery of the Prime Minister). Łukasz is also involved in teaching (including classes for students at the Koźmiński University, postgraduate workshops on quantitative methods in evaluation at Evaluation Academy) and training (e.g. on the use of SPSS package for the Ministry of National Defence). He is a certified Data Analyst for EU Election Observation and Democratic Support and an author or co-author of several books and articles concerning statistical methods in evaluation.

Bibliography:

1. Aggarwal C.C., Reddy C.K., *Data Clustering: Algorithms and Applications*, Chapman and Hall/CRC 2013.
2. Becker O., Egger P., von Ehrlich M., Fenge R., "Going NUTS – The effect of EU structural funds on regional performance", *Journal of Public Economics* 2010, 94 (9–10), pp. 578-590.
3. Crescenzi R., Giua M., *Spatial discontinuity for the impact assessment of the EU Regional Policy. How does the net impact of the Policy differ across countries?*, Paper presented at the Second EU Cohesion Policy Conference, Riga, 5-6 February 2015.
4. Everitt B.S., Landau S., Leese M., Stahl D., *Cluster Analysis*, Wiley 2011.
5. Garson D.G., *Cluster Analysis*, Statistical Associates Publishers 2014.
6. Gorzelak G., "Fakty i mity rozwoju regionalnego", *Studia Regionalne i Lokalne* 2009, No 2 (36), pp. 5-27.
7. Hagen T., Mohl P., "Do EU structural funds promote regional growth? New evidence from various panel data approach", *Regional Science and Urban Economics* 2010, 40 (5), pp. 353-365.
8. Imbens G. W., "Nonparametric estimation of average treatment effects under exogeneity: A review", *The Review of Economics and Statistics*, 2004.
9. Misiąg J., Misiąg W., Tomalak M., *Ocena efektywności wykorzystania pomocy finansowej Unii Europejskiej jako instrumentu polityki spójności społeczno-gospodarczej oraz poprawy warunków życia*, University of Information Technology and Management, Rzeszów 2013.
10. Pellegrini G. et al., "Measuring the Impact of the European Regional Policy on Economic Growth: a Regression Discontinuity Approach", *Papers in Regional Science* 2013, vol. 92, no. 1.
11. Percoco M., Gagliardi L., "The Impact of the European Cohesion Policy in Urban and Rural Regions: a Discontinuity Approach", *Bocconi working paper* 2014, <http://www.unibocconi.it/wps/wcm/connect/5651416a-469d-4d80-af74-a0eaed9d8d4d/Paper+3.pdf?MOD=AJPERES&useDefaultText=0&useDefaultDesc=0> (access on 25.05.2017).
12. Polverari L., J. Bachtler et al., *Balance of Competence Cohesion Review: Literature Review on Cohesion Policy*, Final report to the Department for Business, innovation and Skills, 2014.
13. Rosenbaum P. R., Rubin D. B., "The central role of the propensity score in observational studies for causal effects", *Biometrika* 1983.
14. Rubin D. B., Thomas N., "Combining propensity score matching with additional adjustments for prognostic covariates", *Journal of the American Statistical Association*, 2000.
15. Schafer J. L., Kang J., "Average causal effects from nonrandomized studies: A practical guide and simulated example", *Psychological Methods* 2008.

16. Setoguchi S., Schneeweiss S., Brookhart M. A., Glynn R. J., Cook E. F., "Evaluating uses of data mining techniques in propensity score estimation: A simulation study", *Pharmacoepidemiology and Drug Safety* 2008.
17. Stern E., Stame N., Mayne J., Forss K., Davies R., Befani B., *Broadening the Range of Designs and Methods for Impact Evaluations*, Department of International Development – Working Paper 38, Washington DC 2012.
18. Trzeciński R., "Impact of direct subsidies on SMEs", in: Pokorski J. (ed.), *Towards Innovative Economy Effects of Grants to Enterprises in Poland*, Polish Agency for Enterprise Development, Warsaw 2011.
19. Ward J. H., "Hierarchical Grouping to Optimize an Objective Function", *American Statistical Association Journal* 1963.
20. Wojtowicz D., Kupiec T., "Fundusze unijne szansą na wzrost gospodarczy? Studium przypadku województwa lubelskiego", *Studia Regionalne i Lokalne* 2015, 3(61)/, pp. 15-23.

Other:

1. *Ocena wsparcia udzielonego w projektach realizowanych przez organizacje pozarządowe w ramach komponentu regionalnego POKL dla województwa pomorskiego w kontekście przyszłej perspektywy finansowej. Raport końcowy*, consortium of EGO – Evaluation for Government Organizations s.c. and IDEA for Development Foundation, report commissioned by the Pomorskie Voivodeship Marshal's Office, Warsaw 2013.
2. *Ewaluacja RPO WiM w kontekście konkurencyjności firm, produktów i usług. Raport końcowy*, EGO – Evaluation for Government Organizations s.c., report commissioned by the Warmińsko-Mazurskie Voivodeship Marshal's Office, Warsaw 2012.
3. *Ewaluacja ex post programów Phare 2002 i 2003 SSG, regionalny i sektorowy komponent wsparcia MŚP. Raporty końcowe*, Pentor Research, report commissioned by Polish Agency for Enterprise Development, Warsaw 2006-2007.
4. *Ewaluacje ex post programów Phare 2002 i 2003 SSG, regionalny komponent RZL. Raporty końcowe*, PBS DGA, report commissioned by Polish Agency for Enterprise Development, Sopot 2006-2007.
5. *Badanie efektów netto projektów wspierających wzrost konkurencyjności przedsiębiorstw. Raport końcowy*, PSDB sp. z o.o., report commissioned by Polish Agency for Enterprise Development, Warsaw 2009.
6. *Barometr Innowacyjności – analiza efektów netto wybranych Działań PO IG skierowanych do przedsiębiorstw, z wykorzystaniem podejścia counterfactual impact evaluation. Raport z ewaluacji wewnętrznej*, Polish Agency for Enterprise Development, Warsaw 2013.
7. *Efekty (netto) dofinansowanych studiów podyplomowych. Raport z ewaluacji wewnętrznej*, Polish Agency for Enterprise Development, Warsaw 2014.

8. *Efekty dofinansowanego kształcenia w grupie wiekowej 50 plus. Raport końcowy*, Institute for Socio-Economic Enquiry I-SEE, report commissioned by Polish Agency for Enterprise Development, Warsaw 2014.
9. *Analizy efektów netto wybranych Działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013 z wykorzystaniem podejścia counterfactual impact evaluation w ramach projektu ewaluacyjnego PARP „Barometr Innowacyjności”*, consortium of the Jagiellonian University and EGO – Evaluation for Government Organizations s.c., report commissioned by Polish Agency for Enterprise Development, Warsaw 2015.
10. *Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 na rozwój przedsiębiorczości w Polsce Wschodniej. Raport końcowy*, consortium of Evalu, WYG PSDB and EGO – Evaluation for Government Organizations, report commissioned by Polish Agency for Enterprise Development, Warsaw 2016.
11. *Ocena wpływu PROW 2007-2013 na jakość życia na obszarach wiejskich z uwzględnieniem podejścia LEADER. Raport końcowy*, EGO – Evaluation for Government Organizations s.c., report commissioned by the Ministry of Agriculture and Rural Development, Warsaw 2016.

Theory-based evaluation in studies of the innovative project selection system in Operational Programmes for 2014-2020

Introduction

This chapter presents the lessons learned from the implementation of several studies on evaluation of systems for the selection and evaluation of innovative projects in Operational Programmes implemented in Poland under the European Union Cohesion Policy for 2014-2020. Analysis of these project selection and evaluation systems was based on the theory-based evaluation approach, to ensure as precise a reflection of the support measure intervention logics as possible. This has been necessary in order to formulate argumentative recommendations how to refine the selection systems in terms of their improved adjustment to the specific features of the support instruments¹⁹⁵. Experience acquired in this respect is the key axis of considerations presented herein. It is supplemented by lessons learned from the modelling of the project selection systems in the previous financial perspective (2007-2013). In the studies referred to in this chapter they were the starting point for the improvement of solutions applied under the 2014-2020 perspective programmes. Furthermore, also based on the experience derived from these studies, the theory-based evaluation was assessed as a form of support for the programming process. Finally, several key recommendations were formulated in result of evaluation, with respect to organisation of project selection and evaluation processes under the 2014-2020 perspective Operational Programmes.

¹⁹⁵ In 2015-2016, a large team of researchers, including also the authors of this chapter, carried out two studies on the project selection and evaluation systems in national Operational Programmes for the years 2014-2020 (it was the Smart Growth Operational Programme and the Digital Poland Operational Programme – cf. “Evaluation of the project selection system at SGOP - stages I and II”, the IMAPP Consortium, IDEA of Development Foundation, PAG Uniconsult and Jagiellonian University by order from the Ministry of Development, Warsaw 2016 and 2017, “Evaluation of the project selection system DPOP”, the IMAPP Consortium, IDEA of Development Foundation, PAG Uniconsult and Jagiellonian University by order from the Ministry of Development, Warsaw 2017) and a study under the Eastern Poland macro-regional Operational Programme for 2014-2020 (in this case evaluation of the project selection system included all the intervention instruments – measures and submeasures – under priority axis I of the EPOP– cf. “Evaluation of the project selection system under the Eastern Poland Operational Programme 2014-2020”, The PAG Uniconsult and IMAPP Consortium by order from the Ministry of Development, Warsaw 2016). In each of these studies the theory-based approach was used as a starting point for in-depth evaluation of individual intervention instruments, pertinence, efficiency and effectiveness of the existing system of project selection (with particular focus on innovative projects).

Importance of the intervention theory in the project selection system

Every public intervention scheme should trigger off a specific change within the scope of the subject of intervention, reflected in its key objective and specific goals. Intervention may be perceived as a certain theory of change, imaging the way of solving a given problem according to the assumed logics of proceeding. This approach is particularly useful in programming public aid policies, as it allows seizing the relationships between the goals and the results, through reconstruction of the intervention logics, accounting for different assumptions (also those which are not expressly exposed). It makes it possible to determine why and what mechanisms lead (or not) to the anticipated results in a given context of their implementation¹⁹⁶. In this approach the issue whether the results of intervention were achieved is of less importance than the mechanisms and the cause and effect relationships which created their basis. The experience we acquired shows that the knowledge of these mechanisms and relationships allows refinement of public aid schemes, with particular focus on improvement in construction and modification of solutions for project selection systems which are co-funded and implemented to achieve the assumed goals of intervention. Evaluation-based changes that are introduced may include both legal regulations modelling the selection system as well as substantial and technical issues related to the construction of project evaluation criteria.

The given logics of public intervention is a result of diagnosis of a specific problematic situation and the subsequent challenges (for instance, poor innovative potential of the economy and its consequences). Intervention logics is modelled by predetermined, measurable intervention goals (e.g. growth in innovative potential and competitiveness of the economy), as well as by specified measures envisaged to achieve these goals (e.g. intervention instruments, focused on supporting measures undertaken by the addressees of support) and by the adopted assumptions, concerning conditions required for the expected level of objectives to be achieved. In other words, intervention logics is reflected in the following cause and effect pattern: (1) the implementation of specific measures making up public intervention, for which definite resources (e.g. funds) were allocated, will lead to certain effects (a specific product of intervention – e.g. results of industrial surveys), (2) the use of these products will change the situation of intervention beneficiaries (an immediate result will be observed – e.g. increased competitiveness, thanks to innovation, developed based on surveys results, having been implemented), which will finally (3) bring about definite effects within a broader socio-economic environment (this refers to different strategic results, corresponding to key intervention objective – e.g. generation of new, high quality jobs). Such an approach is based on the assumption that a given support measure, justified by the identified problematic situation or need, is a prerequisite for achievement of the

¹⁹⁶ M. Richie, *Theory Based Evaluation: A wealth of approaches and untapped potential*, European Commission 2012, p. 1.

anticipated (positive) change in the group of public aid beneficiaries, and ultimately this change will have a positive impact on the whole target group and broad socio-economic environment. The presented system, arranged according to the cause and effect relationships, models the intervention logics. However, in evaluation practice, this system must undergo further verification in order to clarify how and why individual components of intervention will (or will not) contribute to the achievement of the pre-determined results¹⁹⁷.

The considerations presented above are clearly visible in the case of intervention in the area of innovation. Namely, the low innovative potential of economic sphere is to be curbed by providing entrepreneurs with financial support, encouraging them to undertake research and development, and also by reducing the risk of failure of research or of economic use of innovative solutions, created based on R&D results. Ultimately, these activities, having reached a definite “critical mass”, should raise the level of innovation across the economy. Obviously, such considerations greatly simplify the whole issue of intervention organisation and process, because in practice the cause and effect relationships in this sequence prove far more complex – they depend on a number of factors which, additionally, do not need to be (and most often are not) unequivocal nor permanent, i.e. they can neither be controlled nor foreseen. Public intervention is delivered in a variable social, economic and legal environment, moreover, intervention itself affects changes taking place in that environment. Mechanisms governing intervention do not come down to just the “input – output” sequence, but they include a number of premises characteristic for intervention environment, as well as behavioural factors affecting the actions taken and choices made¹⁹⁸.

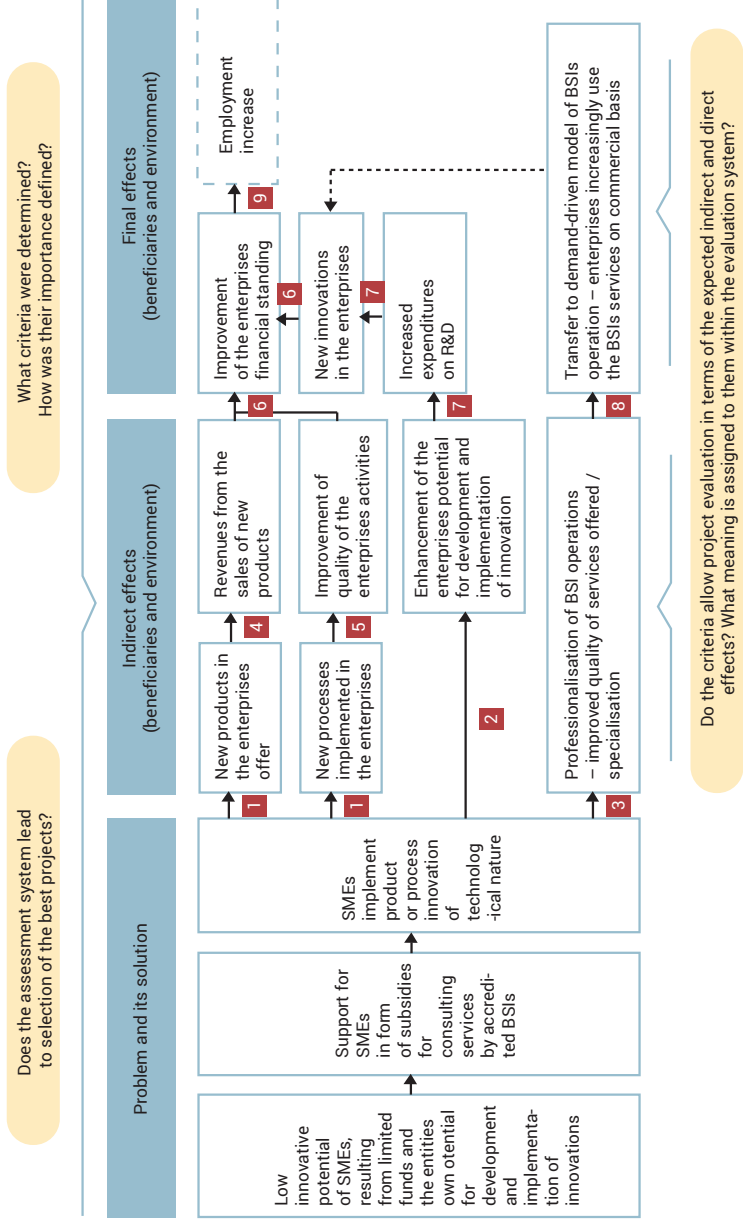
As revealed by results of our studies on project selection and evaluation systems under Operational Programmes, our use of the theory-based approach proved useful for evaluation and formulation of practical recommendations on selection system set-up and contents. Beyond any doubt, one of key success factors, crucial for the public intervention success (efficiency in achievement of objective) consists in proper designing of the entire project selection mechanism and ensuring its efficient operation. The point is, therefore to devise a selecting mechanism guaranteeing that best projects are selected, that operationalize the logics of intervention. Obviously, defining “the best project” and the respective set of selection criteria allowing to identify it as such is not a simple task. Analysis of the selection system – and within this system particularly the sub-system of project evaluation criteria – is particularly important, as it is the critical element of the theory of change, corresponding to the designed intervention.

¹⁹⁷ cf. P. Stronkowski, M. Andrzejewska et al. *Final Report - Evaluation of the support in the domain of social economy through the ESF under OP HC*, Coffey International Development, Warsaw 2013, p. 15.

¹⁹⁸ cf. “Evalsed Sourcebook: Method and Techniques”, European Commission, p. 51
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf
(access on 24.03.2017).

In the practice of our research, the above statement meant the necessity to attempt at reconstructing the intervention theory at different levels, i.e. that of the programme and of its individual instruments, i.e. measures and sub-measures, which provide framework for public aid system (including the project selection and evaluation system). This was necessary because usually the full intervention logics is not explicitly expressed and is not directly commented on in programme documentation – it is hidden under intervention objectives, measures proposed and implementation schemes. Moreover, the possibility to obtain information on the subject directly from the author of a given intervention, most often a number of decision-making centres, is usually very limited. These problems also result from the lapse of time and hence inability to address persons who participated in designing a given intervention. Furthermore, it often turns out that no in-depth descriptions of intervention theory in a form of formalised sources are available. All these make reconstruction of the full theory of change evoked by the intervention a fairly difficult task, but on the other hand such reconstruction is indispensable for analysing the project selection and evaluation system as an element which models that theory and allows for improvement of the system in terms of achievement of intervention goals. This is of particular importance for analysing the selection criteria, as their system should be designed in a way that would ensure that supported undertakings guarantee the best possible achievement of intervention objectives. An analysis of this system should be focused on assessment, whether the system will contribute (and to what extent) to the achievement of goals of support measures. In practice it may turn out that criteria that are applied fail to guarantee that goals are achieved. Only when such situations are identified precisely, it is possible to formulate recommendations how to improve the selection system and the evaluation criteria. The diagram below presents an example of the applied approach, based on reconstruction of the intervention theory, in the context of evaluation if project selection criteria are compatible with intervention objectives .

Diagram 1. Simplified logic model of Sub-measure 2.3.1 SGOP – Innovation promoting services by business environment institutions for SMEs – defined for the needs of analysing the system of project selection and evaluation



Source: Own paper based on documents and analytical work carried out under the project "Evaluation of the project selection system under SGOP – stage II", Consortium IMAPP, IDEA of Development Foundation, PAG Uniconsult and Jagiellonian University by order from the Ministry of Development, Warsaw 2017

Since theory-based evaluation aims to establish the reason why change occurs as well as what is the mechanism behind such change, and also whether a given change leads to anticipated results, the project selection and evaluation system is one of the components of this mechanism. Therefore, the analysis of selection criteria and the resulting conclusions may give an answer to the question why intervention is (or is not) successful. In our studies, the theory-based approach was used solely to identify assumptions and the anticipated results of support at direct and indirect level (as shown in the diagram above). This was the minimum condition which had to be met to be able to assess if criteria are well designed from the point of view of project selection in the analysed intervention. An analysis of deeper relationships and factors modelling intervention (apart from the criteria themselves and the formal pattern of the selection system) was not possible¹⁹⁹. In consequence, the answer to the question about the importance of criteria (their impact) as a component of the mechanism of change occurring within a given intervention may only be partial. The realities of our studies and the complexity of the subject were decisive in this context. It should be noted that the project selection system itself comprises many factors and that is why the system operates in a specific way. All in all, we are dealing here with a very complicated arrangement of different factors with relationships between them, which are all very difficult to control²⁰⁰. For instance, the existence of well designed criteria is a necessary but not a sufficient condition for intervention objectives to be achieved, as the way of intervention implementation – e.g. by experts-evaluators of applications – is always an essential issue. In this situation much will depend on the time assigned to assessments, and the possibility to select experts with specific qualifications and experience – these (exemplary) factors ultimately will have a significant impact on the quality of project assessment.

Moreover, it should be noted that evaluation recommendations indicating the need for sweeping changes in the criteria are barely acceptable by entities that commissioned the study, since that would require designing new criteria, followed by their formal acceptance and introduction into the evaluation process – this may complicate and delay the intervention implementation process. Studies results can also lead to conclusions undermining the legitimacy of using a given support instrument. Apparently, the importance of these issues grows with the engagement (in time and scale) in the implementation of a given programme. This involves the need for appropriate planning of assessment of project selection and evaluation systems. Studies of this type should be carried out as early as possible, because at the moment when the system has already been designed and particularly when

¹⁹⁹ This was due to the realities of our studies, i.e. by the time assigned for the work and the available budget. A similar situation is probably encountered in the case of other evaluations, concerning the project selection and evaluation systems.

²⁰⁰ Rogers P., *Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions, Evaluation*, SAGE Publications, Los Angeles 2008.

it is already in operation – the willingness to implement substantial changes in it is diminishing.

Ultimately, our experience has shown that reference to the theory-based evaluation in studies on project selection and evaluation systems is slightly “exaggerated”. Obviously it is useful, as already mentioned earlier – for identifying assumptions and effects of intervention in terms of checking if a given form of selection system and a given set of criteria will ensure selection of adequate projects. At the same time, however in view of the realities of our study, our analysis of intervention theory of had to be limited to a certain minimum. This was determined both by the contracting authority’s objectives and by the study realities. As to the objectives of studies dedicated to project selection and evaluation systems, our experience shows that when applying the theory-based evaluation it is sufficient to identify key elements of public intervention logics (e.g. measures / sub-measures in the operational programme). This is a sufficient basis for assessment whether the selection system and criteria enhance the probability of selecting projects having features which are desirable in terms of intervention goals. In the course of studies, seeking an answer to this question ultimately translates into analysing many detailed issues, comprising individual elements of the selection system, starting with competition documentation through criteria related aspects: 1) their respective number, considering all the key expectations of a given intervention in terms of products and results; 2) proper substantial adjustment of criteria to the intervention logics; and finally 3) technical aspects of their structure, including the way of calibrating evaluation scales and defining individual criteria.

Additionally, because of the usually expanded catalogues of criteria, it is essential to verify their consistency in the course of study process – the point is, we need to answer the question whether criteria make up a complete, adequately complementary set, thus ensuring the complexity of project evaluation. This is one of key elements of refinement of criteria sets which, if the requirement of complexity is met, become the key factor operationalizing the theory of programme intervention.

The functioning of project selection system is another important aspect here, relating to the way the criteria work in reality (practical effects of their application) – to what extent they focus on evaluating if the application for funding has been correctly prepared (e.g. its completeness, transparency of presentation), and to what extent on the (desirable) evaluation of the quality and development potential of the undertaking to be supported. Also the need to the study the discriminative functions of criteria should be mentioned – for instance certain applicants may be favoured or disfavoured. Without any doubt, as confirmed by our studies, the factors which determine the practical functioning of criteria have become a particularly essential element of the theory of change embedded in public intervention.

Experience in modelling project selection and evaluation systems in the 2007-2013 financial perspective, taking into account the innovation aspect

Despite the relatively many years of implementation of public aid schemes under the EU cohesion policy in Poland, there are still numerous difficulties associated with modelling the project selection criteria, which would well fit the intervention goals, defined in the programme documents. Identification of these difficulties is of particular importance for EU funds disbursement under the 2014-2020 perspective. The legislative solutions and guidelines functioning in the perspective clearly indicate to the need for further, strong orientation of intervention on the results. In these circumstances, lessons learned during the previous programming period become useful for identification of factors and mechanisms which model the theory of change of new public interventions²⁰¹.

Firstly, during the 2007-2013 perspective, proper reflection of the logics behind support measures in project selection criteria was a challenge²⁰². Issues in this scope resulted from decisions comprised in programme documents, which were not always obvious, thus obstructing – apart from everything else – a precise reconstruction of intervention logics. Studies under way in Poland show that problems stemmed from often poor precision in description (vagueness) of assumptions and objectives of intervention, and the lack of short-term action plans which would specify and update the objectives, thus allowing for instantiating the logics of aid measures²⁰³. In result, the processes of aid operationalization disclosed trends to instantiate this logics within the framework of criteria developed in the project selection system, which was disputable, as in these activities attention was mainly focused on ensuring formal conformity with the provisions of programme documents, instead of on seeking solutions which would guarantee the choice of projects with the highest potential for the implementation of programme objectives. In the sphere of supporting innovation in the economy, problems with determination of intervention logics resulted also from the lack of a consistent innovative policy at the national level²⁰⁴.

²⁰¹ In this chapter we present certain selected ones, formulated by the study team at the initial stages of individual studies on project selection and evaluation systems under SGOP, DPOP and EPOP in 2015-2017 (indicated in the introduction to the chapter).

²⁰² F. Gross F, L. Polverari, *Project generation and selection in Cohesion policy in 2014-20: between results orientation and micro-management*, IQ-Net Thematic Paper 34(2), European Policies Research Centre, University of Strathclyde, Glasgow 2014, p. 19-20.

²⁰³ K. Keler, S. Krupnik, Skórska P, Strzebońska A., *Metaevaluation of the research concerning the evaluation of projects selection criteria under Operational Programmes funded from the European funds in Poland in the perspective 2007-2013*, Uniwersytet Jagielloński by order from the Ministry of Regional Development, Cracow 2011, p. 6.

²⁰⁴ T. Klimczak, A. Lis et al., *Evaluation of the impact of Operational Programme Innovative Economy on enhancement of the business innovation*, WYG PSDB by order from the Ministry of Infrastructure and Development, Warsaw 2014, p. 199.

Secondly, focus on ensuring convenience, as well as reduction of logistic and cost burdens of criteria application, instead of maximizing their match to the programme goals, was an obvious weakness of measures striving to reflect intervention goals in the selection system. This problem was visible particularly in the assumptions for interventions concerning innovation. The project selection system used for the purposes of Smart Economy Operational Programme 2007-2013 (OPSE) attached more importance to the simplicity of evaluation. Therefore, as was highlighted in one of the evaluation studies, no efficient selection mechanism, based on assessment of the level of innovation of results brought about by supported project was developed²⁰⁵. The lack of balance between operational and strategic implementation levels of public intervention programmes was noticeable. The processes of project selections were often dictated by the prevailing needs of the operational perspective, not the strategic needs²⁰⁶. On the one hand it was the result of the strive to disburse the funds on time. And apart from that, the focus on effects of support is always more demanding than verifying correctness of the application in terms of individual assessment criteria.

Thirdly, a horizontal impact issue with OPSE consisted in excessive concentration of the forms of providing support in the form of nonreturnable instruments (subsidies). This form should be used less often and only in cases of supporting undertakings with breakthrough innovative features. In the remaining cases repayable (financial) mechanisms should be used on a broader scale, as these may also be reasonable in the case of funding innovative undertakings²⁰⁷. Obviously, the choice of the funding form (subsidy vs. repayable / financial instrument) affects the modelling of project evaluation criteria. In the case of repayable financial instruments, they may have a different, more simplified form.

Fourthly, a problem which appeared fairly often was the formulation of definitions of terms used in the project selection criteria so that they could be applied horizontally (for any type of intervention purposes), i.e. they resulted in uniformity of interpretation and were subject to universal measurement. Problematic issues were particularly noticeable in defining the term innovation and its different levels²⁰⁸, complete with the associated project risk measurements which often favoured already known ideas, which only required some tuning in, not the risky undertakings focused on radical de-

²⁰⁵ *Ibid.*, p. 202.

²⁰⁶ cf. K. Keler, S. Krupnik, P. Skórska, A. Strzebońska, *Metaevaluation of the research concerning the evaluation of projects selection criteria...*, p. 7.

²⁰⁷ M. Gajewski, J. Szczucki, "Financial engineering instruments - an unconventional source of financing innovative projects of the SME sector. Advantages in comparison with grant financing scheme", p. 51 in: J. Pokorski (ed.), *Towards Innovative Economy - Effects of Grants to Enterprises in Poland*, Warsaw 2011.

²⁰⁸ M. Gajewski, J. Szczucki et al., *Mid-term evaluation of the implementation status of priorities 3, 4, 5 and 6 of the OPIE*, PAG Uniconsult by order from the Ministry of Regional Development, Warsaw 2011, p. 8.

velopment and creation of completely new solutions²⁰⁹. In consequence, the selection system often led to supporting undertakings where the innovation level was not very high²¹⁰.

And finally it should be noted that a number of issues were associated with technical problems in designing project assessments criteria. Some matters associated with this problem are of prime importance for project selection, hence they also affect the degree of achievement of the programme goals. The number of criteria in sets is an interesting issue (excessive fragmentation was often noticeable, degrading the consistency of the whole set, sometimes causing “clearing” of certain criteria), as well as the types and ways of projects assessment – expert individual score, panel scores and the applied method for granting points for criteria. For instance, the various evaluation scales used (binary – zero-one, multi-element) were not uniformly matched with the substantial part of specific criteria. Sometimes, however they drastically limited the flexibility of scoring, thus obstructing differentiation in assessing project quality, or they left an excessive margin for interpretation, requiring additional definitions of the meaning of criteria during the process of evaluation.

The herein presented experience from the past 2007-2013 financial perspective shows that many various factors model the actual logics of intervention, but most often distort it. The project selection and evaluation systems, developed and applied in practice, turn out to be a factor of particular importance. Due to limitations in this scope, support is often granted to projects which do not ensure maximisation of intervention goals. Modelling of project selection and evaluation systems is a very important and very useful element operationalizing any public intervention, and thus it should given special attention, both during *ex ante* evaluation and particularly when checking the effects of application of selection criteria (results of projects selected on these grounds), with a view to further optimize the whole system for selection and evaluation of supported undertakings.

Lessons learned from application of the theory-based approach in evaluating project selection system in the 2014-2020 financial perspective

Evaluations of project selection systems in the 2014-2020 financial perspective are a good example of applying the theory-based approach. In the operational programmes for this financial perspective, the support for innovation and competitiveness of enterprises was allocated to one national programme (Smart Growth Operational Programme), one macro regional programme (Operational Programme Eastern Poland),

²⁰⁹ *Ibid.*, p. 126-127.

²¹⁰ P. Gorgol, J. Kotrasiński, A. Weremiuk, *Metaevaluation of the results of the evaluation research of Operational Programme Innovative Economy, 2007-2013*, Jagiellonian University by order from the Ministry of Regional Development, Warsaw 2012, p. 6.

and sixteen regional programmes. In accordance with the Guidelines for evaluation of the cohesion policy in 2014-2020²¹¹, the evaluation of criteria and project selection system should be included in evaluation plans for all operational programmes. Institutions which manage individual programmes were obliged to carry out such analyses within two years from acceptance of the final version of the programme document²¹².

Based on the to-date progress and results of evaluation of project selection systems under operational programmes for 2014-2020, an attempt can be made at summarising initially the process of implementation of such systems and the lessons learned. It should be noted, however that opinions presented below are based on the authors' perspective. Getting to know the contracting authority's perspective would be a very valuable supplement in the future. To start with, we will focus on analysing those conditions of evaluations of project selection system which, in the studies completed, were of crucial importance for the quality and usefulness of conclusions and recommendations.²¹³

1. Greater emphasis placed on the theory of intervention than in evaluations of the earlier programming periods should be basically assessed positively. It should be noted, however that evaluation of the project selection system is most often helpful in fine tuning and operationalization of the intervention logics of a given instrument. In the 2014-2020 financial perspective, the assumptions for aid at programme document level (which was the subject of negotiations with the European Commission) were usually formulated fairly vaguely, with details left to be settled in the so-called Supplement of Priority Axis of the given operational programme. Most often, the participatory evaluation formula included workshops, mainly designed to reconstruct the logics of a given aid instrument. Both evaluators and representatives of managing and intermediary institutions participated in such meetings of this type were attended. The workshops made it possible to clarify the anticipated intervention effects and to formulate postulates for possible changes in criteria or other elements of project selection systems. It should be noted in this context that the optimal solution would be to commence the evaluation of the project selection system as early as possible, even before the first competitions

²¹¹ *Guidelines for evaluation of 2014–2020 cohesion policy* (section 3.1.5 *Thematic scope of evaluation plans*), by order from the Ministry of Infrastructure and Development, Warsaw 2015.

²¹² In practice, evaluation of SGOP and EPOP was launched relatively quickly (as early as 2015), in the regional programmes, the majority were commenced only in 2016 or are planned for 2017.

²¹³ Summary based mainly on the following studies: (1) "Evaluation of the project selection system in SGOP - stage I and II", Consortium IMAPP, IDEA of Development Foundation, PAG Uniconsult and Uniwersytet Jagielloński on order from the Ministry of Development, Warsaw 2016 and 2017, (2) "Evaluation of the project selection system in OPDP", Consortium IMAPP, IDEA of Development Foundation, PAG Uniconsult and Uniwersytet Jagielloński on order from the Ministry of Development, Warsaw 2017, (3) "Evaluation of the project selection system in OPRD", Consortium IMAPP, IDEA of Development Foundation, PAG Uniconsult and Uniwersytet Jagielloński on order from the Ministry of Development 2014-2020", Consortium PAG Uniconsult, IMAPP on order from the Ministry of Development, Warsaw 2016.

are launched; because only then it would still be possible to make any essential corrections in the studied instruments.

2. Regardless of the corrective function mentioned earlier, evaluation of the project selection system does not allow for correcting the key elements of the incorrectly designed aid instrument. In the small number of cases, the analyses already completed lead to a conclusion that the problem does not reside in a single criterion or some other element of the selection system but rather in more fundamental assumptions behind a given intervention (e.g. incorrectly matched form of support, intervention objective or no justification for public aid altogether). Unfortunately, radical recommendations indicating the need for changes in these assumptions are fairly often treated with excessive reserve by the interested institutions.
3. Particularly at the stage beginning before launching of first competitions, analyses of the project selection systems consist not so much in classical evaluation but rather in development of an expert review, within a tight schedule. In practice, this means that both the quantitative and qualitative surveys may only serve as an indicative source of information, while the basis for formulation of conclusions and recommendations should be expertise and experience of the expert team. This does not mean, however that the survey component should be left out from the evaluation of the projects selection system using the theory-based approach. In practice, qualitative surveys (in-depth interviews, focus groups) with representatives of institutions involved in implementing interventions and with experts-members of project evaluating committees turn out to be particularly useful.
4. The quality of on-going cooperation between the contractor and the contracting authority is of fundamental importance for the usability of the results of evaluation of the project selection system. The true commitment, readiness to listen to the arguments of the other party and flexible adjustment of work to the programme implementation realities (though obviously the implementation dimension should not overshadow the much more essential logics of intervention foreseen for a given instrument).
5. It is worth mentioning here that evaluations of the project selection system which use the theory-based approach belong to a specific group of studies, where the final reports are a relatively less essential effect of evaluation efforts. The on-going cooperation between the contracting authority and the contractor, which was mentioned earlier, is of much greater importance, including working meetings, workshops and working opinions exchanged (formally or informally). This allows brainstorming and joint development of optimal (from the adopted intervention logics point of view) and feasible solutions. The study reports themselves are prepared relatively late, when essential arrangements concerning the project selection process have already been made. Nevertheless, it is valuable to formulate a relatively large number of conclusions and recommendations, provided of course they are properly justified and fit for the given implementation realities. Excessive selectivity

and seeking complete consensus between both parties may lead to elimination of creative thinking and premature abandonment of ideas which may have potential to improve the quality of intervention in the future.

6. If the evaluation of the project selection system is to be useful, commitment is required not only of the contracting authority itself (most often it is an evaluation unit at the managing institution) but also of all the other entities responsible for implementing a given intervention: persons responsible for programming at the managing institution and respective structures within intermediate and implementing institutions. Involvement of these entities is necessary not only at the stage of acceptance of results of work, but also in the process. It seems particularly essential for experts who participate in evaluation to have regular access to the information on progress of work associated with the processed changes in the evaluated instruments, as this is of particular importance to the adequacy of formulated conclusions and recommendations. An ideal solution (and feasible in practice) also consists in direct involvement of evaluators in the project selection process as observers (e.g. work in expert panels who meet the applicants face to face).

With respect to supporting innovation and entrepreneurship, based on the completed evaluations of the project selection system in national operational programmes for 2014-2020, several initial key conclusions and recommendations may be formulated. They should be perceived as hypotheses which require verification under the *ex post* evaluation.

- In the case of more complex aid instruments, especially those involving R&D work or implementation of the results of such work, it seems that an expert panel, providing an opportunity of direct contact with the applicant, will be the optimal form of evaluation. Such formula allows verification of key elements of project documentation as well as a detailed discussion on technological and business aspects of projects.
- Usefulness of the panels and their ability to identify the best undertakings, in turn, depends on the extent to which the institutions responsible for project selection succeed in involving competent persons and organising their work properly. The practice of operational programmes 2014-2020 implementation shows that the number of potential experts is limited and this fact should be taken into consideration during drawing up the competition announcement schedules. Improvement of expertise exchange is also of particular importance.
- The practice from the 2007-2013 and 2014-2020 financial perspectives shows that excessive parameterization of project selection criteria has not proved adequate in supporting innovation and competitiveness of enterprises. This is especially true with respect to issues which are subjective or which require expert opinion on a given business idea (product innovation, market potential of the project etc.). Precise determination of score-gaining elements promotes criteria-wise preparation

of applications, thus making it difficult for the experts to evaluate the undertaking as a whole in terms of the intended effects of intervention.

- Over the recent years, the friendliness and transparency of the project selection procedure have been successfully improved from the applicants' point of view, particularly with respect to accessibility of respective materials and the quality and functionality of IT systems. At present, it seems that redefinition of the role and formula of informative-promotional measures is the aspect worth strengthening and tuning in. The goal of these measures should not only consist in transfer of basic information about the announced competitions but also in providing active support to applicants in the process of drawing up high quality applications (e.g. by providing entrepreneurs with a possibility to consult elements of applications for funding with experts – members of the project evaluating committees, without any obligations to follow their advice).

Summary

As presented in the first part of this chapter, reconstruction and consideration of various aspects of the theory of intervention in evaluation of the project selection and evaluation systems may improve the efficiency of public funded support measures. It should be noted, however, that the results of analyses based on the theory-based evaluation, though obviously useful at the stage of identifying assumptions and results of intervention in terms of matching the project selection and evaluation systems, are often attenuated by a limited will to introduce more in-depth changes into the systems already in operation. These limitations also result from the realities of ongoing evaluations of systems and evaluation criteria (particularly the time and budget available for research, and its subsequent available scope). The above means that using this type of research in practice requires a proper definition of its role and its proper design, as well as its synchronization with the process of intervention programming and implementation. The results of evaluations of project selection systems already completed show that the approach applied in the 2014-2020 financial perspective basically meets those expectations. Efficient communication between experts–researchers (contractors) and all the institutions responsible for implementation of the aid schemes is the most important success factor in research of this type. Cooperation between both parties must consist in continuous exchange of opinions and ideas concerning the evaluated instruments, hence extending far beyond the standard evaluation framework, where most often consultation on the final study report is the sole area of cooperation. Evaluations of the project selection systems under operational programmes for 2014-2020 show explicitly that – even given the very limited time for performing the analysis – their strongest possible coordination with the programming process is recommended. Otherwise the application of the theory-based approach in practice will be much more difficult and will not give a chance for actual translation of the evaluation process into improvement of the quality of the studied aid instruments.

About the authors

Maciej Gajewski

PhD in Economic Sciences, lecturer at the Faculty of Management at the University of Warsaw. He has more than 25 years of experience in providing research (evaluation), consulting and training services. He is a partner and expert at the consulting company Policy & Action Group Uniconsult. He specialises in counselling on the creation and functioning of financial instruments aimed at improving the availability of external financing sources, in particular loan, guarantee and equity instruments, for the sector of micro-, small and medium size enterprises. His professional activity also includes the issues of financial support for innovative projects, in all phases of their development. He has many years of experience in designing and conducting research and in managing expert teams involved in the evaluation of public interventions. He is an evaluator and expert in the field of the creation and implementation of operational programmes and EU-financed measures implemented as part of these programmes – both at the national and regional level. He is an author or co-author of many publications presenting the evaluation results of public intervention programmes.

Andrzej Reguński

Since 2013, the Managing Director of the IMAPP company, in the past the Director of Public Policy Evaluation Programme at the Institute for Structural Research. Economist by education, he graduated from the Warsaw School of Economics. So far, he has participated (mostly as a coordinator) in more than 50 evaluation studies and other analytical projects, carried out on behalf of the government and local administration. He specialises in issues related to the institutional development policy system in Poland, innovation and energy. In recent years, he has coordinated, *inter alia*, *ex ante* evaluations of the Partnership Agreement and four operational programmes for 2014-2020 (Infrastructure and Environment Operational Programme, regional operational programmes of the Łódzkie, Podkarpackie and Pomorskie Voivodeships), research related to updating voivodeship development strategies (Małopolskie, Podkarpackie, Pomorskie and Warmińsko-Mazurskie), as well as a number of analyses concerning building a system of indicators in strategy and operational papers at the national and regional level.

Bibliography:

1. Gajewski M., Szczucki J. et al., *Ocena stanu realizacji 3,4,5 i 6 Priorytetu PO IG w połowie okresu programowania*, PAG Uniconsult, Warsaw, 2011.
2. Gajewski M., Szczucki J., "Financial engineering instruments - an unconventional source of financing innovative projects of the SME sector. Advantages in comparison with grant financing scheme", in: Pokorski J. (ed.), *Towards Innovative Economy - Effects of Grants to Enterprises in Poland*, Warsaw, 2011.
3. Gorgol P., Kotrański J., Weremiuk A., *Metaewaluacja wyników badań ewaluacyjnych Programu Operacyjnego Innowacyjna Gospodarka, 2007-2013*, Warsaw, 2012.
4. Gross F., Polverari L., "Project generation and selection in Cohesion policy in 2014-20: between results orientation and micro-management", *IQ-Net Thematic Paper 34(2)*, European Policies Research Centre, University of Strathclyde, Glasgow, 2014.
5. Keler K., Krupnik S., Skórska P., Strzebońska A., *Metaewaluacja badań dotyczących oceny kryteriów wyboru projektów w programach operacyjnych współfinansowanych z funduszy europejskich w Polsce w perspektywie 2007-2013*, Jagiellonian University, Cracow, 2011.
6. Klimczak T., Lis A. et al., *Ocena wpływu Programu Operacyjnego Innowacyjna Gospodarka na zwiększenie innowacyjności przedsiębiorstw*, WYG PSDB, Warsaw, 2014.
7. Richie M., *Theory Based Evaluation: A wealth of approaches and untapped potential*, European Commission, 2012.
8. Stronkowski P., Andrzejewska M. et al., *Raport końcowy - Ocena wsparcia w obszarze ekonomii społecznej udzielonego ze środków EFS w ramach PO KL*, Coffey International Development, Warsaw, 2013.
9. Rogers P., "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions", *Evaluation*, SAGE Publications, Los Angeles, 2008.

Other:

1. *Ewaluacja systemu wyboru projektów w POIR - etap I i II*, IMAPP, IDEA of Development Foundation, PAG Uniconsult, Jagiellonian University, Warsaw, 2016 and 2017.
2. *Ewaluacja systemu wyboru projektów POPC*, IDEA of Development Foundation, PAG Uniconsult, Jagiellonian University, IMAPP, Warsaw, 2017.
3. *Ewaluacja systemu wyboru projektów w Programie Operacyjnym Polska Wschodnia 2014-2020*, PAG Uniconsult, IMAPP, Warsaw, 2016.
4. *Evalsed Sourcebook: Method and Techniques*, European Commission.
5. *Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020*, Ministry of Infrastructure and Development, Warsaw, 2015.

Experience in the application of theory-based evaluation at regional level

Introduction

After the accession to the European Union (EU), the Polish administration had to confront with the requirements applicable in the EU, including those related to the organization of evaluation of programmes supported with the EU funds. Due to the difference of experience and skills in this area it became necessary to i.a. quickly make up for the gap in evaluation potential of institutions which were responsible for EU funds' redistribution. Even though the first two EU financial perspectives implemented in Poland (2004-2006, 2007-2013) helped establish an appropriate national evaluation system (and served its gradual decentralization), the 2014-2020 perspective featured new challenges – related among others with rigorous methodological requirements for evaluation, which „aims at obtaining more reliable and useful information about real effects of cohesion policy”, i.a. development and use of “advanced research methods, including in particular evaluation (...) methods based on theory”²¹⁴.

The aim of this Chapter is to present the first experiences of theory-based evaluation (TBE) application and the resulting challenges, which had to be faced by Polish local government administration. The analysis was conducted based on 15 *ex ante* studies of operational programmes for the period 2014-2020. Moreover, this chapter discusses general difficulties in building evaluation potential in regions, as well as those resulting directly from the application of theory-based evaluation. This is how the main topics were arranged in this article, which is based on a review of the literature and evaluation studies conducted at the regional level in Poland.

Evaluation in regions

Three EU programming perspectives, in which the Polish regions have been participating since the EU accession, have determined the subsequent milestones of evaluation system development at central and regional levels²¹⁵. In the period 2004-2006 the whole evaluation system was designed and implemented by man-

²¹⁴ Cf. Guidelines of the Minister of Infrastructure and Development dated 22 September 2015 on evaluation of cohesion policy for the period 2014-2020, Chapter 1.2 (4) p. 8 (<http://www.funduszeuropejskie.gov.pl/strony/o-funduszach/dokumenty/wytyczne-w-zakresie-ewaluacji-polityki-spojnosci-na-lata-2014-2020> - access on 28.04.2017).

²¹⁵ Cf. Bachtler J., „Ewaluacja regionalnej polityki w Europie: kultura, zaangażowanie i potencjał”, in: Olejniczak K., Kozak M., Ledzion B. (eds.), *Teoria i praktyka ewaluacji interwencji publicznych*, Wydawnictwa akademickie i profesjonalne, Akademia Leona Koźmińskiego, Warsaw 2008,

aging institutions located at the central level. Since 2007, under the decentralization of cohesion policy management system, Polish voivodeships, having prepared regional operational programmes (ROPs) have become responsible for developing evaluation system at a regional level.²¹⁶ Insofar as the evaluation system may be estimated to have been well-established at the central level in 2007, including broadly understood evaluation culture, the regions started to learn how to conduct evaluation and how to develop evaluation culture only during the implementation of ROP 2007-2013, when specialized evaluation units were established in all the 16 voivodeships, along with evaluation steering groups, and when the first evaluation studies were launched. Thus, it can be said that the foundations for the evaluation system were laid, and special work targeted at the development of evaluation potential²¹⁷ and culture at the voivodeship administration level was carried out in the 2007-2013 programming period.

The 2014-2020 financial perspective makes the regions face new challenges, due to new requirements towards methods of evaluation planning and performance. While imposing more demanding regulations concerning operational programme evaluation system on the Member States, the European Commission (EC) forced further development of administrative potential²¹⁸ (including evaluation units at regional level) and a new approach to planning and conducting evaluation studies in Poland. For the first time EC defined the scope and procedure of preparing and

²¹⁶ Public interventions, managed and implemented in regions, and in particular Regional Operational Programmes and the regional component of Human Capital Operational Programme were subject to direct analysis and evaluation by particular voivodeships' administrations (Cf. Guidelines No. 6 of the Minister of Regional Development dated 30 May 2007 concerning evaluation of operational programmes for the period 2007-2013).

²¹⁷ Evaluation potential may be understood as a capability to conduct evaluation studies of quality, the results of which are absorbed and used for public management quality improvement. The sources of those capabilities include evaluation units, implementing units, decision makers, recipients as well as social capital. Cf. M. Szałaj, *Budowa potencjału ewaluacyjnego*, in: *Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka*. Collective work, Marshal Office of Opolskie Voivodeship, Opole 2009, p. 76.

²¹⁸ As at the end of 2015 there were 34 evaluation units operating within cohesion policy evaluation system in Poland. They were located in managing institutions, intermediate bodies, 2nd degree intermediate bodies and implementing institutions. They employed the total of 158 members of staff dealing with evaluation, 70% of whom were the employees of regional units. According to the studies of the Ministry of Development, the number of employees of evaluation units decreased significantly in the period 2013-2015 (in 2014 there were 188 employees, in 2013 - 196). It may be partially explained with changes in the structure of evaluation units at regional level and with the establishment of joint units, responsible for the evaluation of support from the European Regional Development Fund and European Social Fund, provided under Regional Operational Programmes 2014-2020 (Conf. *Rocznik Ewaluacji 2013*, Ministry of Development, Warsaw 2014, p. 8-9; *Rocznik Ewaluacji 2015*, Ministry of Development, Warsaw 2016, p. 6.).

approving evaluation plans for operational programmes in every detail²¹⁹. The increase of evaluation's role within the regional programme management system (including Regional Operational Programmes) is motivated above all by the need of improving the quality (methodology) of evaluation carried out and with the increase of the scope of using the outcomes thereof²²⁰. Increase of methodological requirements aims to obtain more useful information about real policy effects. The main task for evaluation in the region is to indicate which part of change in socio-economic situation (e.g. changes in the value of result indicator) resulted thanks to cohesion policy implementation. The second important task is to explain mechanisms of cohesion policy impact, which means to answer a question why such a change has taken place. A choice of an appropriate methodological approach, which would ensure quality evaluation, is essential here. The guidelines point directly to a need of applying advanced study methods, including in particular counterfactual impact evaluation methods and theory-based evaluation approach²²¹. Consistent use of this approach enables decision makers to better understand regional requirements necessary for their intervention to be successful, and to improve their management quality as a consequence. By making a reference to theories which describe mechanisms and directions of a region's development, it facilitates collation - in the process of evaluation - of selected actions with alternative actions, which may also be justified in light of current state of the art and experience (especially regional ones). Thanks to that ineffective mechanisms may be identified more quickly, their faults may be explained more precisely and they can be eliminated once and for all by substituting them with better solutions. A major challenge here is the transition from general theories of regional development to their regional adjustment achieved thanks to progressive identification of local conditions and linking the latter with the

²¹⁹ *Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund* is the basic document which determines, at the Member State level, the shape of evaluation system of support from European Regional Development Fund and Cohesion Fund for 2014-2020. It was drafted by General Directorate for Regional and Urban Policy in 2014 r. (http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf – access on 19.03.2017). For the remaining funds (e.g. European Social Fund) or other financing mechanisms (e.g. aid schemes), separate horizontal guidelines were additionally developed by particular EC General Directorates (e.g. *Commission Staff Working Document Common methodology for state aid evaluation*, prepared by GD for Competition). At the state level, the EC guidelines concerning evaluation in the new programming period were detailed in the Guidelines of the Minister of Infrastructure and Development dated 22 September 2015 concerning cohesion policy evaluation in 2014-2020).

²²⁰ Chapter 1.2 of the Guidelines of the Minister of Infrastructure and Development provides that: „In the period 2014-2020 the role of evaluation in the process of OPs manager under cohesion policy shall increase. This is reflected in formal legislative clauses which provide the basis for implementing EU funds and EC-formulated guidelines. Increase in the role of evaluation in the process of OPs implementation also means more duties for the EU Member States, which translates into a need to reorganize and adjust evaluation systems to new requirements” (Guidelines of the Minister of Infrastructure and Development dated 22 September 2015 concerning cohesion policy evaluation in the period 2014-2020).

²²¹ Cf. Guidelines of the Minister of Infrastructure and Development dated 22 September 2015 concerning cohesion policy evaluation in the period 2014-2020, Chapter 1.2 (4) p. 8.

intended actions and expected results. Such an approach requires good methodological preparation, a will to engage in broad searching of the most plausible theories to explain the mechanisms of achieving results on a regional scale, a discipline and consistency in the implementation of the interventions adopted, and above all it requires time and open, partner-like attitude of all those who participate in the processes of a public programme development and evaluation.

The meaning of programme logic in the context of theory-oriented evaluation

Thanks to the promotion by EC of theory-based evaluation in the 2014-2020 perspective, decision makers are mobilized to focus on causal relationships triggered when public interventions are pursued. They may be described by applying a metaphor of a programme's „black box“ which contains the mechanism able to change a problem identified „at the input“ in the result achieved „at the output“.

Evaluator's identification of tenets behind the change, which is to be achieved by a given public intervention, is a crucial stage of such evaluations²²². Depending on the concept of theory-oriented evaluation²²³ there are different beliefs as to the content of the „black box“, and as a result, there are different tasks the evaluator must carry out in the process of identifying the conceptual basis of the programme logic²²⁴. As

²²² A. Markiewicz, I. Patrick, *Developing Monitoring and Evaluation Frameworks*, Sage Publications Inc. 2016, p. 35.

²²³ Find more in that publication in the Chapter „Przegląd systematyczny koncepcji ewaluacji opartej na teorii“, by Katarzyna Hermann-Pawłowska and Paulina Skórska. [Systemic Overview of Theory –Based Evaluation Concept]

²²⁴ For example, according to N. Stame, *theory-driven evaluation* (TDE) assumes lack of a theory, which often accompanies public interventions („black box“ is empty). Evaluator's task is thus to provide a set of theories which may explain the achievement of objectives assumed, to give a tip on what may work in a given intervention, based on social sciences. This is carried out by analysis of proposed intervention mechanisms, discussion between stakeholders and evaluators on expected outcomes and study on why and how the programme works, following the existing normative and causal theories. In *realistic evaluation* approach theories lead to assumed results thanks to the engagement of stakeholders in implementation thereof. In this case the „black box“ is *inhabited by the people* who make use of opportunities created by a given intervention, and achieve outcomes (O) by activating change mechanisms (M), which should work in a given context (C). Thus, evaluation is based on reconstruction of a CMO setup, *that means it has to determine how the mechanism could work in a given context and to ask the stakeholders how it worked from their perspective and what evidence for it they can deliver*. The programme context is taken account of in this approach first of all in its sociological environment characteristics, and evaluation of public intervention is obtained through adjudication confirming its actual course and outcomes achieved. In case of *theory-based evaluation* (TBE) the „black box“ is full of theories, called theories of change, which are formulated independently by each stakeholder, and thus often more than one theory for a given programme, which take the form of assumptions, or tacit understandings etc. During TBE one must disclose change mechanisms on which the achievement of expected outcomes of the intervention was based. On one hand it requires a consensus as to which of theories of change assumed by particular stakeholder deserves to be tested by its inclusion to the intervention programme (*ex ante* evaluation). On the other hand, TBE enables to check the effectiveness of the mechanism launched, thanks to evaluation of interventions conducted (*ex post* evaluation), which contributes to generate new theories of change. Cf. N. Stame, *Theory-based Evaluation and Types of Complexity*, Evaluation, 2004, vol. 10(1), pp. 58-76, p. 61, 63.

a rule, the process requires the evaluator working with stakeholders to better understand how the programme is conceptualized to solve the problem. The evaluator helps the decision makers to finely define doubtful areas of change theory (not confirmed by cause-effect relationship) and to select appropriate study methods to dispel those doubts.

In TBE it is helpful to differentiate among *programme theory*, which refers to the intervention's concept, and a more operative programme (intervention) logic - *implementation theory*, which are mutually combined but which constitute different stages of public policy development²²⁵. The programme theory and implementation theory are usually presented graphically as a diagram showing causality relationship of mechanisms which produce change (see Figure 1).

During a programme evaluation, both programme theory and programme logic should be subject to evaluation. For the programme theory one must test the plausibility of intervention project's design and determine its components, because TBE assumes that intervention project is a result of beliefs and of impact of individuals who participate in the project's formulation – also those beliefs which have not been articulated during the development thereof, but have actually been expressed in it. On the other hand, intervention logic evaluation involves a determination of how a programme is implemented on the one hand and of its cause-effect relationships which have been taken into account by the intervention project on the other hand.²²⁶

²²⁵ Programme theory is an explanation how and why it is believed that programme launch will produce the intended outcomes. In contrast, programme logic includes purposeful and sequential process of going from program actions to its outcomes, which takes place within a defined time. Intervention logic includes resources, planned actions, their products and outcomes, identified as milestones of logic intervention model. Conf. A. Markiewicz, I. Patrick, *Developing Monitoring and Evaluation Frameworks*, Sage Publications Inc. 2016, p. 74.

²²⁶ *Ibid.*

Figure 1. Programme theory and intervention logic – an example of a regional programme to increase children’s school attendance



Source: adapted from A. Markiewicz, I. Patrick, Developing Monitoring and Evaluation Frameworks, Sage Publications Inc. 2016, p. 86.

In this way, by studying the programme logic one supports the development of more perfect interventions – based on actually operating mechanisms of change and change implementation – and one improves the programme management. Such an approach fulfils the *evidence-based policy* principle at the same time.

Benefits of TBE for evaluation at a regional level have been summarised in a table below.

Table 1. Importance of TBE for evaluation at regional level

Type of evaluation	TBE component	
	Programme theory	Implementation theory (program logic)
<i>ex ante</i> Evaluation	Programme theory reconstruction <ul style="list-style-type: none"> • Identification of programme assumptions (including the tacit ones) • Evaluation of assumptions' correctness in the light of actual knowledge related to a given area of intervention, i.e. academic theories, experience of other regions (benchmarks) and regional features (context) • Verification of assumptions' coherence in the context of programme goals and conditions 	Implementation theory reconstruction <ul style="list-style-type: none"> • Identification of causal relationships between programme components (mechanisms), supposed to lead to intended results • Evaluation of assumptions' correctness in the light of knowledge, experience and regional context • Verification of internal coherence among all intervention mechanisms to launch • Verification of adequacy of actions planned by intervention implementing institutions against the projected results and regional context

Type of evaluation	TBE component	
	Programme theory	Implementation theory (program logic)
<i>on-going / ex post</i> Evaluation	<p>Linking the programme theory with outcomes achieved</p> <ul style="list-style-type: none"> • Identification of programme theory and a path of its (possible) changes during the intervention course • Assessment of the degree to which the programme theory is confirmed in light of outcomes achieved (products, results, impact) • Identification of conditions for programme theory effectiveness, considering the knowledge, skills and behaviours of all actors involved in the intervention at each level thereof (design, implementation, evaluation) as well as regional context • Assessment of the usefulness of programme theory components to subsequent interventions and justification for continuation and/or modification of intervention (<i>on-going</i> evaluation) or repetition thereof (<i>ex post</i> evaluation). 	<p>Linking the implementation theory (programme logic) with achieved outcomes</p> <ul style="list-style-type: none"> • Identification of programme logic and a path of its (eventual) changes during the intervention course • Assessment of effectiveness of intervention mechanisms launched • Identification of intervention mechanisms' effectiveness' factors, considering the awareness, attitudes and behaviours of all the actors involved in the intervention implementation as well as regional context • Assessment of the usefulness of given mechanisms for subsequent interventions as well as of the justification for the intervention's continuation and/or modification (<i>on-going</i> evaluation) or relaunching thereof (<i>ex post</i> evaluation).

Source: Own study.

In line with *realistic evaluation*, the hitherto design of public interventions features context as an important factor which influences the effectiveness of the change theories launched, and which includes environment, culture and habits of public administration. It is the context of public intervention which seems to be a particularly valuable contribution of that approach to the improvement of quality of Regional Operational Programmes (ROPs). Drawing attention to specific conditions of a given region (e.g. local culture, social capital, administrative culture), which might have significant

influence on results achieved by the change mechanism launched, helps to assess the intervention outcomes more precisely.

TBE-related lack of possibility to fully compare the outcomes of public policies implemented in various situations (limited potential of generalizations), may become its main asset. It forces comparison of various intervention situations (e.g. identical support instruments implemented in various regions), so that to understand what and why works in given circumstances but does not work in every situation. This leads to a deeper understanding of conditions affecting intervention mechanisms (especially local factors) and as a result - to a greater flexibility of adjustment of policies developed on the basis of so called good practices. It also leads to their being more adequate to needs.

At the same time, however, using the capabilities of TBE may become a challenge for the regions in Poland due to still inadequate knowledge of TBE, significantly limited experience of applying thereof, on both sides: of the contracting party (client) and of the evaluators, as well as organizational and financial limitations of evaluation units located in Marshal offices. Building awareness of specific regional character and the need to adjust intervention projects in the heads of the decision makers requires time, as far as both the client and the evaluator are concerned. The importance of factors related to organizational culture of the contracting party and its will to get to actual assumptions, often tacit, which drive the choice of specific actions to the intervention design, may not be underestimated, either. TBE will show its potential and will be useful indeed if it serves to verify the intervention's directions and to identify the most effective problem solving mechanisms in the context of a given region. This also calls for real interest in the outcomes of evaluation to be shown by all the actors of intervention –from top-level decision-makers to beneficiaries. That interest results from the will to learn, also about oneself, and from the will to eliminate constraints related with individual and group decision making processes.

TBE practice at regional level (taking account in particular of *ex ante* ROP 2014-2020 evaluation)

According to the assumptions, the inclusion of TBE approach to *ex ante* evaluation helped improve operational programme at the stage of its concept formulation. The improved setup of causal linkages among needs, objectives, resources and actions taken under public intervention may increase chances for its effective implementation. At the same time, sorting out the programme logic facilitates the performance of *on-going* and *ex post* evaluations by clearly showing which elements have been deemed crucial for the achievement of the intended goals in the programming process. The assessment of changes in these elements during and following intervention may enable us to verify the correctness of the whole concept of operation and acknowledge or reject its being used in the future.

Due to a requirement of conducting *ex ante* evaluations for draft regional operational programmes for the period 2014-2020 (Article 55 of the Council Regulation No 1083/2006²²⁷) so that for all the ROPs all their logic models (implementation theories) were reconstructed and assessed, it is the evaluation reports of regional programmes which show the state of the readiness to apply TBE at regional level. The main goal of *ex ante* evaluation of draft operational programmes was a comprehensive evaluation of appropriateness and coherence (external and internal) of the ROP and programme implementation system, including human and administrative potential, and the monitoring system. Verification of external cohesion required determination of the degree of RPO being in line with European, national and regional horizontal strategies and policies. The evaluators were also faced with a task which required knowledge and experience, as they had to formulate conclusions leading to recommendations (eventually) concerning changes in the operational programme.

Pending negotiations of legal framework for the new perspective, i.a. partnership agreement at the national level were a significant constraint when conducting *ex ante* evaluation of programmes prepared for 2014-2020. As a result, in many cases it was draft ROPs that were subject to evaluation rather than final documents. Moreover, decision making process became undermined, on the side of the contracting party and the evaluator as well. As a result, intervention logic was subject to major changes during the work on the programmes; the changes involved a redefinition of specific goals, and evaluations were carried out as a participative process, involving evaluation unit teams, external experts as well as employees of Managing Authority who were drafting ROP 2014-2020. Participation meant simultaneously conducted processes of programming and evaluation with the application of expert knowledge of the team of evaluators, however, changes in the logic (intervention) model could be introduced under the influence of also other factors than mere conclusions drawn from the evaluation.

Assessment of intervention logic and proposed forms of support followed priority axes. Full reconstruction of the programme logic was not performed in any *ex ante* evaluation for particular ROPs.

²²⁷ Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006.

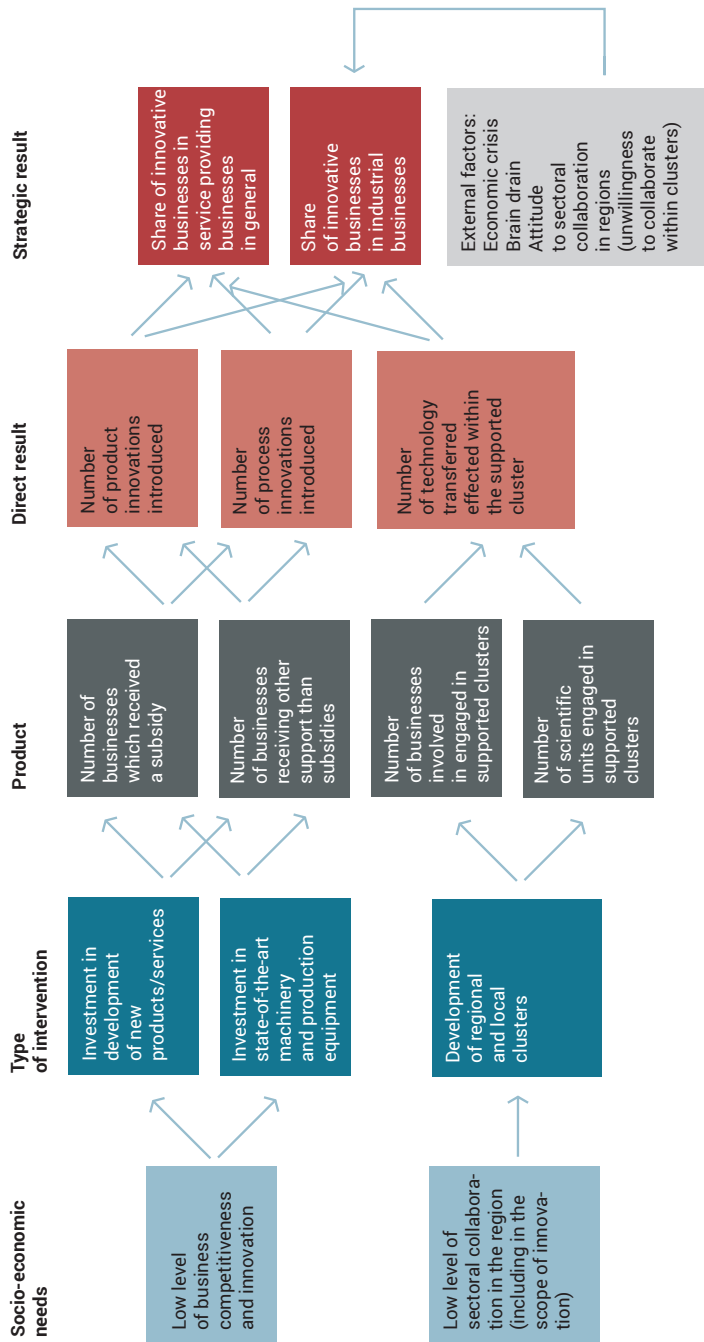
After a detailed overview of 15 final reports on ROP *ex ante* evaluation one can formulate a few major conclusions:

- Some evaluators are prepared to carry out a formally correct TBE. This means that evaluator teams reconstruct correctly the causal relationships, on the basis of the ROP documents, and evaluate them in a broader or narrower scope, but mainly in the context of problems described in the diagnosis – without verifying the rationale for the selection and prioritization of the problems to solve.
- Evaluators demonstrate two divergent approaches to their tasks: on one hand they tend to take the burden of determining base theories, broadening the scope of information and building intervention logic models for the ROP under evaluation; on the other hand there are tendencies to carry out purely formal verification focused on inconsistencies of the material evaluated. In such case evaluators limit themselves to conclude that it is impossible to assess the programme logic in case of some of the axis under analysis. Both approaches limit formative character of each evaluation, which is a condition for launching processes in a region, which involve broadening knowledge about the mechanisms of a given region's development.
- Final reports on ROP evaluation show significant differences in understanding the nature of programme theory and intervention logic. Where the latter is reconstructed, compliance with regional diagnosis is given as the programme cohesion justification. In some evaluations, reconstruction of intervention logic has been boiled down to copying ROP clauses while ignoring actual causal linkages. Other studies reveal lack of relationship between the goals, the actions and the specific types of outcomes as well as the ROP conditions. If ROP conditions are taken account of in logic intervention models, they usually involve external factors with no regard to the internal context. Finally, intervention logic is reduced to appropriateness and cohesion of actions planned.
- Intervention logic is usually illustrated with cause-effect diagrams. In some evaluation reports the programme logic is only described in a form of a paragraph or a table, which is less intelligible than a graphic presentation and may lead to misunderstanding between the parties.
- Otherwise the available *ex ante* evaluations point to good technical preparation of evaluators who are able to identify flaws in the ROPs evaluated and are able to formulate recommendations to enable them to be rectified.
- Repetitiveness of a relatively small number of evaluators makes it possible to follow the dissemination of knowledge about the necessary TBE elements (e.g. about the usefulness of presenting intervention logic in the form of diagrams), and a growing creativity of evaluation teams. They form interesting collations which increase the informative role of reports. One has to point out, however, that tables and block diagrams are not the best form of describing the reconstructed programme logic, because they do not reveal cause-effect relationships which lead to certain defined results.

- The method of justifying the selection of evaluation methods does not look very persuasively against this background. The obvious advantages of some of them (IDI, FGI) are not exploited in order to understand the assumptions behind intervention decision-makers. In general, it seems that the human aspect in ROP *ex ante* evaluation RPO 2014-2020 has been ignored to a broad extent, which is contrary to the nature of TBE. Ignoring the implementation theory as well should also be regarded as an important shortcoming.

During a detailed analysis of an *ex ante* report concerning ROP 2014-2020 for Opole voivodeship it can be noticed that the application of TBE at the regional level encountered certain difficulties. The figure below presents one of diagrams prepared for Priority Axis *PI 3.3* „Support of creating and broadening advanced capabilities for product and services development”, supported under ROP WO 2014-2020, which only very generally describes the recommended intervention logic (see Figure 2.).

Figure 2. Cause-effect diagram describing recommended intervention logic for PI 3.3 „Support of creating and broadening advanced capabilities for product and services development”, which has been taken account of under ROP for Opolskie Voivodeship 2014-2020.



Source: Attachment 3 to Final Report „Ex-ante evaluation of operational programme for Opolskie Voivodeship 2014-2020”, The Marshal Office of the Opolskie Voivodeship, Opole 2014.

It has to be pointed out that intervention logic (implementation theory) presented in the diagram points to outcome achievement mechanisms in existing environmental conditions, which were studied by project developer before planning the intervention. It seems that the main value of evaluator's work, especially at the stage of programme *ex ante* evaluation conducted by means of participative method - should be to support the process of intervention planning through constructive indications referring to possible inconsistencies or irregularities of intervention logic (e.g. making intervention logic more consistent and filling the gaps). Recommendations referring to formal errors in documents prepared, inconsistencies with top level documents are the easiest to formulate and are most frequently used. However, proposals for changes referring to modification of designed instruments (actions), which are to launch an effective change mechanism, pose a bigger challenge to the evaluator, as they require knowledge about effectiveness of various solutions and their conditions. Considering relatively short deadlines for evaluation studies and the simultaneous programming process which was subdued to a separate deadline regime, the analysed ROP evaluations showed that for evaluators a consultancy task remained still a challenge. The consulting function could be manifested by formulating expert proposals for changes on a rolling basis (working recommendations for the purposes of the programming process), relating to flaws and failures discovered in a programme under evaluation. The most difficult task is to recommend modifications to the assumptions of the programme's theory of change, because it requires a great deal of expertise. A lot of evaluation teams which work at the regional level are limited in that they do not have such knowledge (i.e. there is not enough determination or resources to hire adequate experts).

Thus, it has to be concluded that Polish evaluation at regional level features relatively rare examples of reliable restructuring of programme logic at the level of preparing public intervention. In many cases, where programme logic is indeed formulated at the stage of developing the intervention concept, it is very often modified during the decision making process (especially when the project selection system and criteria were formulated) and finally it may be watered down. As a consequence, this makes it difficult to carry out the subsequent evaluation appropriately, as the main objective of evaluation is to verify the efficiency of mechanisms leading to the achievement of results.

Apart from a set of ROP *ex ante* evaluations, the latest final reports available on the website of the Ministry of Development also include studies which fulfil the material requirements of TBE. As far as the development of evaluation concept, and fulfilment of the criteria for identification and description of base theories of the intervention evaluated are concerned, one can point e.g. to an "Evaluation of the application

of equal opportunities principle at all levels of implementing RPO in Podlaskie Voivodeship 2007-2013²²⁸.

Nevertheless one has to underline that the new approach to evaluation remains a big challenge for all regional evaluation units. Relatively low budgets allocated by the regions for evaluation are one of limitations in getting high quality evaluation by means of i.a. TBE. Those voivodeships which have limited funds under ROP 2014-2020 technical assistance, also have relatively small budgets for evaluations. Thus, it is difficult for those regions to plan and then carry out evaluation which takes account of „methodologically rigorous“ EC guidelines and domestic recommendations.

Local administration organizational culture is another very important aspect which matters for evaluation quality in the Polish regions. Organizational problems within the institutions (e.g. people dealing with evaluation also must carry out numerous other assignments) make it very difficult to introduce innovation (the use of evaluation as a support tool in developing regional public policies still remains among innovations). Major problems are encountered by those regional evaluation units, which have experienced significant staff fluctuations or where the number of evaluators have been substantially decreased. One other element which affects the quality of TBE evaluation outcomes is a limited collaboration between the evaluation unit and departments responsible for public interventions development, as TBE requires strict collaboration of an evaluation unit and the evaluator with people responsible for programme design, consideration of context specifics, and looking for evidence that causal linkages have occurred or there is a real chance that they will occur.

Methodology of evaluation is also a major challenge, as even though there is a lot of information and training available from the National Evaluation unit, PARP, Polish Evaluation Society and tertiary education institutions which develop evaluation potential in Poland (e.g. Centre for European Regional and Local Studies EUROREG, Jagiellonian University, Koźmiński University), the TBE approach is relatively rarely analysed in Polish literature (this publication has hitherto been an exception) and included in academic curricula concerning evaluation studies. As a result, TBE is relatively rarely practised at regional level, and if is used (e.g. for *ex ante* evaluation of ROP 2014-2020 as a result of a top-down recommendation), its scope and methodological approach and real support in the programming period leave a lot to be desired.

Systematic and cross-cutting evaluation studies, which take account not only of the degree to which regional problems have been solved, but also of filling the gaps in terms of knowledge about context and its impact on change mechanisms, may shed more light to the critical regional development factors and intervention parameters in

²²⁸ https://www.wrotapodlasia.pl/pl/fundusze_europejskie/fundusze_europejskie/ocena-stosowania-zasady-rownosci-szans-na-wszystkich-etapach-realizacji-rpowp-na-lata-2007-2013.html (10.05.2017r.)

the long-term, resulting in more plausible matching of programmes to local conditions. Identification of effective and ineffective interventions, including a detailed explanation of the causes for successes and failures may help eliminate the latter from the pool of actions taken by regional administrations²²⁹. In this context, however, the method of carrying out evaluation is really important. For *ex ante* evaluation, when TBE is applied, participative approach is especially useful. Thanks to participation, i.e. inclusion of entities and individuals planning public policies in the process of reconstructing the intervention logic, evaluator can better understand the local context of the intervention, assumptions formulated by particular stakeholders and can explain causal relationships fitted into the programme on the rolling basis. As a result, it is possible to determine which elements of intervention require justification or modification (missing or weak elements of intervention logic), which may translate into improving the plans, methodology and effects of intervention as well as evaluation itself. Moreover, the application of participative approach shortens the period of intervention programming due to the possibility of simultaneous carrying out of *ex ante* evaluation and including recommended modifications resulting from that evaluation in the subsequent areas of intervention logic design.

Recapitulation

In the 2007-2013 financial perspective, the foundations for EU cohesion policy evaluation system at regional level were laid in Poland. A question remains to be answered, however, whether in light of new challenges and assignments faced by voivodeships (i.a. TBE), this potential is sufficient. EC requirements were to result in increased application of TBE. It seems, however, that adjusting to those requirements is going to be quite difficult.

While being aware of what a huge challenge it is for the regions to adjust to the new evaluation needs, and knowing that the process of building evaluation potential in Poland has reached a more advanced level of development, one has to remember that participative evaluation should be the greatest advantage of the whole process. Continuous collaboration of evaluator with a contracting party, including those who are/were responsible for given intervention programming, may be the biggest value added of the evaluation - especially when it is conducted based on TBE approach.

²²⁹ More information on the three basic components forming the assumptions for public interventions, namely base theories, theories of change and implementation theories can be found in: K. Olejniczak, „Ewaluacja w sektorze publicznym”, in: „Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka”. Collective work, Marshal Office of Opolskie Voivodeship, Opole 2009, p. 50.

About the authors

Katarzyna Lotko-Czech

PhD in Economic Sciences, graduate of the University of Opole and postgraduate in the field of the evaluation of social programmes at the University of Warsaw. In the years 1999-2014, she worked at the Marshal's Office of the Opolskie Voivodeship, where, in 2007 she created from scratch the Evaluation Unit responsible for coordinating the evaluation process of operational programmes at the regional level. Since 2015, she has been an expert of IDEA of Development Foundation, where she designs and carries out evaluation studies, in particular for public administration. She is an author of scientific articles, *inter alia*, on regional development, European Union Cohesion Policy, public finance. She is a lecturer at scientific conferences and seminars concerning the regional policy and building the evaluation culture at the regional level. She gives lectures on the regional policy at the Wrocław School of Banking (Faculty of Economics in Opole). She is a member of the Polish Economic Society, co-founder and CEO of CG2.

Anna Bruska

PhD in Economic Sciences, employed in the Chair of Logistics and Marketing at the Faculty of Economics of the University of Opole. She is a certified tutor of *Collegium Wratislaviense* and a business coach. She cooperates with local government units, *inter alia*, as an external expert in conducting EU-financed evaluation studies and in developing evaluation plans. Her scientific interests include, *inter alia*, cognitive and methodological determinants of the quality of research, including evaluation studies.

Bibliography:

1. Bachtler J., „Ewaluacja regionalnej polityki w Europie: kultura, zaangażowanie i potencjał”, in: *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, by Olejniczak K., Kozak M., Lendzion B., Academic and Professional Press, Kozminski University, Warsaw 2008.
2. Bober J., „Formalnoprawne podstawy Unii Europejskiej w zakresie ewaluacji”, in: *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Mazur S., University of Economy in Krakow, the Małopolska School of Public Administration, Krakow 2007.
3. Górniak J., „Szacowanie przyczynowego wpływu interwencji: funkcje, walory, ograniczenia oraz relacje z innymi podejściami ewaluacji”, in: *Ewaluacja RPO WK-P. Nowe wyzwania dla ewaluacji programów operacyjnych w perspektywie 2014-2020*, The Marshal Office of the Kujawsko-Pomorskie Voivodeship, Toruń 2016.
4. Markiewicz A., Patrick I., „Developing Monitoring and Evaluation Frameworks”, Sage Publications Inc. 2016.
5. Olejniczak K., „Mechanizmy wykorzystania ewaluacji. Studium ewaluacji Średniookresowych INTERREG III”, Center for European Regional and Local Studies of the University of Warsaw, SCHOLAR Academic Publishing, Warsaw 2008.
6. Olejniczak K., „Teoretyczne podstawy ewaluacji ex-post”, in: *Ewaluacja ex-post. Teoria i praktyka badawcza*, Polish Agency for Enterprise Development, Warsaw 2007.
7. Stame N., „Theory-based Evaluation and Types of Complexity”, *Evaluation*, 2004, vol. 10(1), ss. 58-76.
8. Szałaj M., „Budowa potencjału ewaluacyjnego”, in: *Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka*, Group work. The Marshal Office of the Opolskie Voivodeship, Opole 2009.
9. Żuber P., Bienias S., „System ewaluacji w Polsce – dotychczasowe doświadczenia i wyzwania na przyszłość”, in: *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, by Olejniczak K., Kozak M., Lendzion B., Academic and Professional Press, Kozminski University, Warsaw 2008.

Other:

1. *Guidelines No. 6 concerning evaluation of operational programmes for the period 2007-2013*, Ministry of Regional Development, Warsaw, 30 May 2007.
2. *Rocznik Ewaluacji 2013*, Ministry of Development, Warsaw 2014.
3. *Rocznik Ewaluacji 2015*, Ministry of Development, Warsaw 2014.
4. *Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund*, European Union, 2015, http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf, (access on 19.03.2017).
5. *Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020*, Minister of Infrastructure and Development, Warsaw, 2015, http://www.funduszeuropejskie.gov.pl/media/9318/wytyczne_090915_final.pdf, (access on 19.03.2017).

6. Regulation (EU) No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006.
7. *Recommendations concerning ex-ante evaluation of operational programmes for the period 2014-2020*, Department of Structural Policy Coordination, Ministry of Regional Development, Warsaw, November 2012.

