

# **Ewaluacja oparta na teorii**

w złożonym otoczeniu  
społeczno-ekonomicznym

## **Polska Agencja Rozwoju Przedsiębiorczości**

PARP jest agencją rządową, która od 2000 r. wspiera rozwój małych i średnich firm w Polsce. Aktywność Agencji jest skoncentrowana na wspieraniu powstawania i rozwoju nowych podmiotów gospodarczych, wzmocnieniu pozycji konkurencyjnej firm w oparciu o innowacyjność i nowoczesne technologie, podnoszeniu kwalifikacji kadr przedsiębiorstw, kształtowaniu przyjaznego otoczenia biznesowego oraz tworzeniu warunków do prowadzenia działalności gospodarczej.

Oferując wsparcie finansowe i szkoleniowo-doradcze przedsiębiorcom (a także: instytucjom otoczenia biznesu, jednostkom samorządu terytorialnego, państwowym jednostkom budżetowym, uczelniom), PARP korzysta ze środków budżetu państwa oraz funduszy europejskich. W latach 2014-2020 Agencja jest odpowiedzialna za realizację działań w ramach trzech programów operacyjnych: Inteligentny Rozwój, Polska Wschodnia oraz Wiedza Edukacja Rozwój.

### **Ewaluacje PARP**

PARP posiada unikalne doświadczenie nie tylko w przekazywaniu pomocy przedsiębiorcom. Istotną część aktywności PARP stanowi działalność badawczo-analityczna.

Agencja prowadzi badania ewaluacyjne programów wsparcia konkurencyjności i innowacyjności gospodarki, podnoszenia kwalifikacji kadr sektora prywatnego oraz rozwoju infrastruktury społeczno-gospodarczej. Ich celem jest przygotowywanie wniosków i rekomendacji dla polityki sprzyjającej trwałemu i zrównoważonemu rozwojowi polskiej gospodarki, w szczególności pobudzania przedsiębiorczości i innowacyjności małych i średnich firm oraz ponoszenia efektywności otoczenia biznesu. Realizowane przez PARP ewaluacje dostarczają wiedzy niezbędnej w procesach programowania instrumentów wsparcia i są także istotnym narzędziem weryfikacji skuteczności udzielanej pomocy.

PARP jako jedna z najbardziej doświadczonych w tym zakresie krajowych instytucji, we współpracy z Ministerstwem właściwym ds. rozwoju regionalnego, od wielu lat rozwija również kulturę ewaluacji w Polsce oraz realizuje i upowszechnia ideę polityk publicznych opartych na dowodach. Od 2005 r. Agencja jest współorganizatorem corocznej międzynarodowej konferencji ewaluacyjnej oraz przygotowuje publikacje, poprzez które promuje najnowsze koncepcje teoretyczne i rozwiązania, stosowane w ewaluacji interwencji publicznych (w szczególności w obszarze przedsiębiorczości i innowacyjności) oraz dzieli się doświadczeniem w stosowaniu innowacyjnych metod ewaluacji.

W ramach aktywności badawczo-analitycznej, PARP we współpracy z Partnerami prowadzi również monitoring i ewaluacje strategii inteligentnych specjalizacji w Polsce oraz tworzy Centrum analiz i pilotaży nowych instrumentów – inno\_LAB. Głównym zadaniem inno\_LAB jest tworzenie i testowanie nowych narzędzi wspierania innowacyjności oraz animowanie współpracy i zwiększenie potencjału uczestników Narodowego Systemu Innowacji.

### **Zapraszamy do kontaktu:**

[ewaluacja@parp.gov.pl](mailto:ewaluacja@parp.gov.pl)  
[analizy@parp.gov.pl](mailto:analizy@parp.gov.pl)

[badania.parp.gov.pl](http://badania.parp.gov.pl)

**Ewaluacja  
oparta na teorii**  
w złożonym otoczeniu  
społeczno-ekonomicznym

## Ewaluacja oparta na teorii w złożonym otoczeniu społeczno-ekonomicznym

Redakcja:

Jacek Pokorski, Zuzanna Popis, Teresa Wyszyńska  
(Polska Agencja Rozwoju Przedsiębiorczości)  
Katarzyna Hermann-Pawłowska (Fundacja IDEA Rozwoju)

Autorzy:

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

Poglądy autorów rozdziałów zawartych w publikacji są wyrazem ich własnych opinii i nie odzwierciedlają oficjalnego stanowiska PARP.

Niniejsze wydanie stanowi przekład na język polski publikacji PARP pt. *Theory-based evaluation in complex environments*.

Copyright by Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2017

ISBN 978-83-7633-3335-9

Wydanie I

Egzemplarz bezpłatny

Zakup wybranych utworów autorskich współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Inteligentny Rozwój, 2014-2020.



Dostarczenie utworów autorskich, redakcja techniczna i językowa, szata graficzna i projekt okładki: Fundacja IDEA Rozwoju, we współpracy z CG2 Sp. z o.o.

Tłumaczenie (rozdziały I, IV, V):  
GTC AMG Sp. z o.o.

Przygotowanie do druku, druk i oprawa:  
*Grafpol* Agnieszka Blicharz-Krupińska

# Spis treści

Wprowadzenie .....	5
I. Michael Quinn Patton - <i>Ewaluacja rozwojowa</i> .....	7
II. Katarzyna Hermann-Pawłowska, Paulina Skórska - <i>Przegląd systematyczny koncepcji ewaluacji opartej na teorii</i> .....	23
III. Marcin Kocór, Barbara Worek - <i>Porównawcza analiza jakościowa w ewaluacji wpływu opartej na teorii</i> .....	37
IV. Emanuela Sirtori, Elena Vallino, Silvia Vignetti - <i>Badanie teorii interwencji przy użyciu analizy sieci bayesowskiej: wyniki pilotażowego działania dotyczącego wsparcia MŚP</i> .....	57
V. András Kaszap, Jan-Philipp Kramer, Rafał Trzciński - <i>Zastosowanie analizy kontrybucji jako podejścia ewaluacji opartej na teorii w ewaluacji ex post wsparcia dla dużych przedsiębiorstw w ramach programów Polityki Spójności UE 2007-2013</i> .....	75
VI. Maciej Koniewski, Seweryn Krupnik - <i>Jak lepiej zrozumieć wyniki analiz kontrfaktycznych? Dobór beneficjentów do badań jakościowych, wykorzystujący dopasowanie statystyczne, na przykładzie ewaluacji wpływu Programu Operacyjnego Innowacyjna Gospodarka 2007-2013</i> .....	99
VII. Dominika Wojtowicz, Łukasz Widła-Domaradzki - <i>Stratified Propensity Score Matching – quasi-eksperymentalny schemat badawczy, wykorzystywany na potrzeby ewaluacji wpływu opartej na teorii, w sytuacji braku zmiennej zależnej</i> .....	117
VIII. Maciej Gajewski, Andrzej Regulski - <i>Ewaluacja wspierana teorią w badaniach systemu wyboru projektów innowacyjnych w programach operacyjnych na lata 2014-2020</i> .....	139
IX. Anna Bruska, Katarzyna Lotko-Czech - <i>Doświadczenia w wykorzystaniu ewaluacji opartej na teorii na poziomie regionalnym</i> .....	157



# Wprowadzenie

Z przyjemnością przekazujemy Państwu publikację poświęconą teorii i praktyce ewaluacji interwencji publicznych - „Ewaluacja oparta na teorii w złożonym otoczeniu społeczno-ekonomicznym”.

Inspiracją dla tematyki publikacji były wyzwania związane z prowadzeniem ewaluacji inicjatyw publicznych w złożonym i zmieniającym się otoczeniu. Ich przykładem są m.in. wyzwania realizacji programów proinnowacyjnych, w obliczu różnego typu kryzysów gospodarczych i zmian w postawach konsumentów, w modelach biznesowych czy funkcjonowaniu organizacji. Przy tego typu uwarunkowaniach zarówno założenia programu, jak i logika prowadzonych prac ewaluacyjnych, powinny być na bieżąco dostosowywane, co doskonale opisuje prof. Michael Quinn Patton w rozdziale „Ewaluacja rozwojowa” otwierającym publikację.

Rozwój działań ewaluacyjnych stymuluje również Komisja Europejska. Wskazuje ona na konieczność, w obecnych realiach rynkowych, bardziej rygorystycznego podejścia metodologicznego w ewaluacji (w tym także podejścia innowacyjnego i eksperymentalnego). Pozwoliłoby to na wyjaśnienie faktycznych mechanizmów interwencji i uzyskanie wiarygodnych informacji o rzeczywistych efektach polityk. Jednym z rekomendowanych do ewaluacji polityki spójności Unii Europejskiej (UE) na lata 2014-2020 (oraz ewaluacji *ex post* wsparcia z perspektywy 2007-2013) jest podejście oparte na teorii (*theory-based evaluation, TBE*). Jego stosowanie powinno umożliwić wyjaśnienie zmian zachodzących w wyniku interwencji, ale także przyczynić się do poprawy jakości i użyteczności samej ewaluacji.

Ten kierunek rozwoju ewaluacji pociąga za sobą konieczność dostosowania dotychczasowych zasad realizacji programów poprzez silniejsze powiązanie procesów ewaluacji z programowaniem i wdrażaniem interwencji, przez co wymusza zmiany o charakterze kulturowym w sektorze publicznym. Nawiązuje do tego tekst Macieja Gajewskiego i Andrzeja Regulskiego prezentujący ewaluację systemów wyboru projektów innowacyjnych. Pozostaje to wyzwaniem na poziomie działań ogólnokrajowych, ale zwłaszcza w ewaluacji programów regionalnych (dobrze ilustruje to tekst Anny Bruskiej i Katarzyny Lotko-Czech o doświadczeniach z wykorzystania TBE na poziomie regionalnym). Warto podkreślić, że podejście oparte na teorii zmiany (*theory of change*) nie tylko jest użyteczne dla samej ewaluacji, ale przede wszystkim może być pomocne we wdrażaniu polityk rozwoju, a zwłaszcza ich ciągłego dostrajania do zmieniającego się kontekstu.

Argumentem za obraną tematyką publikacji była nisza w obszarze krajowych opracowań w tym zakresie. Zarówno w obszarze prac o charakterze ideowym, jak i imple-

mentacyjnym nt. studiów konkretnych przypadków, wskazujących czym się kierować i jak w praktyce stosować podejście wspierane teorią.

Dla porównania, w przypadku komplementarnego do ewaluacji opartej na teorii, podejścia kontrfaktycznego (*counterfactual impact evaluation, CIE*) – nie mniej intensywnie promowanego w ewaluacji wpływu funduszy unijnych – dorobek literaturowy (zwłaszcza opracowania eksperckie PARP) można uznać za stosunkowo pokaźny.

W publikacji dokonano uporządkowania licznych koncepcji ewaluacji opartych na teorii (tekst Katarzyny Hermann-Pawłowskiej i Pauliny Skórskiej), scharakteryzowano przykładową metodologię, która doskonale adresuje idee TBE (tekst Marcina Kocóra i Barbary Worek, dotyczący techniki QCA tj. komparatywnej analizy jakościowej, pomocnej w odkrywaniu mechanizmów przyczynowych interwencji), a następnie pokazane zostały konkretne realizacje ewaluacji opartych na teorii lub ewaluacji łączących podejście CIE i TBE (*theory-based impact evaluation, TBIE*).

W publikacji skupiliśmy się na pokazaniu doświadczeń krajowych ekspertów z badań TBIE realizowanych dla PARP (tekst Macieja Koniewskiego i Seweryna Krupnika, „Jak lepiej zrozumieć wyniki analiz kontrfaktycznych? Dobór beneficjentów do badań jakościowych, wykorzystujący dopasowanie statystyczne...” oraz Łukasza Widła-Domadzkiego i Dominiki Wojtowicz, „Stratified Propensity Score Matching - quasi-eksperymentalny schemat badawczy wykorzystywany w sytuacji braku zmiennej zależnej”). Dopełniają je idee i dokonania zagraniczne, pośrednio również dotyczące PARP (dwa teksty odwołujące się do przykładów ewaluacji *ex post* wsparcia polityki spójności UE na lata 2007-2013 - dla małych i średnich przedsiębiorstw oraz dla dużych podmiotów – odpowiednio tekst Emanuela Sirtori, Eleny Vallino i Silvii Vignetti, przedstawiający zastosowanie techniki BNA w testowaniu teorii interwencji, oraz rozdział przygotowany przez Andrása Kaszapa, Jan-Philippa Kramera i Rafała Trzcińskiego, prezentujący m.in. zastosowanie analizy kontrybucji w TBE).

Liczymy, że pozwoli to w większym stopniu upowszechnić wartościowe doświadczenia PARP w tym zakresie - z jednej strony wynikające ze stosowania innowacyjnych podejść w ewaluacji, z drugiej dotyczące proinnowacyjnego obszaru przedmiotowego, tj. ewaluacji różnego typu wsparcia innowacyjności i przedsiębiorczości w ramach polityki spójności UE.

Zachęcając Państwa do lektury przygotowanej publikacji, i jednocześnie życząc skutecznego wprowadzania w życie TBE lub korzystania z wyników tego typu ewaluacji, chcielibyśmy podziękować Fundacji IDEA Rozwoju, która wspierała nas w tym przedsięwzięciu oraz autorom poszczególnych rozdziałów publikacji. Mamy nadzieję, że doświadczenia i konkluzje, które w nich Państwo znajdziecie, będą inspirujące do udoskonalania realizowanych działań w obszarze ewaluacji i polityk rozwoju.

**Zespół redakcyjny**



## Ewaluacja rozwojowa

### Wprowadzenie

„Ewaluacja rozwojowa” (*developmental evaluation, DE*) wspiera „rozwój” innowacyjnych interwencji, aby ukierunkować ich dostosowanie do wyłaniających się i dynamicznych realiów złożonego otoczenia<sup>1</sup>. Ewaluacja rozwojowa jest „skoncentrowana na wykorzystaniu” (*utilization-focused*)<sup>2</sup> w tym sensie, że skupia się na określonym, zamierzonym użyciu – rozwoju – na rzecz konkretnych przewidywanych użytkowników: innowatorów społecznych, dostosowujących swoje interwencje do dynamicznego, złożonego otoczenia (*complex environments*).

Ewaluacja na rzecz równości szans i wspierania praw człowieka, jako części osiągniętych rezultatów w zakresie rozwoju, często występuje w złożonych systemach adaptacyjnych. Złożony system cechuje się wysoką liczbą wzajemnie oddziałujących na siebie i współzależnych elementów, bez centralnej kontroli. Złożone otoczenie w przypadku interwencji i innowacji społecznych to takie, w którym nie jest pewne, jakie działania należy podjąć w celu rozwiązania problemów a kluczowi interesariusze nie zgadzają się w kwestii sposobu postępowania. Sytuacja ta jest typowa w przypadku wspierania praw człowieka. Rozwiązania, które sprawdziły się w danym miejscu, mogą nie być przydatne w innym. Kontekst ma znaczenie. Różnice kulturowe, polityczne, różnice zasobów, potencjału i różnice historyczne wpływają na to, w jaki sposób funkcjonują inicjatywy wspierające rozwój oraz na to, jak te inicjatywy kierują uwagę na kwestie równości szans i praw człowieka. W takich sytuacjach, wspartych myśleniem systemowym (*systems thinking*) i wrażliwością na złożoną, nieliniową dynamikę, ewaluacja rozwojowa sprzyja wyższej efektywności interwencji, innowacjom społecznym, adaptacyjnemu zarządzaniu i ciągłemu procesowi uczenia się.

„Ewaluator rozwojowy” (*developmental evaluator*) nierzadko jest częścią „rozwojowego zespołu”, którego członkowie współpracują ze sobą w celu konceptualizacji, projektowania i testowania nowych podejść w długofalowym, ciągłym procesie stałego rozwoju, adaptacji i eksperymentowania, mocno uwrażliwionym na niezamierzone rezultaty i skutki uboczne. Podstawową rolą ewaluatora w takim zespole jest stymulowanie dyskusji ewaluacyjnymi pytaniami, sposobem myślenia i danymi oraz facylitacja systematycznej refleksji opartej na danych, jak również podejmowanie w procesie rozwojowym decyzji w czasie rzeczywistym.

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

<sup>2</sup> *Ibid.*

## Ulepszenia kontra rozwój

Istnieje wiele podejść do ewaluacji. Każde z nich, także ewaluacja rozwojowa, spełnia konkretny cel i dodaje określony rodzaj wartości. Jak wspomniano powyżej, DE okazała się szczególnie adekwatna i atrakcyjna dla zainteresowanych zmianą systemową i innowacjami społecznymi w złożonych, dynamicznych systemach. Są to systemy, w których ludzie próbują doprowadzić do istotnych przemian społecznych poprzez zwalczanie ubóstwa, bezdomności, nierówności, naruszania praw człowieka, przemocy w społecznościach i w rodzinie, a także pomóc osobom dotkniętym AIDS, wysokim stopniem niepełnosprawności oraz ofiarom klęsk żywiołowych i wojen. Wiele z tych interwencji i inicjatyw na rzecz zmian systemowych, wzmacnianych jest przez intensywne zaangażowanie we wspieranie praw człowieka i równości szans.

Wraz z moimi kanadyjskimi koleżankami, Frances Westley i Brendą Zimmerman, zajmuję się badaniem innowacji społecznych, które odniosły sukces. W książce zatytułowanej *Getting to Maybe: How the World Is Changed*<sup>3</sup> prezentujemy wyniki naszych badań. Bycie „agentem zmiany” wymaga odważnego myślenia i śmiałej wizji. Teoria złożoności (*complexity theory*) pokazuje, że wielkie zmiany mogą wyłonić się z niewielkich działań. Zakłada to wiarę w to, co możliwe, a nawet w to, co „niemożliwe”. Ponadto, znaczące innowacje społeczne nie wiodą prostą, linearną drogą zmian. Napotykać wzloty i upadki, *roller coaster* pędzi przez kaskady dynamicznych interakcji, nieoczekiwanych i nieprzewidywanych rozbieżności, punktów przełomowych i zwrotów masy krytycznej, a gdy zmiana systemowa napotyka na opór, sprawy często przybierają zły obrót, zanim ulegną poprawie.

Tradycyjne podejścia do ewaluacji nie są dobrze przystosowane do takich wstrząsów. Tradycyjna ewaluacja ma na celu kontrolę i przewidywanie, uporządkowanie chaosu poprzez staranne określanie i mierzenie ustalonego sposobu wdrażania oraz osiągnięcia uprzednio zdefiniowanych priorytetowych rezultatów. Natomiast ewaluacja rozwojowa akceptuje zawirowania i niepewność jako sposób zorganizowania świata, jako sposób kształtowania się innowacji społecznych, w obliczu złożoności (*complexity*). Ewaluacja rozwojowa dostosowuje się do realiów złożonej nieliniowej dynamiki, nie próbując narzucać porządku i pewności w nieuporządkowanym i niepewnym świecie. DR dokonuje tego poprzez śledzenie i dokumentowanie nowych i dynamicznych adaptacji procesu wdrażania oraz uzyskanych rezultatów.

Wiele z osób pracujących na rzecz wspierania praw człowieka mówi mi, że miały do czynienia z metodami ewaluacji zupełnie nieprzystosowanymi do natury realizowanych przez nich inicjatyw. Na przykład, identyfikowanie przejrzystych, konkretnych i możliwych do zmierzenia rezultatów na samym początku innowacyjnego projektu może być nie tylko trudne, ale też bezproduktywne. W warunkach dużej niepewności,

---

<sup>3</sup> F. Westley, B. Zimmerman, M. Q. Patton, *Getting to Maybe: How the World Is Changed*, Random House, Kanada 2006.

rezultaty mogą wyłonić się dzięki określönemu zaangażowaniu, raczej jako część procesu na rzecz zmiany. Tak zwane „cele SMART”<sup>4</sup>, przyjęte zbyt wcześnie, nie są „inteligentne” i w istocie mogą zaszkodzić, ograniczając reaktywność i możliwość adaptacji. Ewaluację rozwojową opracowano tak, aby była zgodna z procesami rozwojowymi, wyłaniającymi się, innowacyjnymi i transformatywnymi, oraz umacniała je.

## **Ewaluacja rozwojowa a teoria złożoności**

Złożoność jako pojęcie jest rozległą mozaiką łączącą w sobie wiele elementów odnoszących się do innowacji i ewaluacji: nieliniowość, wyłanianie się (*emergence*), systemy dynamiczne, adaptacyjność, niepewność oraz procesy koewolucyjne<sup>5</sup>. Podobnie ewaluacja rozwojowa, skupiająca się na wrażliwości sytuacyjnej, reaktywności i adaptacji, jest podejściem do ewaluacji szczególnie adekwatnym w przypadku sytuacji o wysokim stopniu niepewności, w których to, co może się pojawić i w istocie pojawia się, jest stosunkowo nieprzewidywalne i niemożliwe do kontrolowania. Ewaluacja rozwojowa śledzi i stara się zrozumieć to, co wyłania się w warunkach złożoności, dokumentując i interpretując dynamikę, interakcje i współzależności, jakie pojawiają się w procesach innowacji i zmian systemowych.

## **Złożone systemy mające zdolność adaptacji**

Opracowania dotyczące złożoności są pełne metafor, które próbują uczynić złożone zjawiska zrozumiałymi dla ludzkiego mózgu, wykazującego wrodzoną potrzebę porządku, znaczeń, schematów, nadawania sensu i kontroli, niezmiennie podtrzymującego złudzenie, że „wiemy, co jest grane”. Często nie wiemy. Jednak nasze udawanie, że wiemy jest pocieszające i czasem niezbędne do podjęcia jakiegokolwiek działania. Dlatego teoretycy złożoności mówią o uderzeniach motyli skrzydeł, które zmieniają układy pogodowe i wywołują huragany; o indywidualnych formacjach pleśni, które w sposób zadziwiający organizują się samoczynnie w organiczne całości; o koloniach mrówek, których szalenie usługiwanie królowej hipnotyzuje nas swoją zbiorową inteligencją; o lawinach, które przekształcają górskie ekosystemy; o bakteriach „znających” systemy, do których przynależą, pomimo braku samoświadomości; oraz o „czarnych łabędziach”, które pojawiają się nagle i w nieprzewidywalny sposób, żeby zmienić świat. Nauka o złożoności daje wgląd w miliardy interakcji na światowym rynku giełdowym; rozprzestrzenianie się choroby na świecie; niestabilne układy pogodowe; ewolucję gatunków; przemiany środowiskowe na wielką skalę; oraz gromadzenie się stad migrujących ptaków. Teoretycy złożoności objaśniają wzrost i upadek cywilizacji a także wzlot i wygasanie romantycznego zauroczenia. Dziedzina jest, jak widać, pojemna. Może ona, a nawet powinna, obejmować kwestię rozwoju i upadku ewaluacji.

---

<sup>4</sup> Cele SMART: konkretne, dające się zmierzyć, osiągalne, realistyczne, terminowe (specific, measurable, attainable, realistic, timely)

<sup>5</sup> M. Q. Patton, *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, Guilford, Nowy Jork 2011 <http://www.guilford.com/excerpts/patton.pdf>

## Jak postępować w nieoczekiwanych okolicznościach

W odniesieniu do ewaluacji wiele mówi się o poszukiwaniu nieprzewidzianych konsekwencji i ocenie skutków ubocznych. W rzeczywistości są to symboliczne elementy, typowe dla projektów ewaluacji, niewłaściwie ujęte w budżecie, którym rzadko poświęca się wiele czasu i uwagi, ponieważ dominującym przedmiotem zainteresowania jest mierzenie osiągnięcia zamierzonych rezultatów i śledzenie z góry określonych wskaźników wydajności. Tymczasem wykrywanie nieprzewidzianych konsekwencji odbywa się poprzez wyjście do realnego świata, prowadzenie prac w terenie, zadawanie otwartych pytań, rozmowy z uczestnikami oraz obserwowanie, co się dzieje wraz z rozwojem interwencji i innowacji. Uważam, że ewaluatorzy zazwyczaj podchodzą do tego, co nieoczekiwane i nieprzewidziane w sposób zwyczajny, nie nadając wysokiego priorytetu, jak gdyby w zasadzie myśleli: „poszukamy nieprzewidzianych konsekwencji i wyłaniających się rezultatów po zrobieniu całej reszty, jeśli będziemy mieć czas i środki”. Jednak oczywiście czas i środki rzadko udaje się znaleźć. Niemniej, prawdopodobieństwo wystąpienia nieoczekiwanych skutków staje się bardzo wysokie w warunkach złożoności, dlatego w przypadku ewaluatorów rozwojowych, przygotowanie się na to co nieoczekiwane, ma fundamentalne znaczenie dla ich bieżącej pracy.

## Ewaluacja rozwojowa a uczenie się

Ewaluacja rozwojowa wspiera uczenie się w celu nadawania świadomego wymiaru działaniom, które wiele zmieniają. Często oznacza to dokonywanie zmian systemowych, co z kolei zakłada wyjście poza powierzchowne uczenie się na rzecz głębokiego rozumienia tego, co się dzieje w danym systemie. Społeczni innowatorzy i przedsiębiorcy społeczni, szczególnie pracujący nad zagadnieniami praw człowieka i równości szans, zazwyczaj starają się doprowadzić do fundamentalnych zmian w systemach, żeby zmienić świat. W tym celu muszą zrozumieć, w jaki sposób funkcjonuje system, który chcą zmienić i przeprowadzić zmiany, które zmieniają system jako taki poprzez wyjście poza rozwiązania tymczasowe i powierzchowne. Wymaga to „dwupętłowego uczenia się” (*double loop learning*).

Od dziesięcioleci powtarzano bez końca trzy historie: jedną o ambulansach czekających na dole klifu zamiast budowania ogrodzenia na jego szczycie; kolejną o licznych zwłokach płynących rzeką i jedynie tworzeniu coraz bardziej imponujących sposobów ich wyławiania; oraz historię o podarowaniu komuś ryby w opozycji do wartości nauki łowienia. Dokonując przeglądu tych historii, utytułowana australijska uczona, badaczka i praktyk, Yolande Wadsworth<sup>6</sup> skomentowała je twierdząc, że przypominają one o naszej powtarzającej się tendencji do wybierania krótkoterminowych, szybkich rozwiązań (zamiast analizy, zrozumienia i podjęcia działania w celu zmiany funkcjonowania systemu), które właśnie tworzą problem, który próbujemy rozwiązać. „Dwupętłowe

---

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

uczenie się” obejmuje rozwiązania systemowe i znajduje wsparcie w ewaluacji dostosowanej do poszukiwania wyjaśnień systemowych i oferującej systemowy wgląd (*systemic insights*).

W ramach „jednopętlowego uczenia się” osoby modyfikują swoje działania, ewaluując różnicę pomiędzy rezultatami pożądanymi a rzeczywistymi, i wprowadzają zmiany w celu zwiększenia osiągania pożądaných wyników. W istocie proces wykrywania i naprawiania problemu, podobnie jak ewaluacja formatywna, jest „jednopętlowym uczeniem się”. W ramach „dwupętlowego uczenia się” osoby w nie zaangażowane wykraczają poza pojedynczą „pętlę” identyfikacji problemu i znalezienia rozwiązania, przechodząc do drugiej „pętli”, zakładającej kwestionowanie założeń, polityk, praktyk, wartości i dynamiki systemu, które doprowadziły do powstania problemu w pierwszej kolejności, a następnie ingerują w funkcjonowanie i relacje w obrębie systemu leżące u podstaw danego problemu. Wprowadzanie zmian w celu poprawy natychmiastowych rezultatów jest „jednopętlowym uczeniem się”; wprowadzanie zmian do systemu w celu zapobiegania powstawaniu problemu lub wbudowanie rozwiązania w poddany zmianom system - oznacza „dwupętlowe uczenie się”. „Trzypętlowe uczenie się” zakłada naukę sposobu uczenia się i jest zawarte w procesach ewaluacji rozwojowej.

## **Ewaluacja rozwojowa w kontekście ewaluacji rozwoju**

Ewaluację rozwojową łatwo mylimy z ewaluacją rozwoju. Nie są tym samym, chociaż ewaluację rozwojową można wykorzystywać do ewaluacji rozwoju.

„Ewaluacja rozwoju” (*development evaluation*) jest terminem ogólnym określającym ewaluacje prowadzone w krajach rozwijających się, zazwyczaj skupiające się na skuteczności programów i agencji pomocy międzynarodowej. *The Road to Results: Designing and Conducting Development Evaluations*<sup>7</sup> jest przykładem tego gatunku. To książka oparta na niezwykle udanym programie szkoleniowym Banku Światowego - International Programme for Development Evaluation Training (IPDET) - który autorzy książki założyli i którym kierowali, i w oparciu o który napisano tę książkę.

„Ewaluacja rozwojowa”, zgodnie z jej definicją i opisem w Encyklopedii Ewaluacji<sup>8</sup>, ma na celu pomoc w rozwoju innowacji, interwencji, programu lub zmiany systemu. Ewaluator wykorzystuje metody ewaluacyjne w celu facylitacji trwającego programu, projektu, produktu, personelu i/lub rozwoju organizacyjnego. Podstawową funkcją ewaluatora w zespole jest animowanie i objaśnianie dyskusji poprzez wprowadzenie ewaluacyjnych pytań, danych i logiki, oraz wspieranie podejmowania w procesie rozwojowym decyzji opartych na danych.

---

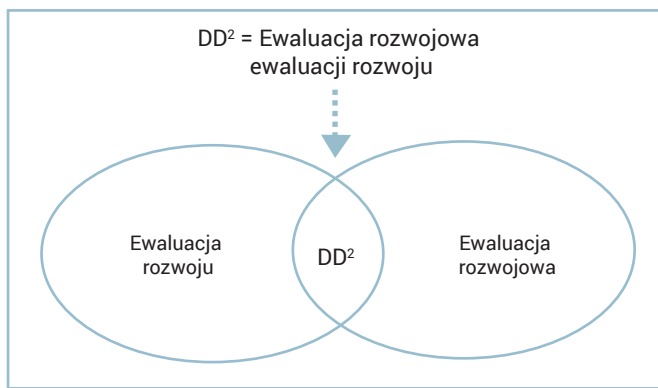
<sup>7</sup> L. G. M. Imas, R. C. Rist, *The Road to Results: Designing and Conducting Development Evaluations*, The World Bank, Washington, D.C 2009.

<sup>8</sup> S. Mathison (pod red), *Developmental evaluation. Encyclopedia of evaluation*, Sage, Thousand Oaks, CA 2005, s. 116.

Ewaluacja skupiająca się na pomocy rozwojowej w krajach rozwijających się, mogłaby wykorzystywać podejście ewaluacji rozwojowej, w szczególności w przypadkach, gdy taka pomoc rozwojowa jest postrzegana jako występująca w warunkach złożoności, z naciskiem na dostosowanie do lokalnego kontekstu. Ewaluację rozwojową można wykorzystywać wszędzie tam, gdzie innowatorzy społeczni angażują się w doprowadzenie do zmian systemowych w warunkach złożoności.

Końcówkę „-owa” w „rozwojowa” łatwo przeoczyć, jednak ma ona kluczowe znaczenie dla rozróżnienia ewaluacji rozwoju i ewaluacji rozwojowej.

### Rysunek 1. Ewaluacja rozwojowa wykorzystywana na potrzeby ewaluacji rozwoju



Kiedy po raz pierwszy nazwałem i opisałem „ewaluację rozwojową” 15 lat temu<sup>9</sup>, ewaluacja rozwoju nie stanowiła odrębnej i widocznej kategorii w praktyce i nauce ewaluacji. Ewaluacje w krajach rozwijających się były, rzecz jasna, prowadzone, jednak możliwy do zidentyfikowania korpus literatury przedmiotu skupiający się na ewaluacji pomocy rozwojowej, nie przyciągał powszechnej uwagi specjalistów. Jedną z najistotniejszych tendencji ostatniej dekady jest szybkie rozpowszechnianie się ewaluacji na całym świecie, w szczególności w krajach rozwijających się, co widać na przykładzie utworzenia International Development Evaluation Association (Międzynarodowe Stowarzyszenie Ewaluacji Rozwoju). Nieporozumienia, co do odrębnych, a czasem nakładających się na siebie, nisz ewaluacji rozwoju i ewaluacji rozwojowej, są już, obawiam się, obecnie częścią złożonego krajobrazu międzynarodowej ewaluacji. Mam nadzieję, że niniejszy rozdział pomoże w objaśnieniu, zarówno wspomnianych różnic jak i obszarów wspólnych.

<sup>9</sup> M. Q. Patton, *Developmental Evaluation. Evaluation Practice* 15, 1994, s. 311-320.

## Przykłady ewaluacji rozwojowej w kontekstach rozwoju

- Współpraca z naukowcami w dziedzinie rolnictwa na rzecz przyjęcia zintegrowanego, systemowego podejścia do „upraw sierocych”, zakładałaby pracę z agronomami, gleboznawcami, hodowcami roślin, specjalistami do spraw nawadniania, pracownikami dziedzin pokrewnych, zdrowia, żywienia, badaczami płci oraz z rolnikami, w celu opracowania koncepcji innowacji w rolnictwie jako złożonego systemu adaptacyjnego oraz określenia bieżących wskaźników interakcji (*real time indicators*) oraz dynamiki w obrębie systemów, kiedy nowe podejście do rolnictwa zaczyna wpływać na wykorzystanie środków produkcji rolnej, technik produkcji, siły roboczej w gospodarstwie rolnym oraz dynamiki rodzin rolniczych. Taki rodzaj holistycznej interwencji zakłada zmiany w sposobie w jaki tradycyjnie odrębni naukowcy w dziedzinie rolnictwa i żywienia pracują z rolnikami, i wpłynęłby na podejmowanie decyzji oraz interakcje w rodzinach rolniczych.
- Interwencja w dziedzinie mikrofinansowania, zbadana przy zastosowaniu ewaluacji rozwojowej, skupiałaby się na zasilaniu kapitałem jako czynnika uruchamiającym dźwignię w złożonym systemie adaptacyjnym. Niosłoby to ze sobą następstwa dla szeregu kalkulacji i decyzji biznesowych; wzajemnych zależności wśród kredytobiorców; relacji z konsumentami; oraz dla finansów rodzinnych i dynamiki interpersonalnej. Obserwowanie i dostosowywanie się do pojawiających się rezultatów, wykraczających poza zwykłe korzystanie z niewielkich funduszy pożyczkowych, byłoby wbudowane w projekt ewaluacji i informację zwrotną w czasie rzeczywistym, wraz z rozwojem systemu mikrokredytów.

## Przykłady ewaluacji rozwojowej w ewaluacjach skoncentrowanych na równości szans

Ewaluacje rozwojowe skupiające się na marginalizowanych i wykluczonych grupach ludności pomagają w dostosowaniu się do szybko zmieniających się warunków. Oto kilka przykładów.

- Osoby żyjące w stanie ubóstwa znajdują się na krawędzi przetrwania. Nagłe zmiany dostępności żywności mogą zmienić sytuację całej grupy ludności, od posiadania środków utrzymania do głodu. Brak bezpieczeństwa żywnościowego może być wynikiem pogody (ostra susza lub powódź), niepokojów politycznych (przerwy w transporcie żywności) i przemian gospodarczych (wzrost cen żywności). Czasem wszystkie trzy czynniki, czyli zakłócenia pogodowe, polityczne i gospodarcze, występują równocześnie, tworząc napędzającą się również pochyłą wiodącą do rosnącej desperacji. Sytuacje takie wymagają dysponowania danymi w czasie rzeczywistym, dotyczącymi tego, co dzieje się z dotkniętymi taką sytuacją ludźmi i na ile oparte na dobrej wierze interwencje faktycznie się sprawdzają.
- Marginalizowane i wykluczone grupy ludności są szczególnie podatne na choroby zakaźne. Na przykład kampanie szczepień przeciwko polio, wymagają dostosowania do konkretnych kontekstów rozwoju. Na obszarach, na których starania na

rzecz wykorzenia polio nie powiodły się, np. w niektórych regionach Nigerii, Pakistanu czy Indii, mogą pojawić się nowe ogniska choroby, które rozprzestrzenia się szybko na obszary, na których - jak sądzono - choroba została zwalczona. Na przykład, ewaluator rozwojowy pomógłby w monitorowaniu doniesień o oporze wobec kampanii szczepień, tak jak w przypadku muzułmańskich matek, które ukrywały dzieci przed osobami wykonującymi szczepienia, z powodu plotek głoszących, że szczepionka jest częścią hinduistycznego spisku mającego na celu sterylizację dzieci. Wykrywanie i prostowanie takich plotek w czasie rzeczywistym, w miarę ich pojawiania się, może ratować życie.

- Kampania na rzecz praw człowieka prowadzona w dowolnym miejscu na świecie może wymagać istotnych dostosowań, jako że uliczne demonstracje wzywające do demokratycznych reform w Tunezji i Egipcie (2011) zmieniają globalny kontekst, w którym podejmowane są inicjatywy na rzecz praw człowieka. Marginalizowane, pozbawione mocnej pozycji i wykluczone grupy ludności mogą stać się bezdomnymi uchodźcami w momencie, gdy zawirowania polityczne przyspieszą i rozprzestrzenia się.
- Reagowanie na katastrofę humanitarną, taką jak trzęsienie ziemi w Haiti (2010), wymaga dysponowania danymi w czasie rzeczywistym, dotyczącymi lokalnych skupisk osób dotkniętych tragedią; przejezdności dróg; zagrożeń stabilności budynków pozostałych na obszarach występowania silnych opadów deszczu; ognisk cholery; obszarów najbardziej potrzebujących żywności, czystej wody i leków; i tak dalej. Starania na rzecz międzynarodowej koordynacji reakcji humanitarnej są nieodzownie staraniami „rozwojowymi” (*developmental*), jako że kontekst katastrofy jest złożony i ujawnia się (*emergent*) w czasie rzeczywistym. Ewaluacja powinna być również rozwojowa pod względem wspierania bieżącego podejmowania decyzji w sprawie pomocy humanitarnej. Marginalizowane, pozbawione mocnej pozycji i wykluczone grupy ludności często są szczególnie bezbronne w sytuacji katastrofy, ponieważ zazwyczaj żyją na obszarach o wysokim stopniu narażenia na zagrożenia, na których brakuje podstawowej infrastruktury. Dlatego dostarczenie pomocy na czas jest jeszcze trudniejsze. Ewaluacja rozwojowa może prześledzić zarówno powstające punkty wrażliwości na zagrożenia jak i rozwój interwencji.

## **Dynamiczne kontra statyczne projekty ewaluacji wpływu**

Jak pokazują powyższe przykłady, w ewaluacji rozwojowej interwencje w zakresie rozwoju postrzegane są jako dynamiczne i na bieżąco kształtujące się, w złożonych systemach adaptacyjnych (*complex adaptive systems*). Zarówno interwencja jak i ewaluacja, są dynamiczne i adaptacyjne. Stanowi to dobitny kontrast dla ewaluacji wpływu, które jako ramy metodologiczne wykorzystują eksperymenty losowe (*randomized controlled trials*, RCTs). RCTs dokonują konceptualizacji interwencji jakoby występowała w systemach zamkniętych i badają ją jako statyczną i mechaniczną przyczynę, nakierowaną na przyjęte z góry skutki w prostym liniowym modelu przyczynowo-skutkowym. Takie



projekty mają na celu standaryzację interwencji i kontrolę zmienności, co ogranicza użyteczność i możliwość uogólnienia wniosków<sup>10</sup>. Natomiast ewaluacje rozwojowe zakładają, że rozwój częściej występuje w złożonych dynamicznych systemach i stawiają głównie na rozumienie kontekstu i adaptacyjność w czasie rzeczywistym, nie zaś na wygenerowaniu modelu o wysokim stopniu wierności i na stosowanie rygorystycznych praktyk. Różnice te wykraczają poza metodologiczne preferencje i debaty. Obejmują fundamentalnie różne poglądy na naturę rozwoju, konteksty, w jakich rozwój następuje, sposoby przebiegu zmian, a także różnice epistemologiczne, dotyczące istoty możliwej do wykorzystania wiedzy.

## **Ewaluacja rozwojowa a rozliczanie odpowiedzialności**

Tradycyjne podejście do rozliczania odpowiedzialności (*accountability*) zakłada ewaluację wykorzystania zasobów zgodnie z planem oraz osiągnięcia założonych rezultatów. Jest to statyczne i mechaniczne podejście do zagadnienia *accountability*, które zakłada, że osoby projektujące ewaluację wiedzą z trzyletnim lub pięcioletnim wyprzedzeniem, w jakie istotne rezultaty należy celować i w jaki sposób postępować, aby te pożądane rezultaty osiągnąć. Odejście od planowanego wdrażania jest uznawane za niepowodzenie wdrożenia. Skupienie się na nowych, wyłaniających się szansach, uznaje się za odejście od realizowanej misji (*mission drift*). Mantrą tradycyjnego statycznego podejścia do rozliczania odpowiedzialności jest planowanie pracy, pracowanie według planu i ewaluacja tego czy plan został wykonany. Jednak organizacje osiągające wysokie wyniki nie podchodzą tak ani do rozwoju ani do rozliczenia z wykonanego przedsięwzięcia.

Henry Mintzberg jest jednym z czołowych uczonych na świecie, zajmujących się myśleniem strategicznym, rozwojem organizacyjnym oraz cechami przedsiębiorstw o wysokiej wydajności (*high performing business*). Stwierdził on, że wdrażanie strategii jest zawsze kombinacją procesów celowych i nieplanowanych. Badając przez lata setki firm odkrył, że nie istnieje perfekcyjnie kontrolowany, celowy proces, w którym zamierzenia prowadzą do sformułowania planów, wdrażania oraz pełnej realizacji zamierzonych rezultatów. Rzeczywisty świat nie działa w ten sposób. Jak pokazano na rysunku poniżej, zrealizowana strategia (na niej po pewnym czasie się kończy) zaczyna się jako „strategia zamierzona” (planowanie), jednak nie wszystko z tego, co zamierzono, zostaje zrealizowane. Z pewnych rzeczy się rezygnuje albo nie przeprowadza ich, ponieważ założenia z etapu planowania okazały się wadliwe w obliczu procesów realnego świata; nazywa to „strategią niezrealizowaną”. Pozostała część strategii zamierzonej jest przez niego określana mianem „strategii celowej”, która łączy się ze „strategią wyłaniającą się” (*emergent strategy*), aby stać się „strategią zrealizowaną”. „Strategia wyłaniająca” się powstaje w wyniku pochwycenia nowych szans, co jest kolejnym powodem, dla którego pewne planowane rzeczy pozostają niewykonane,

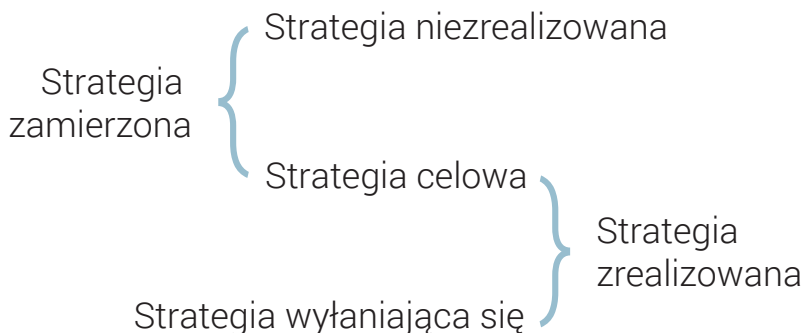
---

<sup>10</sup> Więcej informacji o założeniach mechanicznych i liniowych RCTs w: M. Q. Patton *Utilization-Focused Evaluation*, 4<sup>th</sup> ed., Thousand Oaks, CA 2008, rozdział 12.

wraz z pojawieniem się nowych i lepszych możliwości<sup>11</sup>. W istocie organizacja osiąga-  
jąca wysokie wyniki, zwracająca uwagę na świat, w którym funkcjonuje, nie oczekuje  
sztywnego trzymania się wysoce nakazowego planu. Plan jest punktem wyjścia. Po  
rozpoczęciu wdrażania planu, musi on, i powinien, zostać dostosowany do tego, co  
zaobserwowano i czego się nauczono w interakcji ze złożonym, adaptacyjnym syste-  
mem, dynamiki rzeczywistego świata.

Spostrzeżenia Mintzberga na temat wdrażania strategii w rzeczywistym świecie do-  
bitnie kontrastują z klasycznym podejściem do ewaluacji, zorientowanym na rozlicza-  
nie odpowiedzialności (*accountability-oriented approach to evaluation*), gdzie wdrażanie  
programu i rezultaty są mierzone i oceniane w oparciu o to, co zaplanowano do zrobie-  
nia i to, co udało się osiągnąć (zamierzone rezultaty). W takich ramach, innowacyjny  
i adaptacyjny program, który zakłada wykorzystywanie nowych szans i dopasowywa-  
nie się do zmieniających się warunków, zostanie oceniony negatywnie. Ewaluacja roz-  
wojowa oczekuje natomiast, że część tego, co planowano, zostanie niezrealizowana,  
część zostanie wdrożona mniej więcej zgodnie z oczekiwaniami, a przy tym pojawią  
się pewne nowe elementy. Ewaluacja rozwojowa śledzi i dokumentuje te różne aspek-  
ty strategicznych innowacji - oraz ich następstwa dla dalszych innowacji i rozwoju.  
*Accountability* opiera się na starannym, systematycznym i gruntownym dokumento-  
waniu tych zwrotów rozwojowych, nadawaniu przejrzystości danym, których zmiany  
dotyczą, a także śledzeniu implikacji odstępstw od pierwotnego planu - zarówno od-  
stępstw pod względem wdrażania, jak i wyłaniających się rezultatów.

## Rysunek 2. Mintzberg o strategii



Źródło: H. Mintzberg, *Tracking Strategies*, Oxford University Press, Nowy Jork 2007.

Ewaluacja rozwojowa oparta na złożoności (*complexity-based developmental evaluation*) dokonuje przesunięcia środka ciężkości i przedmiotu rozliczania odpowiedzialności. W ewaluacji rozwojowej *accountability* oznacza dokumentowanie dostosowań

<sup>11</sup> H. Mintzberg, *Tracking Strategies*, Oxford University Press, Nowy Jork 2007, rozdział 1.

i ich następstw, nie zaś ocenę sztywnego trzymania się zaplanowanego wdrażania i z góry założonych rezultatów. Dlaczego? Dlatego, że wrażliwa na kryterium złożoności ewaluacja rozwojowa zakłada, iż plany są zawodne, oparte o niedoskonałe informacje i założenia, które okażą się niesłuszne, a także przewiduje, że rozwój występuje w kontekstach dynamicznych, w których nawet dobre plany będą wymagać dostosowania do zmieniających się realiów. Tym samym, zamiast stawać się barierą na drodze do adaptacji, jak ma to miejsce w tradycyjnych, sztywnych miarach rozliczania odpowiedzialności, w ramach których programy uznaje się za zakończone niepowodzeniem jeśli odbiegają od tego, co zostało zaplanowane, ewaluacja rozwojowa dopuszcza dynamicznie działający świat, z odstępstwami od początkowych planów. Ewaluacja rozwojowa kładzie nacisk na rozumienie, wspieranie i dokumentowanie dostosowań oraz ich następstw.

### **Ewaluacja rozwojowa jako skoncentrowana na wykorzystaniu**

Zagadnienie złożoności każe mi zastanowić się nad jego przydatnością w zrozumieniu pewnych wyzwań ewaluacyjnych. Koncepcje złożoności można wykorzystać do zidentyfikowania i ujęcia w określone ramy okoliczności związanych z interwencją. Dają się one sprowadzić do konkretnej, odpowiedniej dla danej sytuacji, reakcji ewaluacyjnej (*evaluation response*), czyli tego, co nazywam tutaj „ewaluacją rozwojową”. Uwzględnienie złożoności staje się zatem definiującą cechą niszy, jaką wypełnia ewaluacja rozwojowa. Zasady funkcjonowania w złożonych systemach adaptacyjnych nadają świadomy wymiar ewaluacji rozwojowej. Kontrowersje i wyzwania towarzyszące ideom dotyczącym złożoności również, w sposób nieunikniony, będą dotyczyć ewaluacji rozwojowej. Myślenie w kategoriach złożoności, które przyciąga uwagę i wywołuje entuzjazm innowatorów społecznych, będzie obejmować również ewaluację rozwojową i będzie stanowić źródło jej przydatności.

Ewaluacja rozwojowa ma pokazywać, że istnieje opcja w postaci podejścia do prowadzenia ewaluacji, która obejmuje w szczególności „adaptację rozwojową” (*developmental adaptation*). Tym samym umieszczam to podejście w szerszym kontekście „ewaluacji skoncentrowanej na wykorzystaniu”<sup>12</sup>. *Utilization-focused evaluation* jest ewaluacją zarówno dla - jak również realizowaną z udziałem - konkretnych, głównych i przewidzianych użytkowników, na potrzeby konkretnych, zamierzonych zastosowań. „Ewaluacja skoncentrowana na wykorzystaniu” zaczyna się od przesłanki, zgodnie z którą ewaluację należy oceniać z punktu widzenia ich użyteczności i rzeczywistego wykorzystania; dlatego ewaluatorzy powinni ułatwiać proces ewaluacji i opracowywać projekt ewaluacji ze starannością, ponieważ sposób przeprowadzenia wszystkiego, od początku do końca, wpłynie na wykorzystanie. „Wykorzystanie” (*use*) dotyczy zarówno sposobu w jaki rzeczywiste osoby, w rzeczywistym świecie stosują wnioski z ewalu-

---

<sup>12</sup> M. Q. Patton, *Utilization-Focused Evaluation*, 4<sup>th</sup> ed.(...), Patton M. Q., *Utilization-Focused Evaluation*, w: M. Segone (pod red.), *From Policies to Results: Developing Capacities for Country Monitoring & Evaluation Systems*, UNICEF 2009, ss. 252-277, M. Q. Patton, *Essentials of Utilization-Focused* (...)

acji, jak również sposobu w jaki doświadczają procesu ewaluacji. Dlatego w przypadku „ewaluacji skoncentrowanej na wykorzystaniu”, główny ciężar spoczywa na osiągnięciu „zamierzonego użycia przez przewidzianych użytkowników”. W ewaluacji rozwojowej zamierzonym użyciem jest rozwój, co określiłem tutaj jako odrębny i istotny cel ewaluacji. Głównymi przewidzianymi użytkownikami są innowatorzy w dziedzinie rozwoju i inni pracujący nad dokonaniem istotnej zmiany społecznej.

## Rozpoznanie sytuacji a ewaluacja rozwojowa

Wnikliwe rozpoznanie sytuacji jest istotą „ewaluacji skoncentrowanej na wykorzystaniu”. Nie ma lepszego sposobu na przeprowadzenie ewaluacji. Krytyczne znaczenie ma ten wgląd (*insight*). Projekt danej ewaluacji zależy od uczestniczących w niej osób i ich sytuacji. Standardy Komitetu Pomocy Rozwojowej<sup>13</sup> dostarczają ogólne wytyczne, podstawę dla etycznych wskazówek oraz zobowiązanie do kompetencji zawodowej i uczciwości. Jednak nie ma absolutnych zasad jakimi mógłby kierować się ewaluator i dzięki którym dowiedziałby się dokładnie, jak postąpić z konkretnymi użytkownikami, w określonej sytuacji. Rozpoznanie tego problemu, analiza sytuacji, jest jedną z „zasadniczych kompetencji ewaluatorów programu”<sup>14</sup>.

Ideałem byłoby dopasowanie typu ewaluacji do sytuacji i potrzeb przewidzianych użytkowników, w celu osiągnięcia jej zamierzonego wykorzystania. Oznacza to, i chciałbym podkreślić ten aspekt, że „ewaluacja rozwojowa nie jest odpowiednia dla każdej sytuacji”. W żadnym wypadku. Nie sprawdzi się, jeśli warunki i relacje nie będą odpowiednie. Każda ewaluacja obejmuje bowiem wyzwanie jakim jest dopasowanie procesu i podejścia do okoliczności, zasobów, harmonogramów, wymogów dotyczących danych, do polityki, przewidzianych użytkowników oraz do celów w danej sytuacji. Dopasowanie tych elementów ewaluacji, wymaga wnikliwego rozpoznania sytuacji. Ewaluacja rozwojowa jest odpowiednia tam, gdzie sytuacja ma obejmować interwencje i innowacje w złożonych, adaptacyjnych sytuacjach rozwojowych<sup>15</sup>.

---

<sup>13</sup> DAC (Komitet Pomocy Rozwojowej), *Quality Standards for Development Evaluation, 2010*, <http://www.oecd.org/dataoecd/55/0/44798177.pdf>

<sup>14</sup> CES (The Canadian Evaluation Society), *Competencies for Canadian Evaluation Practice, 2010*. [http://www.evaluationcanada.ca/txt/2\\_competencies\\_cdn\\_evaluation\\_practice.pdf](http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf)

<sup>15</sup> M. Q. Patton, *Developmental Evaluation: Applying Complexity (...)*

### Rysunek 3. Wykorzystanie ewaluacji



Written by Mark M. Rogers and illustrated by Lawson Sworh

*"I can honestly say that not a day goes by when we don't use those evaluations in one way or another."*

Źródło: Opracowanie - Mark M. Rogers, ilustracje - Lawson Sworh

## O autorze

### Michael Quinn Patton

---

Założyciel i prezes *Utilization-Focused Evaluation*, niezależnej organizacji rozwoju organizacyjnego i ewaluacji programów. Po uzyskaniu tytułu doktora socjologii Uniwersytetu Wisconsin-Madison przez 18 lat pracował na wydziale Uniwersytetu Minnesota (1973-1991), w tym pięć lat jako dyrektor Centrum Badań Społecznych w Minnesocie i dziesięć lat w Centrum Doradztwa w Minnesocie.

Patton napisał wiele książek na temat sztuki i nauki ewaluacji programów, w których podkreśla znaczenie przydatności ewaluacji, w opozycji do praktyki sporządzania długich raportów, które nigdy nie zostaną przeczytane lub nigdy nie wywołają żadnych zmian. Pisał o ewaluacji i działał w tej dziedzinie na początku lat 70-tych, gdy ewaluacja w sektorze non-profit stanowiła relatywnie nowy obszar.

Patton to jeden z zaledwie dwóch laureatów nagrody Alvy i Gunnara Myrdal (przyznawanej przez Towarzystwo Badań Ewaluacyjnych, a następnie przez Amerykańskie Stowarzyszenie Ewaluacji) za „wybitny wkład w zastosowanie i praktyczne wykorzystanie ewaluacji”, oraz nagrody Paula F. Lazarsfelda za wkład w teorię ewaluacji (przyznawaną przez Amerykańskie Stowarzyszenie Ewaluacji). Towarzystwo Socjologii Stosowanej uhonorowało go w 2001 nagrodą Lestera F. Warda za wybitny wkład do socjologii stosowanej. Był prezesem Amerykańskiego Stowarzyszenia Ewaluacji w 1988 i współprzewodniczącym międzynarodowej konferencji ewaluacyjnej w 2005 w Toronto, sponsorowanej wspólnie przez amerykańskie i kanadyjskie stowarzyszenia ewaluacji. Zasiada w Redakcyjnej Radzie Doradczej czasopisma *The Foundation Review*.

## **Bibliografia:**

1. CES (The Canadian Evaluation Society), *Competencies for Canadian Evaluation Practice, 2010*. [http://www.evaluationcanada.ca/txt/2\\_competencies\\_cdn\\_evaluation\\_practice.pdf](http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf)
2. DAC (Komitet Pomocy Rozwojowej), *Quality Standards for Development Evaluation, 2010*, <http://www.oecd.org/dataoecd/55/0/44798177.pdf>
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. (red.), *Developmental evaluation. Encyclopedia of evaluation*, Sage, Thousand Oaks, CA 2005, s. 116.
5. Mintzberg H., *Tracking Strategies*, Oxford University Press, Nowy Jork 2007.
6. Patton M. Q., Developmental Evaluation, *Evaluation Practice* 15, 1994, s. 311-320.
7. Patton M. Q. *Utilization-Focused Evaluation, 4<sup>th</sup> ed.*, Sage, Thousand Oaks, CA 2008
8. Patton M. Q., Utilization-Focused Evaluation, w: Segone M. (ed.), *From Policies to Results: Developing Capacities for Country Monitoring & Evaluation Systems*, UNICEF 2009, ss. 252-277.
9. Patton M. Q., *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford, Nowy Jork 2011 <http://www.guilford.com/excerpts/patton.pdf>
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, Kanada 2006.





## Przegląd systematyczny koncepcji ewaluacji opartej na teorii

### Wprowadzenie

Celem rozdziału jest przedstawienie wyników przeglądu systematycznego najbardziej popularnych i powszechnie dyskutowanych podejść ewaluacyjnych opartych na teorii programu (*theory-based evaluation*, TBE). Mnogość istniejących w literaturze teorii wymaga usystematyzowania kwestii terminologicznych poprzez identyfikację wspólnych elementów różnych definicji i koncepcji, które przesądzą o oryginalności tego podejścia. Następnie rozdział koncentruje się na analizie wybranych koncepcji TBE pod kątem podejścia do kwestii tzw. czarnej skrzynki programu, problemu atrybucji oraz sposobów odtwarzania i testowania teorii. W końcu w rozdziale wskazane zostanie, na ile to podejście jest stosowane w praktyce oraz jakie możliwości daje wykorzystanie TBE w badaniach polityk publicznych, w szczególności dotyczących kompleksowych programów finansowanych ze środków Unii Europejskiej.

Ewaluacja oparta na teorii programu stawia sobie za cel udzielenie odpowiedzi na pytanie: dlaczego i w jaki sposób dana interwencja powoduje określone, zamierzone i niezamierzone, efekty. Innymi słowy, podejście to nie koncentruje się na zmierzeniu efektów, ale na zidentyfikowaniu mechanizmu zmiany – określeniu, dlaczego i jak działa interwencja<sup>16</sup>. W tym celu sięga ona do wnętrza interwencji publicznej, czyli tzw. czarnej skrzynki programu. Pojęcie „czarnej skrzynki” odnosi się do tej części logiki interwencji, która znajduje się pomiędzy tym, co „na wejściu” (zasoby zaangażowane w program), i tym, co „na wyjściu” (spodziewane wyniki interwencji)<sup>17</sup>. Stame definiuje „czarną skrzynkę” jako przestrzeń między faktycznym wkładem a oczekiwanym rezultatem programu<sup>18</sup>.

Podejścia do ewaluacji oparte na teorii programu pojawiły się ze względu na ograniczenia poprzednio dominujących podejść zorientowanych metodologicznie (*method-oriented approaches*). Badania quasi-eksperymentalne nie przyniosły bowiem rozwiązania problemu otwarcia „czarnej skrzynki” programu czy interwencji publicznej<sup>19</sup>. Dzięki tym schematom badawczym można co najwyżej odpowiedzieć na pytanie, czy program działa, a więc czy efekt interwencji jest pozytywny czy negatywny i jaka jest

<sup>16</sup> Komisja Europejska, *Monitoring and evaluation of European cohesion policy: concepts and recommendations*, Bruksela 2011.

<sup>17</sup> R. Pawson, N. Tilley, *Realistic Evaluation*, Sage, Londyn 1997.

<sup>18</sup> N. Stame, „Theory-based Evaluation and Types of Complexity”, *Evaluation* 2014, vol. 10, no. 1, s. 58.

<sup>19</sup> Tamże, s. 58.

wielkość tego efektu. Podejścia te nie pozwalają samodzielnie pójść dalej w analizie polityk publicznych, a więc wyjaśnić, dlaczego i jak program działa. W związku z tym podejścia te nadają się do podsumowania efektów (funkcja konkluzywna i rozliczeniowa ewaluacji), ale mają niewielki potencjał dla ewaluacji formatywnej (modyfikacja lub tworzenie nowej interwencji). Wynika to z tego, że w niewielkim stopniu dostarczają one informacji istotnych z punktu widzenia podejmowania decyzji na temat przyszłego kształtu programów i polityk publicznych<sup>20</sup>.

Podjęta w latach 80. krytyka podejść ewaluacyjnych opartych wyłącznie na ilościowych metodach szacowania efektów interwencji doprowadziła do zwiększenia zainteresowania ewaluacją opartą na teorii. Choć źródeł tego podejścia niektórzy autorzy<sup>21</sup> doszukują się już w pracach dotyczących testowania teorii programów z lat 30.<sup>22</sup>, rozwijanych w kolejnych trzech dekadach, wyraźna zmiana paradygmatu w podejściu do ewaluacji pojawiła się wraz z publikacją Chen i Rossiego z 1989 roku<sup>23</sup>. Od tego czasu rośnie zainteresowanie TBE, co objawia się zarówno zwiększeniem liczby artykułów naukowych na ten temat, jak i proliferacją terminów i różnych strategii realizacji TBE.

W odniesieniu do podejścia ewaluacyjnego opartego na teorii programu stosowane są różnorodne terminy, których zakres definicyjny niekoniecznie jest rozdzielny<sup>24</sup>. Do najpopularniejszych należą<sup>25</sup>: ewaluacja wiedzona teorią (*theory-driven evaluation*), ewaluacja oparta na teorii programu (*programme-theory evaluation*), ewaluacja kierowana teorią (*theory-guided evaluation*), teoria akcji (*theory of action*), teoria zmiany (*theory of change*), logika programu (*program logic*), ramy logiczne (*logical frameworks*), realistyczna ewaluacja (*realist/realistic evaluation*), nauka ewaluacji wiedzona teorią programu (*program theory-driven evaluation science*)<sup>26</sup>.

Różnorodność nazewnictwa skłania do zadania pytania o to, czy mamy do czynienia z jednym podejściem, czy też z wieloma różnymi. A w konsekwencji, czy można określić wspólny mianownik dla koncepcji opisywanych przez różnych autorów, który odróżniałby TBE od innych podejść ewaluacyjnych. Choć jednoznaczność definicyjna i operacyjna TBE bywa kwestionowana<sup>27</sup>, wskazać można wspólne cechy TBE stanowiące o spójności tej koncepcji:

---

<sup>20</sup> Tamże.

<sup>21</sup> C. L. S. Coryn i in., „A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009”, *American Journal of Evaluation* 2011, vol. 32, no. 2, s. 200.

<sup>22</sup> Chodzi tu o pracę Tylera dotyczącą kwestii formułowania i testowania teorii programu dla celów ewaluacyjnych. w: Tamże.

<sup>23</sup> H. T. Chen, P. H. Rossi, „Issues in the theory-driven perspective”, *Evaluation and Program Planning* 1989, vol. 12, no. 4.

<sup>24</sup> 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.

<sup>25</sup> Ze względu na to, że pojęcia te pochodzą z literatury anglojęzycznej i dotąd nie funkcjonowały powszechnie w języku polskim, podane są one w autorskim tłumaczeniu auterek rozdziału.

<sup>26</sup> C. L. S. Coryn i in., *op.cit.*

<sup>27</sup> Tamże.

- a) otwarcie „czarnej skrzynki” programu celem odpowiedzi na pytanie, dlaczego i jak interwencja działa (określenie mechanizmów zmiany),
- b) uwzględnienie w ewaluacji kontekstu, w tym środowiska organizacyjnego, w którym działają główni gracze (aktorzy), i szerszego kontekstu kulturowego i politycznego programu,
- c) metodologiczna neutralność, tzn. TBE nie preferuje żadnych konkretnych metod, lecz uzależnia wybór metod od potrzeb,
- d) przywiązywanie wagi do ustanowienia przyczynowości, czyli powiązania oczekiwanych/zaistniałych efektów z interwencją dla rozwiązania problemu atrybucji (w jakim stopniu dana interwencja przyczyni się/przyczyniła się do zaplanowanej/ obserwowanej zmiany).

Warto podkreślić, że według autorów rozwijających teoretyczne podstawy tego podejścia schemat badania w oparciu o TBE powinien składać się z dwóch komponentów:

- a) konceptualnego – którego celem jest sformułowanie teorii programu, tzn. logiki leżącej u podłoża programu, zgodnie z którą interwencja ma osiągać założone rezultaty,
- b) empirycznego – którego zadaniem jest testowanie tej teorii, a więc sprawdzenie, czy, jak i dlaczego interwencja powoduje obserwowalne zmiany (rezultaty).

Według Rogers<sup>28</sup>, jeśli logika programu została właściwie odtworzona (komponent konceptualny), ale nie jest ona wykorzystywana jako model prowadzący proces ewaluacji (komponent empiryczny), nie mamy do czynienia z ewaluacją opartą na teorii.

Teorie programu są najczęściej opisywane lub prezentowane w formie graficznej. Zawiera ona albo prosty, linearny model teorii uwzględniający co najmniej trzy elementy: wkład/zasoby – działania/procesy – produkty/rezultaty/efekty, albo bardziej rozwinięty, złożony model uwzględniający szereg czynników wynikających z kontekstu, w którym program jest osadzony (np. potrzeby, problemy). W przedstawionym poniżej przykładzie linearnego przedstawienia teorii poszczególne jej elementy należy rozumieć tak<sup>29</sup>:

- a) wkład to zasoby (ludzkie, finansowe, rzeczowe) niezbędne do wdrożenia programu,
- b) działania to czynności podjęte w celu osiągnięcia zakładanego celu (np. przeprowadzenie szkolenia),
- c) produkty to natychmiastowy wynik podjętych działań (np. liczba przeszkolonych osób),
- d) wstępne rezultaty (initial outcomes) to zmiany w wiedzy i umiejętnościach,
- e) pośrednie rezultaty (intermediate outcomes) to zmiany w zachowaniu (*behavioral changes*),

<sup>28</sup> P. J. Rogers, „Program theory evaluation: Not whether programs work but how they work”, w: D. L. Stufflebeam i in. (red.), *Evaluation models: Viewpoints on educational and human services evaluation*, Kluwer, Boston 2000, cyt. za: Coryn i in., *op.cit.*

<sup>29</sup> C. L. S. Coryn i in., *op.cit.* s. 201.

f) długoterminowe rezultaty (*long-term outcomes*) to rozwiązanie społecznego problemu lub zaspokojenie potrzeb danej grupy społecznej objętej interwencją.

### Rysunek 1. Graficzne przedstawienie elementów teorii programu – przykład



Źródło: Na podstawie schematu S.L. Donaldsona, przedstawionego w: Coryn C.L.S. i in., "A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009".

Neutralność metodologiczna TBE oznacza brak sztywnej hierarchii metod używanych w ewaluacji. Jak zauważył White: „Ewaluacje powinny kierować się (ewaluowaną) materią, a nie metodami” („Evaluations should be issues-led not methods-led”)<sup>30</sup>. Według tego autora wybór metod powinien zależeć od pytań ewaluacyjnych i możliwości różnych metod dostarczenia odpowiedzi na nie. W praktyce TBE zazwyczaj łączy wykorzystanie zarówno metod ilościowych, jak również jakościowych.

### Wyniki przeglądu systematycznego

Wyniki przeglądu systematycznego przedstawione w niniejszym rozdziale odnoszą się do dwóch podstawowych kwestii: podejścia poszczególnych koncepcji TBE do „czarnej skrzynki” i jej zawartości (1) oraz do sposobów odtwarzania i testowania teorii programu (2). Przekonania o naturze „czarnej skrzynki” warunkują rolę i zadania ewaluatora. Przede wszystkim wpływa to na sposób, w jaki konstruowany jest schemat badawczy, i na to, jakie metody zostaną wybrane do badania ewaluacyjnego. Przegląd koncentruje się na najbardziej popularnych i powszechnie dyskutowanych podejściach, które występują w syntetycznych zestawieniach i przewodnikach po ewaluacji opartej na teorii.

Sięgając korzeni podejść TBE, Chen i Rossi<sup>31</sup> w koncepcji ewaluacji wiedzonej teorią (theory-driven evaluation) stwierdzili, że jednym z głównych problemów badań ewaluacyjnych jest to, że koncentrują się one na relacjach między wkładem a wpływem interwencji, nie uwzględniając zawartości samego programu. Z tego względu są to „ewaluacje czarnej skrzynki”. Postulowane przez tych autorów przejście do nowego paradygmatu w badaniach ewaluacyjnych, biorącego pod uwagę teorię programu, nie jest jednak łatwe. W opinii tych autorów programy nie poddają się łatwo badaniu, gdyż proces ich przygotowania nie jest oparty na teoriach społecznych. W efekcie cele programów, założone efekty i działania nie są jasno określone. W związku z tym samo badanie ewaluacyjne sprowadza się co najwyżej do „policzenia interesariuszy, opisu

<sup>30</sup> H. White, „Theory-based impact evaluation: principles and practice”, *Journal of Development Effectiveness* 2009, vol. 1, no. 3, s. 162.

<sup>31</sup> H. T. Chen, P. H. Rossi, *op.cit.*

programów i czasami oszacowania efektów"<sup>32</sup>, a „czarna skrzynka” programu jest po prostu pusta – brakuje w niej teorii<sup>33</sup>. W tym kontekście rolę ewaluacji jest dostarczenie tej brakującej teorii, czyli określenie, w jaki sposób interwencja mogłaby prowadzić do określonych skutków. Przy czym autorzy ci proponują, aby przy tworzeniu teorii programu korzystać nie tylko z dostępnych wyników badań oraz teorii społecznych, na których programy publiczne powinny być budowane, ale również z perspektywy twórców programu oraz analizy procesu wdrażania programu.

Podejście ewaluacji opartej na teorii (*theory-based evaluation*, TBE) zaproponowane przez Weiss<sup>34</sup> podkreśla rolę praktyki: w tym ujęciu programy co prawda posiadają swoje założone teorie, ale są one zniekształcone przez proces podejmowania decyzji, który w gruncie rzeczy ma charakter polityczny (strategiczny). W związku z tym dobra teoria programu wyjaśnia, w jaki sposób ten proces decyzyjny zachodzi w praktyce. Według definicji Weiss z 1995 roku teoria zmiany to zestaw założeń dotyczących „minikroków”, które prowadzą zarówno do osiągnięcia długoterminowego celu programu, jak i powiązań między działaniami programu a rezultatami, które pojawiają się na każdym z kroków podjętych w kierunku celu (ostatecznego efektu). Co więcej, badaczka wskazała, że „czarna skrzynka” programu jest pełna różnych teorii zmiany<sup>35</sup> – są to założenia i przeświadczenia różnych interesariuszy na temat tego, jak program powinien działać. Stąd podstawowym zadaniem ewaluatora jest ścisła współpraca z szerokim zakresem interesariuszy przy odtwarzaniu teorii zmiany. Rolą ewaluacji jest konfrontowanie i integrowanie tych różnych teorii celem zbudowania jednej, syntetycznej teorii zmiany. Pierwszym krokiem na tej drodze jest odkrycie teorii implementacji (*implementation theory*), a więc opisanie kroków wdrożenia programu. Dopiero kolejnym etapem jest wskazanie teorii programowej (*programmatic theory*) odkrywającej rzeczywisty mechanizm, który miał miejsce w ramach interwencji. Przy rekonstruowaniu teorii programu zadaniem ewaluatora jest ustalenie w pierwszej kolejności długoterminowej wizji programu – ostatecznego celu oraz powiązania jej z istniejącym problemem wymagającym rozwiązania. Następnie konieczne jest określenie oczekiwanych efektów programu, rezultatów i produktów programu, które są „minikrokami” do osiągnięcia celu. W końcu interesariusze programu wskazują zasoby, które można zaangażować do realizacji planowanych interwencji. Jak widać, proces ustanawiania teorii jest odwrócony – ewaluator, rekonstruując teorię, cofa się krok po kroku od celu

---

<sup>32</sup> Tamże, s. 299.

<sup>33</sup> N. Stame, *op.cit.*

<sup>34</sup> C. H. Weiss, „Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families”, w: J. Connell i in. (red.), *New approaches to evaluating community*, The Aspen Institute, Washington DC 1995, cyt. za: Komisja Europejska, *Evalsed Sourcebook: Methods and Technics*, Bruksela 2013, dokument dostępny na stronie: [http://ec.europa.eu/regional\\_policy/en/newsroom/news/2013/11/evaluation-guidance-evaled-guide-and-sourcebook-updates](http://ec.europa.eu/regional_policy/en/newsroom/news/2013/11/evaluation-guidance-evaled-guide-and-sourcebook-updates) (dostęp 7.04.2017 r.).

<sup>35</sup> C. Weiss, „Theory-Based Evaluation: Past, Present, and Future”, *New Directions for Evaluation* 1997, no. 76.

(ostatecznego efektu) do najwcześniejszych zmian uruchomionych (lub planowanych do uruchomienia) przez interwencję.

Trzecim podejściem w tej dyskusji jest tzw. realistyczna ewaluacja (*realistic evaluation*) zaproponowana przez Pawsona i Tilley'ego<sup>36</sup>. Podejście realistyczne do ewaluacji zakłada, że we właściwej teorii interwencji rezultat jest wypadkową kontekstu i mechanizmu. Sam program ma potencjał generowania określonych rezultatów, ale w końcu to ludzie (interesariusze programu) zanurzeni w kontekście (np. organizacyjnym) podlegający programowi (publicznej interwencji) są w stanie lub nie zainicjować określony mechanizm zmiany. Wyjaśnienie dylematu „czarnej skrzynki” jest tu takie: składa się ona z ludzi działających w określonym kontekście i zgodnie z pewnym mechanizmem<sup>37</sup>. Ponieważ program przynosi różne wyniki dla różnych interesariuszy i w różnym kontekście, dla procesów decyzyjnych ważne jest ustanowienie w trakcie ewaluacji, co działa dla kogo i w jakim kontekście<sup>38</sup>. Pozwala to na właściwe replikowanie programów w przyszłości.

Schemat badania zakłada tu, że najpierw ewaluator ustanawia, jak mechanizm mógłby zadziałać, a następnie konfrontuje go z opiniami interesariuszy biorących udział w powstawaniu i wdrażaniu programu. Odtworzenie teorii programu rozpoczyna się od analizy dokumentacji programowej oraz od przeprowadzenia wywiadów z twórcami programu celem identyfikacji teorii programu. Następnie takie teorie podlegają kodyfikacji poprzez zastosowanie różnych technik (tj. mapowanie koncepcji, tworzenie drzew problemów i rozwiązań, budowanie scenariuszy) i przedstawiane są w efekcie w formie graficznej (diagramu). Kolejnym krokiem jest formalizacja teorii programu, tj. przełożenie jej na hipotezy nadające się do empirycznego testowania. Ostatnim krokiem jest badanie empiryczne wykorzystujące metody jakościowe i ilościowe, które pozwala przetestować i doprecyzować faktyczną teorię programu.

Jak widać z powyższego zestawienia, teoria zmiany i realistyczna ewaluacja różnią się w podejściu co do sposobów konceptualizacji teorii programu. W przypadku teorii zmiany odtworzenie teorii następuje poprzez budowanie konsensusu między interesariuszami, tak aby następnie testować zakładane przez nich wyniki. Realisci z kolei zakładają kumulatywny proces budowania, testowania i doskonalenia teorii. W trakcie tego procesu poszukiwane są relacje między interwencjami a ich wynikami, a mechanizm ustanawiania tych relacji jest generatywny.

Kolejne podejście TBE koncentruje się przede wszystkim na strategiach rekonstrukcji teorii programu. Leew i Vaessen traktują „czarną skrzynkę” programu jako zbiór behawioralnych i społecznych mechanizmów wpływających na zachowania odbiorców

---

<sup>36</sup> R. Pawson, N. Tilley, *op.cit.*

<sup>37</sup> N. Stame, *op.cit.*

<sup>38</sup> Komisja Europejska, *Evalsed Sourcebook: Methods and Technics*, Bruksela 2013, s. 52-53.

programu<sup>39</sup>. Stąd teorie programu składają się z zestawu behawioralnych, społecznych i instytucjonalnych założeń dotyczących tego, jak działania i zasoby programu przyniosą oczekiwaną zmianę. Leeuw i Vaessen zakładają, że ustanawiając teorię programu, należy zidentyfikować założenia dotyczące tego, jak i dlaczego działania programu przyniosą oczekiwane zmiany. Leeuw wskazuje również trzy autonomiczne strategie dotyczące sekwencji kroków niezbędnych do odtworzenia tych założeń oraz metod rekonstrukcji mechanizmów zmiany, tj.: polityczno-naukową (*policy-scientific approach*), oceny strategicznej (*strategic assessment approach*) i metodę aktywizacji (*elicitation method*). Każda z tych strategii korzysta z innych metod zbierania danych i ich analizy<sup>40</sup>.

W metodzie polityczno-naukowej zakłada się, że odtworzenie teorii opiera się na analizie formalnych i nieformalnych dokumentów oraz transkrypcji wywiadów celem identyfikacji założeń dotyczących celów programów oraz przyjętych sposobów ich osiągnięcia. Na tej podstawie odtwarzane są założone mechanizmy, które mają rozwiązać problem. Do oceny spójności odtworzonej logiki interwencji służy analiza argumentacji stosowana w filozofii i logice. W tym podejściu ewaluacja teorii programu może zachodzić poprzez konfrontację różnych teorii, empiryczne testowanie teorii programu lub przez proces ciągłego udoskonalania teorii z wykorzystaniem różnych technik zbierania danych<sup>41</sup>.

Podejście oceny strategicznej z kolei koncentruje się na rekonstrukcji założeń leżących u podstaw utworzenia programu. W tym celu wyłaniane są grupy interesariuszy względnie homogeniczne ze względu na ich charakterystyki i przekonania dotyczące interwencji. W ramach grup odbywa się analiza głównych aktorów biorących udział w programie, a następnie identyfikowane są założenia dotyczące zachowań aktorów przy wdrażaniu interwencji, prowadzące do powodzenia programu. Odtworzone założenia podlegają następnie rankingowaniu w celu wyłonienia tych najważniejszych. W końcu debata między grupami oparta na argumentacji złożonej z odtworzonych założeń pozwolić ma na zrozumienie strategii różnych grup i ich syntezę na ostatnim etapie prac nad odtworzonymi założeniami. Drogą do ustalenia założeń stojących u podstaw interwencji jest tu konsensus między różnymi grupami jej interesariuszy.

W przypadku metody aktywizacji podstawą do rekonstrukcji teorii są mentalne modele lub, inaczej mówiąc, mapy kognitywne ludzi działających w organizacjach realizujących programy. Mapy te odnoszą się do wyobrażeń osób na temat organizacji zjawisk w przestrzeni, wpływając na ich zachowania. Stąd kluczowym zadaniem ewaluatora jest tu odtworzenie przekonań interesariuszy na temat organizacji, w której pracują,

---

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

<sup>40</sup> F. L. Leeuw, „Reconstructing Program Theories: Methods Available and Problems to be Solved”, *American Journal of Evaluation* 2003, vol. 24, no. 1.

<sup>41</sup> Komisja Europejska, *Evalsed Sourcebook...*, *op.cit.*, s. 58.

i jej otoczenia, co pozwala na ich późniejsze porównanie z dowodami pochodzącymi z badań naukowych dotyczących podobnych organizacji. Takie podejście przede wszystkim ułatwia organizacjom uczenie się z dotychczasowych doświadczeń wdrażania programu.

Ostatnie z omawianych tutaj podejść TBE – analiza kontrybucji (*contribution analysis*) – koncentruje się nie na procesie rekonstrukcji założeń leżących u podstaw teorii interwencji, ale na ocenie ich wiarygodności. Podstawą tego podejścia jest przekonanie, że chociaż wyniki badań wskazują często na to, że rezultaty interwencji zaistniały, to nie jest to równoważne z tym, że to interwencja przyczyniła się do osiągnięcia tych rezultatów<sup>42</sup>. Z tego względu w podejściu tym kluczowe jest pozyskanie dowodów na to, że to właśnie program przyczynił się do obserwowanej zmiany. Podstawą do określenia kontrybucji programu jest odtworzenie wstępnej teorii zmiany na podstawie analizy przyczyn i skutków interwencji, a następnie weryfikacja jej prawdziwości poprzez zebranie istniejących dowodów. Dopiero takie wstępne przetestowanie teorii w świetle istniejących dowodów pozwala na jej rozwinięcie do krytycznej oceny. Kolejnym krokiem jest zebranie danych pierwotnych tam, gdzie dodatkowe dowody są potrzebne. W końcu pozyskane empirycznie dane pozwalają na zbudowanie historii kontrybucji, która jest tym bardziej wiarygodna, im silniejsze są dowody na zaistnienie powiązań między interwencją a obserwowanymi zmianami. W podejściu tym „czarna skrzynka programu” składa się z działań, które prowadzą do ciągu rezultatów powiązanych związkami przyczynowo-skutkowymi. W związku z tym podejście to koncentruje się na sposobach oceny wiarygodności związków między przyczynami uruchomienia interwencji a jej skutkami.

Dyskusja na temat różnych podejść TBE do określania teorii programu i mierzenia wkładu (kontrybucji) interwencji do uzyskanych efektów doprowadziła do pytania o to, na ile ewaluacja oparta na teorii jest podejściem odrębnym od ewaluacji wykorzystującej metody (quasi)eksperymentalne. Niektórzy autorzy postrzegają oba podejścia jako różne ze względu na inne podstawy przyjmowane do określania przyczynowości<sup>43</sup>. Inni zaprzeczają tej dychotomii i proponują rozwiązania łączące TBE z podejściem wykorzystującym badania eksperymentalne. White zaproponował ramy teoretyczne nazwane ewaluacją wpływu opartą na teorii (TBIE)<sup>44</sup>, integrującą oba te podejścia. Sześć zasad TBIE obejmuje: (1) mapowanie łańcucha przyczynowego (teoria programu), (2) zrozumienie kontekstu, (3) antycypowanie heterogeniczności, (4) rygorystyczną ewaluację wpływu wykorzystującą wiarygodny schemat kontrfaktyczny, (5) rygorystyczną analizę kontrfaktyczną, (6) zastosowanie mieszanych metod. Wykorzystanie podejścia opartego na teorii oraz podejścia kontrfaktycznego łącznie jest postrzegane

<sup>42</sup> J. Mayne, „Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly”, *The Canadian Journal of Program Evaluation*, 2001, vol. 16, no. 1, s. 3.

<sup>43</sup> E. Stern i in., *Broadening The Range of Designs and Methods for Impact Evaluations, Report of a study commissioned by the Department for International Development, DFID 2012.*

<sup>44</sup> H. White, *op.cit.*



przez ekspertów i interesariuszy<sup>45</sup> jako najlepsze podejście do analizy wpływu polityk publicznych (*theory-based counterfactual impact evaluation*) w ramach ewaluacji *ex post*.

## Praktyczne zastosowanie podejścia

Podejście TBE stosowane jest i promowane przez liczne organizacje międzynarodowe, np. Bank Światowy, Międzynarodową Inicjatywę dla Ewaluacji Wpływu, Grupę Ewaluacyjną Zgromadzenia Narodów<sup>46</sup>, Komisję Europejską<sup>47</sup>. Niemniej zastosowanie TBE w praktyce nie odpowiada zainteresowaniu teoretycznemu, jakie poświęcane jest temu zagadnieniu w literaturze<sup>48</sup>. Co więcej, przegląd systematyczny badań wykorzystujących TBE w latach 1990-2009<sup>49</sup> wskazuje, że tylko w 13% z nich badacze sformułowali teorie programu na podstawie obserwacji programu, a tylko 36% ewaluacji faktycznie mierzyło wyniki postulowane w teorii programu. Dodatkowo Nakrošis twierdzi, że jak dotąd brakuje konkretnych przykładów dobrych praktyk i potencjalnych korzyści zastosowania podejścia ewaluacji opartej na teorii w programach finansujących rozwój regionalny w krajach UE<sup>50</sup>.

Jeśli chodzi o samą ewaluację wpływu opartą na teorii, to nadal niewiele badań wydaje się sprostać założeniom tego podejścia w praktyce<sup>51</sup>. Pomimo rosnącego zainteresowania i szybko rosnącej liczby artykułów naukowych poświęconych temu tematowi nie ma konsensusu dotyczącego tego, jak zastosować to podejście w praktyce<sup>52</sup>.

W przypadku programów i projektów współfinansowanych ze środków Unii Europejskiej TBE jest uznawana przez Komisję Europejską jako użyteczne narzędzie ewaluacji na każdym etapie „cyklu życia” programu/projektu<sup>53</sup>. Na etapie ewaluacji *ex ante* instytucje zarządzające powinny zadbać o wyspecyfikowanie teorii programu. Wskazanie logiki programu powinno polegać na opisanu explicite zakładanego mechanizmu zmiany, leżącego u podłoża interwencji, poprzez jednoznaczne określenie:

- a) w jaki sposób oczekiwana zmiana przyczyni się do osiągnięcia celów lokalnych i europejskich (tzw. *programme theory*),
- b) w jaki sposób planowane działania mają przyczynić się do osiągnięcia założonych rezultatów (tzw. *implementation theory*).

---

<sup>45</sup> Tamże, Komisja Europejska, *Evalsed Sourcebook...*, *op.cit.*

<sup>46</sup> C. L. S. Coryn i in., *op.cit.*

<sup>47</sup> Komisja Europejska, *Evalsed Sourcebook...*, *op.cit.*

<sup>48</sup> C. L. S. Coryn i in., *op.cit.*, s. 217.

<sup>49</sup> Tamże.

<sup>50</sup> V. Nakrošis, „Theory-based evaluation of capacity-building interventions”, *Evaluation* 2014, vol. 20, no. 1.

<sup>51</sup> H. White, *op.cit.*

<sup>52</sup> V. Nakrošis, *op.cit.*

<sup>53</sup> M. Riché, *Theory Based Evaluation: A wealth of approaches and an untapped potential*, Komisja Europejska 2012.

Podczas ewaluacji *on-going* celem zastosowania ewaluacji opartej na teorii programu jest ocena tego, czy teoria implementacji jest realizowana skutecznie, a więc czy prowadzi do osiągnięcia rezultatów średniookresowych (*mid term outcomes*).

Z kolei wykorzystanie podejścia TBE na poziomie ewaluacji *ex post* służy przyjrzeniu się kontekstowi i mechanizmowi zmiany, by określić sposób osiągnięcia (lub nieosiągnięcia) poszczególnych celów interwencji.

Jak podkreśla Stame<sup>54</sup>, ewaluacja oparta na teorii dostarcza rozwiązań dla deficytów związanych z przeprowadzaniem oceny programów realizowanych w wieloszczeblowych systemach zarządzania. W przypadku ewaluacji takich kompleksowych programów założyć należy, że mamy do czynienia nie z jednym mechanizmem zmiany, lecz kilkoma, które oddziałują wzajemnie na siebie. W warunkach złożoności ewaluowanej materii należy zwrócić uwagę na relacje i zależności między różnymi mechanizmami, co pozwala uniknąć błędów przy stosowaniu TBE. Zwracają na to uwagę zarówno Hedström i Ylikoski<sup>55</sup>, Rogers<sup>56</sup>, jak również Rogers i Funnell<sup>57</sup>. Inne problemy, z którymi może borykać się ewaluator stosujący TBE, to niewłaściwie zrekonstruowane teorie<sup>58</sup> lub pułapka „braku teorii”<sup>59</sup>. Niewłaściwie zrekonstruowane teorie mogą wynikać z problemu quasi-wzbogacenia teorii<sup>60</sup>. Taka sytuacja zachodzi wtedy, kiedy twórcy polityk dążą do uwzględnienia w teorii, zrekonstruowanej przez ewaluatorów, założeń, które *de facto* nie występowały, kiedy program był tworzony. W przypadku „braku teorii” mamy do czynienia z graficznymi przedstawieniami działań i rezultatów bez określenia logicznych związków między nimi. Problemem może być również niewykorzystanie odtworzonej teorii przy realizacji badań empirycznych, czyli realizacja badań w oderwaniu od zrekonstruowanej teorii<sup>61</sup>. Należy to mieć na uwadze zwłaszcza w ewaluacji polityk publicznych, ponieważ nierzadko w ramach realizowanych interwencji nie istnieją przemyślane i dobrze ustrukturyzowane teorie programów (tj. spisane w ramach formalnej dokumentacji programowej). W takich sytuacjach teorie są często odtwarzane *post factum*, a w skrajnej sytuacji – tworzone *ad hoc* („dorabiane”) na poziomie badania ewaluacyjnego. W związku z tym takie badania wykorzystujące TBE mogą być podatne na wystąpienie

---

<sup>54</sup> N. Stame, *op.cit.*

<sup>55</sup> P. Hedström, P. Ylikoski, „Causal Mechanisms in the Social Sciences”, *Annual Review of Sociology* 2010, vol. 36 no. 1.

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

<sup>57</sup> S. C. Funnell, P.J. Rogers, *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*, Jossey Bass 2011, cyt. za: Komisja Europejska, *Evalsed Sourcebook...*, *op.cit.*

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

<sup>59</sup> S. C. Funnell, P. J. Rogers, *op.cit.*

<sup>60</sup> Komisja Europejska, *Evalsed Sourcebook...*, *op.cit.*

<sup>61</sup> S. C. Funnell, P. J. Rogers, *op.cit.*

opisanych wyżej problemów związanych z (od)tworzeniem teorii programu i powinny być przedmiotem szczególnej uwagi.

## **Podsumowanie**

Podsumowując, ewaluacja oparta na teorii programu od dziesięcioleci budzi zainteresowanie teoretyków i praktyków. Decyduje o tym komplementarność TBE względem podejść zorientowanych głównie na oszacowanie efektów interwencji i wysoka wartość praktyczna podejścia (użyteczność w ocenie skuteczności interwencji publicznych). Brak jest jednak zintegrowanej koncepcji całościowej dla tego paradygmatu, co ilustruje mnogość podejść teoretycznych i metodycznych do tego zagadnienia. To z kolei przekłada się na relatywnie nieczęste jak dotąd wykorzystanie podejścia TBE w praktyce badawczej.

Ewaluacja oparta na teorii dostarcza ram koncepcyjnych i narzędzi mających prowadzić do otwarcia „czarnej skrzynki” kompleksowych programów, do których należą – przykładowo – programy finansowane ze środków unijnych. Pozwala ona na odtworzenie związków przyczynowo-skutkowych zachodzących między nakładami (środkami finansowymi zaangażowanymi w program) a zakładanymi rezultatami wdrożenia zaplanowanych działań. Jej użyteczność należy postrzegać przede wszystkim poprzez pryzmat informacji, których dostarcza twórcom i realizatorom polityk i programów publicznych. TBE pozwala na odtworzenie mechanizmów zmiany przewidzianych w programie, a w konsekwencji zasila procesy decyzyjne informacjami na temat tego, dlaczego i jak dana interwencja działa (lub nie działa). Dzięki wykorzystaniu tego podejścia na etapie programowania, wdrażania i oceny efektów programu możliwe jest zarówno lepsze zaplanowanie kolejnych interwencji, jak i skuteczne podejmowanie działań naprawczych w trakcie wdrażania programu.

## O autorkach

### Katarzyna Hermann-Pawłowska

---

Analitik, ewaluator, ekspert w dziedzinie planowania strategicznego, funduszy unijnych i partycypacji publicznej. Członek Zarządu i Dyrektor ds. Projektów Strategicznych w Fundacji IDEA Rozwoju. Z wykształcenia europeistka, absolwentka Uniwersytetu im. A. Mickiewicza w Poznaniu i Kolegium Europejskiego w Natolinie, tytuł doktora nauk ekonomicznych uzyskała w Szkole Głównej Handlowej w Warszawie. Posiada wieloletnie doświadczenie w planowaniu i wdrażaniu interwencji publicznych, zdobyte w trakcie pracy w administracji publicznej, w tym w Ministerstwie Rozwoju Regionalnego. Brała udział w przygotowaniu systemu zarządzania polityki spójności UE w Polsce na lata 2007-2013 oraz była odpowiedzialna za budowanie odpowiedniej zdolności administracyjnej do zarządzania i wdrażania funduszy unijnych w tym okresie. Ponadto, realizowała projekty rozwojowe w Tunezji, badania ewaluacyjne w Afryce oraz projekty naukowe i watchdogowe, nastawione na upowszechnianie idei polityk opartych na wiedzy. Autorka artykułów i publikacji w zakresie procesu europeizacji polskiej administracji publicznej, wieloszczeblowego rządzenia, planowania strategicznego, wykorzystania wiedzy płynącej z ewaluacji. Prowadzi szkolenia dotyczące m.in. polityki spójności UE, planowania strategicznego czy partycypacji publicznej.

### Paulina Skórska

---

Absolwentka socjologii i psychologii Uniwersytetu Jagiellońskiego w Krakowie. Stypendystka i absolwentka programu *Educational measurement*. Ukończyła szkoły z zakresu wykorzystania metod kontrfaktycznych - *Summer School on Counterfactual Impact Evaluation* (2013) i *Fundamentals and Methods for Impact Evaluation of Public Policies* (2014). Absolwentka *ICPSR Summer Program in Quantitative Methods of Social Research* (2014, 2015) na Uniwersytecie Michigan. Dyrektor ds. badań i statystycznej analizy danych w Fundacji IDEA Rozwoju. Posiada doświadczenie w projektowaniu i prowadzeniu badań społecznych i ewaluacyjnych, zwłaszcza realizowanych na rzecz administracji publicznej. Od 2013 r. w Instytucie Badań Edukacyjnych pełni funkcję eksperta ds. metodologii i statystycznej analizy danych. Współpracuje od kilku lat z Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego, jako koordynator i ekspert w projektach z obszaru polityk publicznych, ukierunkowanych na rozwój regionalny, planowanie strategiczne oraz rozwój systemu oświaty. Od 2014 r. konsultuje budowę systemu ewaluacji oświaty dla agencji rządowej Królestwa Arabii Saudyjskiej (*Public Education Evaluation Commission*). Szkoli pracowników administracji publicznej i prowadzi zajęcia akademickie z zakresu metodologii badań społecznych.

## Bibliografija:

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. i in., "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, Nowy Jork 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. Komisja Europejska, *Monitoring and evaluation of European cohesion policy: concepts and recommendations*, Bruksela 2011.
7. Komisja Europejska, *Evalsed Sourcebook: Methods and Technics*, Bruksela 2013. Pobrano z [http://ec.europa.eu/regional\\_policy/en/newsroom/news/2013/11/evaluation-guidance-evalsed-guide-and-sourcebook-updates](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, Londyn 1997.
13. Riché M., *Theory Based Evaluation: A wealth of approaches and an untapped potential*, Komisja Europejska 2012.
14. Rogers P.J., *Program theory evaluation: Not whether programs work but how they work*, w: D.L. Stufflebeam i in. (red.), *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. i in., *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*, w: J. Connell I in. (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.

## Porównawcza analiza jakościowa w ewaluacji wpływu opartej na teorii

### Wprowadzenie

Celem niniejszego rozdziału jest przedstawienie jakościowej analizy porównawczej (*qualitative comparative analysis*, QCA) jako techniki, która może stanowić istotne uzupełnienie metod badawczych wykorzystywanych w ewaluacji wpływu opartej na teorii. Specyfika tej metody, leżące u jej podstaw założenia, a także schemat postępowania badawczego sprawiają, że może ona być z powodzeniem wykorzystywana wszędzie tam, gdzie interesuje nas odpowiedź na pytanie, jakie czynniki są decydujące dla powodzenia interwencji publicznej, jaką rolę odgrywa ta interwencja w uzyskiwaniu pożądanych zmian czy wreszcie – jak kontekst i uwarunkowania zewnętrzne wpływają na efekty uzyskiwane w wyniku jej podjęcia. Bez wątplenia więc charakter pytań, na jakie pozwalają odpowiedzieć wyniki QCA, lokuje tę technikę na ważnym miejscu w paście metod, które warto wykorzystać w ewaluacji wpływu.

Jakościowa analiza porównawcza jest przykładem techniki, która łączy podejście jakościowe z ilościowym. Bazując na pogłębionej analizie pojedynczych przypadków i na dobrej ich znajomości, przechodzi do analizy porównawczej, zmierzającej do identyfikacji konfiguracji czynników, które można określić mianem warunków koniecznych lub wystarczających dla wystąpienia danego efektu. To wydawałoby się proste połączenie zalet metod jakościowych (właściwa dla metod idiograficznych<sup>62</sup> dogłębna znajomość przypadku) z zasadami logiki formalnej (algebra Boole'a)<sup>63</sup> pozwala uzyskać rezultaty w dużej mierze przewyższające ograniczenia metod jakościowych i ilościowych. Takim ograniczeniem w przypadku metod jakościowych jest ich niekonkluzywność i nadmierna koncentracja na opisie danych, zaś w przypadku

<sup>62</sup> Metody idiograficzne zorientowane są na szczegółowy opis pojedynczego przypadku, dążąc do pełnej rekonstrukcji przyczyn i uwarunkowań występowania danego zjawiska. Zazwyczaj stosuje się w nich jakościowe metody badań. Często też metody jakościowe jako takie określa się mianem idiograficznych. Podejście idiograficzne jest przeciwstawiane podejściu nomotetycznemu, zorientowanemu na ogólne wyjaśnienie danej kategorii zjawisk z wykorzystaniem niewielu czynników przyczynowych (por. E. Babbie, *Badania społeczne w praktyce*, PWN, Warszawa, 2004).

<sup>63</sup> Algebra Boole'a to algebra ogólna stosowana w matematyce, informatyce teoretycznej oraz elektronice cyfrowej. Jej nazwa pochodzi od nazwiska matematyka, filozofa i logika George'a Boole'a. W algebrze Boole'a mamy do czynienia z danymi binarnymi, które mogą przyjąć tylko jedną z dwóch możliwych wartości: 1 – prawda, 0 – fałsz. Na tych danych można wykonywać trzy podstawowe i bazowe dla QCA operacje logiczne – negacja/zaprzeczenie (NIE); alternatywa/suma logiczna (LUB); koniunkcja/iloczyn logiczny (I) (por. B. Rihoux, Ch. C. Ragin (red.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, Thousand Oaks, CA, Sage 2009).

metod ilościowych – mechaniczne podejście do analizowanych przypadków i pomiaranie ich specyfiki<sup>64</sup>.

Ponieważ w Polsce do tej pory QCA w badaniach ewaluacyjnych stosowano bardzo rzadko<sup>65</sup>, warto przybliżyć czytelnikom ogólne założenia tej metody, przedstawić procedurę postępowania analitycznego związanego z jej stosowaniem oraz przykładowe wyniki, jakie przynosi. Realizacji takich celów podporządkowana jest struktura niniejszego rozdziału, który podzielono na trzy zasadnicze części. W pierwszej przedstawiamy założenia teoretyczne QCA, w tym koncepcję przyczynowości opartej na współwystępowaniu wielu czynników, w drugiej opisujemy QCA jako podejście badawcze, w trzeciej zaś przedstawiamy procedurę badawczą związaną z zastosowaniem tej metody. Mamy nadzieję, że informacje zawarte w tym rozdziale zachęcą czytelników do samodzielnych poszukiwań i do podjęcia próby zastosowania QCA w badaniach ewaluacyjnych<sup>66</sup>.

### **Teoria programu i złożona przyczynowość – punkty wspólne ewaluacji wpływu opartej na teorii i jakościowej analizie porównawczej**

Zanim przejdziemy do przedstawienia specyfiki jakościowej analizie porównawczej i etapów postępowania badawczego związanych z zastosowaniem tej metody, warto zwrócić uwagę na funkcję, jaką może ona pełnić w ewaluacji wpływu opartej na teorii. Według wskazań Komisji Europejskiej pytania, jakie należy uwzględnić podczas oceny efektów polityk publicznych, można podzielić na dwie ogólne kategorie: pierwsza dotyczy kwantyfikacji tych efektów, druga – ich wyjaśnienia<sup>67</sup>. Tym dwóm ogólnym rodzajom pytań odpowiadają dwa ogólne typy metod badawczych. Na pytania dotyczące tego, czy dana interwencja przynosi założone efekty w określonych warunkach,

---

<sup>64</sup> Twórcą jakościowej analizie porównawczej Charles Ragin lokuje ją wśród metod zorientowanych na przypadek (*case oriented methods*), odróżniając je od metod zorientowanych na zmienne (*variable oriented methods*). Pierwsze cechuje to, że pozostają niejako blisko przypadków, pozwalając widzieć ich złożoność i specyfikę. Drugie natomiast w niewielkim stopniu skupiają się na przypadkach (osobach, programach, organizacjach, przedsiębiorstwach itp.), interesując się jedynie zmiennymi i relacjami między nimi. Por. Ch. C. Ragin, H. S. Becker (red.), *What is a case. Exploring the foundation of social inquiry*, Cambridge University Press, Cambridge, New York 1992.

<sup>65</sup> Jedyna znana nam próba zastosowania QCA w ewaluacji wpływu w Polsce miała miejsce w projekcie „SPIN – Model transferu innowacji w Małopolsce”, realizowanym przez Centrum Ewaluacji i Analiz Polityk Publicznych UJ. Jako dane wejściowe zamierzano wykorzystać wyniki studiów przypadków. Jednakże ze względu na małą liczbę tych przypadków (4) i dużą liczbę zmiennych zrezygnowano z zastosowania tej techniki, zamieniając ją na mniej sformalizowaną porównawczą analizę przypadków. Szerzej na ten temat w: D. Szklarczyk i in., *Configurational Analysis in the Evaluation of Complex Public Programs: Application in the Area of Knowledge Transfer*, w: E. S. C. Berger, A. Kuckertz (red.), *Complexity in Entrepreneurship, Innovation and Technology Research*, Springer International Publishing, 2016, s. 371-395.

<sup>66</sup> W takich poszukiwaniach bardzo pomocna może być strona internetowa <http://www.compass.org/>, zawierająca zarówno pełną bibliografię metody QCA, teksty źródłowe, jak i darmowe oprogramowanie wykorzystywane w analizie.

<sup>67</sup> Komisja Europejska *Evalsed Sourcebook: Method and Techniques*, Bruksela 2013, s. 47, [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/evaluation\\_sourcebook.pdf](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf) (dostęp 24.02.2017 r.).



odpowiadają przede wszystkim metody kontrfaktyczne, zorientowane na wyabstrahowanie swobodnego efektu interwencji z całości zmian zaobserwowanych w danej grupie, organizacji czy na danym obszarze. Podejście ewaluacyjne wykorzystujące te metody określa się jako *conterfactual impact evaluation* (CIE). Druga kategoria metod, związana z ewaluacją wpływu opartą na teorii (*theory based impact evaluation*, TBIE), zorientowana jest przede wszystkim na zrozumienie, w jaki sposób interwencja prowadzi do wystąpienia zamierzonych i niezamierzonych efektów oraz jak efekty interwencji może różnicować kontekst, w którym ma ona miejsce. W przypadku tych metod najważniejszym celem badań jest odpowiedź na pytanie: dlaczego i w oparciu o jakie mechanizmy interwencja wywołuje określone efekty? Kluczowa jest identyfikacja teorii zmiany, która znajduje się u podstaw tej interwencji, oraz ocena efektów jej implementacji w danym kontekście<sup>68</sup>.

Tym, co łączy oba te podejścia – *conterfactual impact evaluation* i *theory-based impact evaluation* – jest centralne miejsce, jakie zajmuje w nich przyczynowość<sup>69</sup>. Sposób rozumienia i analizy przyczynowości w każdym z tych podejść jest jednak inny<sup>70</sup>. W przypadku CIE kluczowa jest identyfikacja związku przyczynowo-skutkowego pomiędzy interwencją a jej efektem, streszczająca się w odpowiedzi na pytanie, czy i w jakim stopniu uzyskane efekty mogą być przypisane do programu (*attribution*). Natomiast w przypadku TBIE na pierwsze miejsce wysuwa się odpowiedź na pytanie, czy analizowana interwencja jest, czy nie jest jedną z przyczyn obserwowanych zmian (*causal contribution analysis*)<sup>71</sup>. Zdaniem Sterna i innych badaczy można nawet mówić o stopniowym przechodzeniu od koncentracji na pytaniach zorientowanych na atrybucję (*attribution-oriented questions*) do pytań zorientowanych na kontrybucję (*contribution-oriented questions*)<sup>72</sup>. Zainteresowanie badaczy zwraca się też coraz częściej w kierunku analizy mechanizmów przyczynowych, czego dobrym przykładem jest rozwinię-

---

<sup>68</sup> Tamże.

<sup>69</sup> 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, s. 8.

<sup>70</sup> Sama przyczynowość i jej rozumienie w badaniach ewaluacyjnych są obecnie szeroko dyskusyjne. Por. np.: 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, s. 105-117, S. I. Donaldson, C. A. Christie, M. M. Mark (red.), *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, s. 11-24, E. Stern, O. W. Andersen, H. Hansen, „Editorial: Special issue: What can case studies do?” *Evaluation* 2013, 19 (3), s. 213-216, E. Gates, L. Dyson, „Implications of the Changing Conversation About Causality for Evaluators”, *American Journal of Evaluation* 2016, volume: 38 issue: 1, s. 29-46.

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

<sup>72</sup> 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.

ta przez Pawsona i Tilley'go koncepcja ewaluacji realistycznej, koncentrującej się na opracowaniu i weryfikacji teorii dotyczącej tego, w jaki sposób określone mechanizmy działają w zróżnicowanych kontekstach<sup>73</sup>. Te zmiany akcentów mają swoje implikacje dla metodologii ewaluacji – obserwuje się wzrastające zainteresowanie innymi niż eksperymentalne metodami i podejściami badawczymi pozwalającymi na analizę związków przyczynowo-skutkowych. Wykorzystuje się więc na coraz szerszą skalę zarówno nieeksperymentalne metody ilościowe – jak na przykład modelowanie strukturalne czy metody ekonometryczne – jak i jakościowe, pozwalające wyjaśniać mechanizmy przyczynowo-skutkowe w obrębie pojedynczego przypadku (np. śledzenie procesów) lub wielu przypadków (np. jakościowa analiza porównawcza)<sup>74</sup>.

Ponieważ to ta ostatnia jest przedmiotem naszego zainteresowania, warto wskazać czynniki, które sprawiają, że może ona być szczególnie przydatna w ewaluacji wpływu opartej na teorii. Do czynników tych należą:

1. Centralna rola teorii, rozumianej jako wskazanie relacji przyczynowo-skutkowych zachodzących pomiędzy czynnikami przyczynowymi a efektami programu. Co oczywiste, teoria znajduje się zarówno u podstaw TBIE, jak i w centrum QCA. Bez opracowania adekwatnego modelu przyczynowego trudno jest prowadzić analizę. Za pomocą QCA można więc testować teorię programu znajdującą się u podstaw danej interwencji.
2. Właściwe zarówno dla TBIE, jak i QCA dążenie przede wszystkim do identyfikacji mechanizmów decydujących o sukcesie bądź porażce danego przedsięwzięcia, a nie do pomiaru wielkości efektów.
3. Dostrzeganie znaczenia kontekstu i lokalnych uwarunkowań jako czynników decydujących o sukcesie bądź porażce interwencji. Wiąże się to z zakwestionowaniem możliwości istnienia uniwersalnej, niezależnej od kontekstu teorii programu. W QCA uwidacznia się to w dążeniu do identyfikacji różnych ścieżek prowadzących do wystąpienia określonych efektów. W zależności od kontekstu i lokalnych uwarunkowań ścieżki te mogą być różne.
4. Zwrócenie uwagi na możliwość osiągania założonych efektów za pomocą różnych działań czy czynników. W tym kontekście w QCA mówi się o ekwifinalności – wskazując na to, że do występowania tych samych efektów mogą prowadzić różnorakie ścieżki.
5. Zwracanie uwagi nie tylko na teorię programu, lecz i na teorię wdrożenia – QCA można wykorzystać jako technikę pozwalającą sprawdzić, jak różne czynniki charakteryzujące sposób wdrażania interwencji mogą wpływać na jej efekty.

---

<sup>73</sup> R. Pawson, N. Tilley, *Realistic evaluation*, Sage, London 1997, R. Pawson, N. Tilley, *Realist evaluation*, [http://www.communitymatters.com.au/RE\\_chapter.pdf](http://www.communitymatters.com.au/RE_chapter.pdf) (dostęp 25.02.2017 r.).

<sup>74</sup> E. Gates, L. Dyson, *op.cit.*, s. 3.

Przyczynowość oparta na współwystępowaniu wielu czynników – podstawowe założenie jakościowej analizy porównawczej.

Jak piszą Gates i Dyson, poszczególne metody służące ocenie efektów i wpływu polityk publicznych oparte są na leżących u ich podstaw różnych sposobach rozumienia przyczynowości<sup>75</sup>. Zdaniem tych autorów można wskazać co najmniej pięć różnych sposobów ujmowania przyczynowości:

1. podejście akcentujące następstwo w czasie, które uwzględni regularność i logikę kontrfaktyczną (*successionist framework that underlies regularity and counterfactual logics*);
2. podejście bazujące na analizie narracji interesariuszy (*narrative stakeholder accounts*);
3. podejście generatywne, uwzględniające analizę procesów i mechanizmów (*generative accounts of processes and mechanisms*);
4. podejście analizujące zestawy czynników przyczynowych i kontrybucję (*causal packages and contributory accounts*);
5. podejście uwzględniające nielinearność, wielokierunkowość i dynamikę relacji występujących w złożonym systemie (*nonlinear, multidirectional, and dynamical accounts of relations as found in complex systems*)<sup>76</sup>.

Rozumienie przyczynowości znajdujące się u podstaw QCA można zaliczyć do trzeciej i czwartej spośród wymienianych wyżej kategorii podejść. Podstawowym założeniem tej metody jest bowiem koncepcja złożonej przyczynowości – współwystępowania wielu czynników, które mogą być albo nie być warunkami koniecznymi lub wystarczającymi do wywołania określonych efektów. Podejście to pozwala więc analizować rolę poszczególnych elementów interwencji lub konfiguracji tych elementów w uzyskaniu jej efektów. Można je też wykorzystać do analizy wpływu poszczególnych instrumentów na występowanie określonego rodzaju efektów, np. wpływu dotacji inwestycyjnych dla przedsiębiorstw i dotacji rozwojowych na wzrost innowacyjności tych przedsiębiorstw. Może ono też być pomocne w identyfikacji tych czynników, które wpływają na skuteczność realizacji danej interwencji. Jego zastosowanie może więc pokazać, że np. interwencja jest skuteczna w tych przypadkach, gdy jej wprowadzenie poprzedzono odpowiednim przygotowaniem odbiorców, a nieskuteczna wtedy, gdy takiego przygotowania zabrakło. QCA może być też użyteczna wówczas, gdy chcemy pokazać, jak dane mechanizmy działają w różnych kontekstach, prowadząc do wystąpienia określonych efektów. Odpowiednie zastosowanie tej analizy pozwala bowiem odpowiedzieć na pytania, w jaki sposób, dla czego, dla kogo i pod jakimi warunkami dana interwencja będzie prowadzić do wystąpienia określonych efektów. Ten sposób myślenia, właściwy dla trzeciej spośród wymienionych wyżej koncepcji przyczynowości oraz bliski realistycznej ewaluacji, zakłada, że istnieje wiele ścieżek przyczynowych łączących interwencję z jej efektem.

---

<sup>75</sup> Tamże, s. 7.

<sup>76</sup> Tamże.

Takie podejście do rozumienia przyczynowości stanowi podstawę QCA. W metodzie tej używa się pojęcia „złożonej przyczynowości” czy – jak nazywają to Ragin i Rihoux – „przyczynowości opartej na współwystępowaniu wielu czynników” (*multiple conjunctural causation*)<sup>77</sup>. Istotą tak ujmowanej przyczynowości jest zwrócenie uwagi na fakt, że różne kombinacje czynników czy też różne ścieżki przyczynowe mogą prowadzić do wystąpienia tych samych efektów. Charakterystyczna jest tutaj zasada ekwifinalności (*equifinality*), co oznacza, że każda z tych ścieżek jest tak samo istotna, czyli że mechanizmy prowadzące do wystąpienia tych samych efektów mogą być różne. W tym rozumieniu przyczynowości istotne jest zarówno pojęcie *multiple*, jak i *conjunctural* – *multiple* odnosi się do liczby ścieżek, zaś *conjunctural* do tego, że każda ścieżka zawiera kombinację czynników. Streszczenie istoty założeń przyczynowości opartej na współwystępowaniu wielu czynników, znajdującej się u podstaw jakościowej analizy porównawczej, przedstawia tabela 1.

**Tabela 1. Najważniejsze założenia koncepcji wielokrotnie złożonej przyczynowości (*multiple conjunctural causation*)**

The most important assumptions of the concept of multiple conjunctural causation

1. Efekt powstaje najczęściej w wyniku kombinacji istotnych czynników przyczynowych ( $AB \rightarrow Y$ )
2. Kilka różnych kombinacji czynników może prowadzić do wystąpienia tego samego efektu ( $AB + CD \rightarrow Y$ )
3. W zależności od kontekstu efekt może powstać wówczas, gdy dany czynnik jest obecny lub gdy jest nieobecny ( $AB \rightarrow Y$ , ale również  $aC \rightarrow Y$ ). W tym przypadku [A] występujące razem z [B] prowadzi do wystąpienia wyniku [Y], lecz także jego nieobecność [a] występująca z [C] może prowadzić do wystąpienia efektu [Y]

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

Stosując jakościową analizę porównawczą, badacz nie buduje jednego modelu przyczynowego, który najlepiej pasuje do danych, jak to ma miejsce w przypadku większości technik statystycznych, lecz identyfikuje liczbę i charakter różnych modeli przyczynowych, które występują w porównywanych przypadkach<sup>78</sup>. Ponadto jakościowa analiza porównawcza koncentruje się na wyjaśnieniu, jak powstaje dany efekt, a nie

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

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

– jak ma to miejsce w przypadku technik opartych na regresji – na analizie wpływu netto danego czynnika przyczynowego na występowanie tego efektu.

## **Podejście badawcze jakościowej analizy porównawczej**

Pomimo że opracowana przez Ragina QCA<sup>79</sup> jest wykorzystywana na świecie w różnych obszarach badań od końca lat 80., w Polsce wciąż jest mało popularna. W literaturze polskojęzycznej trudno znaleźć publikacje zawierające sam opis metody czy przykłady jej zastosowania w badaniach<sup>80</sup>. Lukę tę niewątpliwie warto wypełnić, gdyż metoda ta może być cennym uzupełnieniem podejść analitycznych służących do wyjaśniania mechanizmów przyczynowych zachodzących zjawisk. W dalszej części przedstawimy w syntetyczny sposób podstawowe zasady metodologiczne projektowania badań i analizy w oparciu o założenia metody QCA, wyjaśniając przy tym najważniejsze pojęcia stosowane w tym podejściu.

Jak wspomniano wcześniej, jakościowa analiza porównawcza służy ogólnie do wnioskowania o złożonych przyczynach określonych zjawisk. Jednak w odróżnieniu od klasycznych metod wnioskowania przyczynowego, np. regresji liniowej, modelowanie opiera się nie na układach zmiennych i powiązaniach między nimi, lecz na teorii zbiorów i analizie relacji pomiędzy obiektami należącymi do tych zbiorów i ich wpływie na określony wynik (*output*). Celem jest zbudowanie wyjaśnienia pewnego zjawiska za pomocą konfiguracji zależności przyczynowych, które mogą do niego prowadzić. Nie bez powodu wspomnieliśmy o regresji liniowej, gdyż QCA dzieli z nią logikę postępowania, jednocześnie w wielu kwestiach przyjmując inne rozstrzygnięcia. Dlatego też, aby ułatwić czytelnikowi zrozumienie QCA, czasami będziemy się odwoływać do porównań z regresją liniową jako techniką stosunkowo dobrze znaną.

Na potrzeby wyjaśnienia ogólnych zasad QCA weźmy przykład, w którym chcemy ocenić skuteczność różnych programów działań mających na celu wprowadzenie osób nieaktywnych zawodowo na rynek pracy. Efektem takich działań byłby w uproszczeniu sukces lub porażka<sup>81</sup> i to determinowałoby przynależność do zbioru wynikowego – tzn. efekt programu przyjmowałby dwie wartości. W toku prowadzonych analiz chcielibyśmy, w oparciu o pewne kryteria charakteryzujące poszczególne programy, określić, co decyduje o ich sukcesie. Tutaj pojawia się pierwsza zasadnicza różnica w porówna-

<sup>79</sup> Tamże.

<sup>80</sup> Nie znaczy to, iż metoda ta nie jest w Polsce w ogóle znana i stosowana. Sporadycznie wykorzystuje się ją np. w badaniach związanych z zarządzaniem, czego przykładem jest publikacja: A. Kwiotkowska, „Jakościowa analiza porównawcza jako koncepcja metodologiczna w naukach o zarządzaniu”, w: J. Lichtarski, S. Nowisielski, G. Osbert-Pociecha, E. Tabaszewska-Zajbert (red.), *Nowe kierunki w zarządzaniu przedsiębiorstwem – wiodące tendencje, Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 2014, nr 340, s. 65-77.

<sup>81</sup> Nie wchodzimy tutaj w szczegóły tego, jak jest oceniany sukces bądź porażka całego programu. W rzeczywistości wskaźnikiem jednego bądź drugiego byłby odsetek nieaktywnych zawodowo wprowadzonych na rynek, przy czym punkt odcięcia wyznaczający sukces musiałby być wyznaczony w oparciu o określone kryteria. Problemem tym zajmiemy się jeszcze, omawiając tzw. kalibrację przynależności do zbiorów.

niu do regresji – ta ostatnia pozwala modelować tylko jeden określony stan zmiennej wynikowej (byłby to prawdopodobnie sukces programu), natomiast QCA jest metodą asymetryczną i pozwala na wnioskowanie zarówno o jednym, jak i o drugim wyniku programu. Oznacza to, że możemy również sprawdzić, co decyduje o porażce programu. W podejściu QCA wręcz zalecane jest sprawdzenie tego, jakie czynniki prowadzą do każdego wyniku, gdyż może się okazać, że zgodnie z zasadami złożonej przyczynowości różne konfiguracje warunków prowadzić będą do innych rezultatów. W naszym przypadku różne mogą być przyczyny porażki czy sukcesu programów aktywizacji osób na rynku pracy. Założmy następnie, że analizowane programy oceniane byłyby pod względem trzech cech: tego, czy osobom, do których były skierowane, oferowane było doradztwo zawodowe; czy beneficjentami były osoby długotrwale nieaktywne (powyżej 24 miesięcy), czy też krócej; oraz czy w ramach programów oferowano im staż u pracodawców. W efekcie analizowane jest, jak te trzy warunki lub ich rozmaite konfiguracje przyczyniają się do sukcesu lub porażki programu. Jednym z pierwszych kroków w modelowaniu QCA jest sporządzenie tzw. tabeli prawdy (*truth table*). Pokazuje ona wszystkie kombinacje czynników wraz z osiąganym efektem. Ogólnie w takiej prostej tabeli jedynki oznaczają przynależność do danego zbioru (występowanie danego stanu), a zera – brak takiej przynależności.

**Tabela 2. Przykładowa „tabela prawdy” uwzględniająca efekty pięciu programów wsparcia osób nieaktywnych zawodowo w postaci powrotu na rynek pracy** (dane fikcyjne)

Program	Doradztwo	Okres nieaktywności	Staż	Efekt programu
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

Źródło: opracowanie własne.

W przedstawionym powyżej (tabela) najprostszym rozwiązaniu przynależność do zbiorów wyznaczonych przez atrybuty poszczególnych czynników i wyniku programów określona jest dychotomicznie (0-1) – dany program oferował uczestnikom doradztwo (1) bądź też nie (0), skierowany był do długotrwale nieaktywnych (1) lub też nie (0), oferowany był w nim staż (1) bądź też nie (0) i zakończył się sukcesem (1) albo porażką (0). Celem wnioskowania jest określenie dwóch relacji pomiędzy czynnikami i wynikiem, analizowanymi na podstawie odpowiadających im zbiorów – tego, czy dany czynnik (lub kombinacja czynników) jest warunkiem wystarczającym lub koniecznym dla wystąpienia rozpatrywanego efektu. To, że jakiś czynnik lub kombi-

nacja kilku stanowią warunek wystarczający do pojawienia się wyniku, stwierdzamy wówczas, jeśli każdemu wystąpieniu czynnika odpowiada wystąpienie rozpatrywanego zjawiska (jeżeli  $x$  to  $y$ ). W naszym przykładzie z taką sytuacją mamy do czynienia w relacji pomiędzy stażem i sukcesem programu – zawsze, jeśli oferowany był staż, program kończył się sukcesem. W przypadku pozostałych czynników nie możemy mówić, że były one wystarczające, aby program zakończył się sukcesem. Natomiast czynnik jest warunkiem koniecznym wystąpienia określonego efektu programu w sytuacji, kiedy musi on być obecny dla wystąpienia danego efektu – efekt występuje wyłącznie w sytuacji wystąpienia określonego warunku lub kombinacji warunków ( $y$  pociąga za sobą  $x$ ). W rozpatrywanym przypadku takim warunkiem koniecznym byłby również staż, gdyż zawsze, kiedy program kończył się sukcesem, oferowany był staż. Można zatem powiedzieć, że analiza pięciu programów prowadzi do wniosku, że warunkiem wystarczającym i koniecznym, aby interwencja zakończyła się wprowadzeniem osób nieaktywnych na rynek pracy, był staż oferowany uczestnikom programu.

Oczywiście przykład ten jest znacznym uproszczeniem pokazującym ogólne mechanizmy QCA, które w rzeczywistości dają znacznie bogatsze możliwości wnioskowania o przyczynowości analizowanych zjawisk<sup>82</sup>. Wykorzystując algebrę Boole'a w odniesieniu do analizowanych zbiorów, można uzyskać szereg konfiguracji warunków prowadzących do określonego wyniku. Uwzględniając asymetryczność analiz w ramach QCA, można również wprowadzić negacje rozpatrywanych warunków lub nawet samego wyniku, co jest niewątpliwą zaletą w porównaniu np. do modelowania regresyjnego<sup>83</sup>. W omawianym przypadku rozpatrywaliśmy jedynie tzw. zbiory klasyczne (*crisp set*) i odpowiadające im czynniki binarne, gdzie przynależność do zbiorów warunków była wyznaczana w oparciu o prostą klasyfikację dychotomiczną – czy określony obiekt posiada daną cechę, czy też nie. W rozszerzonej formie QCA, zwanej *fsQCA* (od *fuzzy set* QCA – QCA na zbiorach rozmytych), przynależność obiektów do zbiorów warunków określana jest gradacyjnie i może przyjmować wiele poziomów, choć zazwyczaj stosuje się porządek trzy-, cztero- lub pięcioelementowy. W naszym przypadku taki zbiór rozmyty warunku „okres nieaktywności zawodowej” mógłby przyjmować na przykład trzy stany: program skierowany do osób nieaktywnych na rynku pracy mniej niż rok, od roku do dwóch i więcej niż dwa lata.

Do praktycznego zastosowania QCA we wnioskowaniu przyczynowym można wykorzystać liczne programy komputerowe opracowane w tym celu. W większości są one

<sup>82</sup> Po szczegółowe wyjaśnienia różnego rodzaju relacji, jakie mogą zachodzić pomiędzy warunkami i wynikiem – w tym tzw. warunków INUS i SUIN – odsyłamy do: 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.

<sup>83</sup> Wprawdzie w regresji również można uwzględniać kombinacje zmiennych niezależnych jako czynników interakcyjnych, ale przy takim złożeniu już przy kilku zmiennych interpretacja takiego czynnika interakcyjnego jest bardzo utrudniona.

darmowe, przykładowo można wymienić opracowany przez samego Ragina fsQCA<sup>84</sup>. Jego rozwojową wersją stosowaną dla wielokategorialnych (*multivariate*) zbiorów wynikowych, czyli dopuszczających, że wnioskowany wynik może przyjmować więcej niż dwa stany, jest opracowana przez Cronqvista Tosmana<sup>85</sup>. Oprócz tego istnieją odpowiednie moduły dla środowiska R i pakietu Stata<sup>86</sup>.

## Procedura QCA

W dalszej części przedstawiamy krok po kroku procedurę QCA ze zwróceniem uwagi na kluczowe momenty prowadzonych analiz, a następnie prezentujemy wyniki, jakie można uzyskać za pomocą tej techniki.

### Tabela 3. Etapy QCA

1. Budowa modelu w oparciu o odpowiednie przypadki i układy przyczynowe
2. Przygotowanie bazy danych z kalibracją zbiorów czynników (kodowanie)
3. Opracowanie tabeli prawdy i sprawdzenie wyników pod kątem spójności i występujących niezgodności czy sprzecznych warunków
4. Analiza tabeli prawdy z wykorzystaniem algorytmów minimalizujących układy przyczynowe
5. Ocena uzyskanych wyników
6. Powrót do kroku 3. i powtórzenie analiz dla negacji zbioru wynikowego

Źródło: opracowanie własne.

Przed przystąpieniem do analizy jak w przypadku każdego modelowania należy oczywiście osadzić całą procedurę w określonym kontekście teoretycznym, czyli znaleźć uzasadnienie dla przeprowadzonego wnioskowania. Najczęściej więc analizę zaczynamy od zbudowania modelu, ale w przeciwieństwie do klasycznego modelowania regresyjnego wymagane tu jest znacznie bardziej elastyczne, jakościowe podejście. Chodzi bowiem o odpowiedni dobór przypadków do analizy, aby można było w oparciu o nie wyjaśnić określony wynik za pomocą układów przyczynowych. Dużą zaletą QCA jako podejścia mieszanego jest możliwość jednoczesnego wykorzystania danych ja-

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

<sup>85</sup> L. Cronqvist, *Tosmana: Tool for small-n analysis, Version 1.522* [program komputerowy], University of Trier, Trier 2016, <https://www.tosmana.net/> (dostęp 25.05.2017 r.)

<sup>86</sup> Omówienie różnych rozwiązań można znaleźć w: 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, s. 505-521.



kościowych i ilościowych. Przykładowo w celu wyjaśnienia czynników warunkujących innowacyjność przedsiębiorstw za pomocą QCA można wykorzystać dane ilościowe zebrane w trakcie wywiadów standaryzowanych oraz uzupełnić je informacjami pozyskanymi w trakcie wywiadów pogłębionych z przedstawicielami tych podmiotów. Do takiej analizy można również włączyć dane zastane z dostępnych statystyk publicznych, informujące o kondycji rozpatrywanych firm. Prowadzone wnioskowanie powinno przebiegać w oparciu o odpowiednio wybrane przypadki z całego zbioru wszystkich dostępnych. Nic nie stoi na przeszkodzie, aby przeanalizować wszystkie istniejące przypadki, należy jednak pamiętać o tym, aby były one odpowiednio zróżnicowane. W sytuacji wspomnianej analizy przyczyn innowacyjności firm, gdyby większość przypadków obejmowała podmioty, które są innowacyjne, to brak zróżnicowania zbioru tak wybranych przypadków utrudniłby lub wręcz uniemożliwił wykonanie wnioskowania. Jest to sytuacja analogiczna do modelowania regresyjnego, w którym niepożądane jest, aby zmienna zależna cechowała się małą wariancją (zróżnicowaniem wartości). W tym pierwszym kroku oprócz wybrania odpowiednich przypadków do analizy należy również zbudować odpowiedni model przyczynowy, który będzie podstawą do wyboru czynników mogących wpływać na sprawdzany wynik. To, jakie będą to czynniki, zdecyduje o sposobie ich kalibracji w kolejnym kroku. Konstruowanie takiego modelu analitycznego można prowadzić na dwa sposoby. Po pierwsze, można zrobić to metodą aprioryczną, wychodząc od teorii i z niej wyprowadzając zależności, które później będą testowane. Po drugie, można również postępować eksploracyjnie i dopiero na etapie samej analizy budować model zależności przyczynowych. W praktyce z reguły łączy się te dwa sposoby – wychodzi się z pewnych założeń wynikających z natury badanych zjawisk, a w trakcie analizy sprawdza się, które czynniki najlepiej wyjaśniają przyczyny tych zjawisk.

Kalibracja czynników warunkujących wynik, jak i samego wyniku, jest po prostu ich kodowaniem według określonych reguł. Ten etap będzie wyglądał inaczej, kiedy chcemy przeprowadzić wnioskowanie QCA w oparciu o zbiory klasyczne (czyli csQCA), a inaczej w przypadku zbiorów rozmytych (czyli fsQCA). Przypomnijmy, że w tym pierwszym przypadku przynależność do zbioru ma charakter dychotomiczny – czyli rozpatrywane czynniki mogą posiadać cechę determinującą przynależność do tego zbioru lub jej nie posiadać. Na przykład jednym z czynników mogących wpływać na innowacyjność przedsiębiorstw jest posiadanie działu badawczo-rozwojowego. Wówczas czynnik ten przyjmie dwie wartości: 1 – posiadanie takiego działu – lub 0 – jego brak. W przypadku zbiorów rozmytych przynależność do zbioru jest stopniowalna – dany obiekt może do takiego zbioru należeć w większym, mniejszym stopniu lub w ogóle nie należeć. Prowadząc wnioskowanie oparte na zbiorach rozmytych, z reguły stosuje się trzy, cztery lub pięć poziomów gradacji przynależności do zbioru. W przypadku trzech stopni są one kodowane jako 1 – pełna przynależność, 0,5 – przynależność częściowa, 0 – brak przynależności. Natomiast kiedy chcemy zastosować cztery stopnie, to możliwe są wartości, które umownie można określić: 1 – pełna przynależność,

0,67 – średnia przynależność, 0,33 – niewielka przynależność, 0 – brak przynależności. To, czy zastosujemy trzy czy cztery poziomy, zależy od danych, jakimi dysponujemy, i natury samego czynnika. Dla przykładu: takim zbiorem rozmytym, który może być warunkiem innowacyjności przedsiębiorstw, jest wielkość podmiotu. Możemy wówczas zakodować te informacje według klasyfikacji wielkości prowadzonej działalności na firmy mikro (do 10 pracowników) jako „0”, firmy małe (od 11 do 49 zatrudnionych) jako „0,33”, firmy średnie (od 50 do 249 pracowników) jako „0,67” i wreszcie firmy duże (powyżej 250 zatrudnionych) jako „1”. Z jakichś przyczyn, najczęściej wynikających z przeprowadzonych analiz, moglibyśmy się zdecydować na pozostawienie tylko trzech stopni, powiedzmy: w wyniku połączenia firm mikro i małych oraz zakodowania ich jako „0”, pozostawienia firm średnich jako „0,5”, a dużych jako „1”. W tym przypadku kalibracja takiego czynnika jest stosunkowo prosta, można by powiedzieć: naturalnie wynikająca z charakteru samych danych. Nieco inaczej jest w przypadku cech o charakterze ilościowym (ciągłym). Wówczas można wykorzystać rozkład takiej cechy do określenia rozpiętości jej wartości i w oparciu o pewne kryteria określić wymagane punkty podziału (*threshold*). Wyznaczenie tych punktów podziału nie powinno być jednak automatyczne, lecz oparte na jakimś konkretnym modelu teoretycznym, np. podziale wieku na kohorty według obowiązującego standardu. Jeśli brak takiego koncepcyjnego punktu odniesienia, to można się oprzeć na populacyjnym rozkładzie cech w oparciu o wiarygodne statystyki publiczne.

Znacznie trudniej jest skalibrować zbiór rozmyty w oparciu o dane jakościowe. W takiej sytuacji nie można się bowiem posłużyć empirycznymi rozkładami cech i należy takie kodowanie przeprowadzić w oparciu o dobrze przemyślany schemat koncepcyjny. Dobry sposób postępowania zaproponowali w tym przypadku Basurto i Speer (tabela 3)<sup>87</sup>. Jest on oparty na sześciu krokach przedstawionych poniżej. Ten sposób postępowania został opracowany na potrzeby procedury gromadzenia danych jakościowych za pomocą wywiadów pogłębionych, ale może być równie dobrze zastosowany do kodowania jakościowych danych zastanych.

---

<sup>87</sup> X. Basurto, J. Speer, „Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA)”, *Field Methods* 2012, 24, s. 155-174.

#### Tabela 4. Procedura kalibrowania danych jakościowych

- Krok 1. Opis operacjonalizacji czynników i wyniku użytych w modelu
- Krok 2. Opracowanie wstępnych punktów odniesienia dla tworzonych zbiorów
- Krok 3. Kodowanie danych
- Krok 4. Sprawdzenie spójności kodów i ich dopasowania do zbiorów
- Krok 5. Ocena precyzji zbiorów i zdefiniowanych wartości punktów podziału
- Krok 6. Ostateczne przypisanie wartości kodowanym danym

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

Podsumowując zalecenia dotyczące kalibracji danych w QCA, należy podkreślić, że proces ten wymaga namysłu i powinien być oparty na przyjętym modelu teoretycznym bądź koncepcyjnym oraz odpowiadać potrzebom budowanego modelu. Dobry model znacznie ułatwia dobór samych czynników, ale też przypisanie im odpowiednich poziomów. Przy kalibracji czynników należy unikać zbytniego automatyzmu, gdyż od tego, jak przygotowujemy dane do analizy, w dużym stopniu zależą uzyskane wyniki. Szczególnie że przy prowadzeniu analizy na niewielkiej liczbie przypadków jakakolwiek zmiana danych może istotnie zmieniać cały model. Warto wypróbować różne konfiguracje czynników i ich poziomów oraz zobaczyć, jak wpływają one na wyniki wnioskowania. W każdym przypadku jednak ważne jest, aby postępować w sposób przejrzysty i dobrze opisać przyjęte zasady kalibracji, aby możliwa była ocena całej procedury i sprawdzenie jej trafności oraz rzetelności.

Po przygotowaniu danych można przystąpić do analizy. Pierwszą czynnością jest przygotowanie tabeli prawdy. Tak jak pisaliśmy wyżej, tabela ta zawiera wszystkie możliwe konfiguracje przyczynowe i odpowiadające im wyniki. Liczba wierszy w tej tabeli zależy od liczby warunków (czynników) wykorzystanych do zbudowania modelu. Przed przystąpieniem do minimalizacji konfiguracji przyczynowych należy sprawdzić taką tabelę pod kątem tzw. reszt logicznych (*logical reminders*) oraz sprzecznych konfiguracji warunków (*contradictory rows*). Reszty logiczne to po prostu takie konfiguracje warunków (czyli wiersze w tabeli prawdy), które nie występują w analizowanych przypadkach – są empirycznie puste. Natomiast sprzeczne konfiguracje pojawiają się w sytuacjach, kiedy układ warunków w wierszu tabeli prowadzi do wyniku, który może być jednocześnie prawdą lub fałszem (1 lub 0), dając niespójne rezultaty. Do diagnozowania takich sytuacji służy miara spójności (*consistency*), która – w dużym skrócie – jest oszacowaniem tego, jak dana konfiguracja warunków odbiega od idealnego

rozwiązania reprezentowanego przez wynik równy 0 lub 1. Przyjmuje się również, że sprzeczne konfiguracje to te o spójności mniejszej niż 0,75<sup>88</sup>. Zaleca się, aby problemy reszt logicznych i sprzecznych konfiguracji traktować indywidualnie i przed podjęciem decyzji o ich wykluczeniu z tabeli zweryfikować je w oparciu o analizę poszczególnych przypadków – czy takie sytuacje są możliwe i co za nimi stoi. Istnieją również trzy ogólne strategie, które mogą pomóc w przypadku napotkania takich problemów. Chodzi o: 1) dodanie przypadków do analizy – co może zasilić danymi reszty logiczne i usunąć niespójności, 2) dodanie dodatkowego warunku lub rozłożenie jednego warunku na kilka, 3) zmianę kalibracji warunków lub wyniku. Zalecenia te są podobne do spotykanych w problemach statystycznej analizy danych, np. w sytuacji braku danych lub niskich liczebności.

Kiedy tabela prawdy jest już wyczyszczona, można przystąpić do właściwej analizy, czyli minimalizacji tej tabeli za pomocą algorytmów algebry Boole'a (w przypadku csQCA) lub algorytmu Quine-McCluskey<sup>89</sup> (kiedy prowadzimy fsQCA). Minimalizacja sprowadza wszystkie konfiguracje warunków do uproszczonej postaci, łącząc i wykluczając podobne lub sprzeczne warunki. Dla przybliżenia działania QCA wystarczy powiedzieć, że w rezultacie tych analiz otrzymujemy trzy rozwiązania: złożone, oszczędne i pośrednie. W złożonym rozwiązaniu prezentowane są finalne ścieżki przyczynowe, przy wyznaczaniu których uwzględnione zostają wszystkie konfiguracje warunków, również te nieznajdujące pokrycia w danych empirycznych (czyli z uwzględnieniem reszt logicznych). W rozwiązaniu oszczędnym wręcz przeciwnie, reszty logiczne zostają wyłączone z analizy, co prowadzi do prostszych ścieżek przyczynowych. I wreszcie rozwiązanie pośrednie, które jest z reguły zalecane jako najlepsze, uwzględnia niektóre reszty logiczne, ale zgodnie z intencją użytkownika (to on decyduje, które z takich warunków włączyć).

Ostatecznie uzyskujemy ścieżki przyczynowe na podstawie czynników uwzględnionych w modelu, które wskazują konfiguracje warunków prowadzących do wystąpienia analizowanych efektów. Dla przykładu: w przypadku oszczędnego rozwiązania przyczyn leżących u podstaw rozwoju przedsiębiorstw uzyskać można następującą ścieżkę:

---

<sup>88</sup> Dokładne wyjaśnienie spójności i sprzecznych warunków można znaleźć w: C. Q. Schneider, C. Wagemann, *Set-Theoretic Methods...*

<sup>89</sup> Algorytmy opracowane przez Quine i McCluskeya są uznawane za najbardziej efektywne sposoby prowadzenia obliczeń w oparciu o zasady algebry Boole'a. Zostały opracowane w celu efektywnego wykorzystania komputerowych technik minimalizacji funkcji prawdy (por. 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).

R&D\*~PUBLIC SUPPORT  
 + SIZE\*STRATEGIC MANAGEMENT  
 + R&D\*RISK TAKING  
 → DEVELOPMENT

Należy ją rozumieć w ten sposób, że do rozwoju firm mogą prowadzić trzy alternatywne konfiguracje warunków: 1) posiadanie przez firmy działu badawczo-rozwojowego (R&D) i niewykorzystywanie pomocy publicznej (PUBLIC SUPPORT, gdzie „~” oznacza negację warunku), 2) w przypadku dużych firm (SIZE) – posiadanie strategicznego zarządzania (STRATEGIC MANAGEMENT – myślenie w kategoriach przyszłościowych) lub 3) posiadanie działu badawczo-rozwojowego (R&D) i gotowość do podejmowania ryzyka (RISK TAKING). Wyniki, zwłaszcza jeśli ścieżek przyczynowych jest wiele, można również przedstawić w postaci tabel zawierających informacje o wkładzie czynników w poszczególne ścieżki. Przykład takiej tabeli wynikowej uzyskanej w analizie przyczyn rozwoju firm przedstawiono poniżej (tabela 5).

**Tabela 5. Przykładowa tabela ścieżek przyczynowych rozwiązania pośredniego wyjaśniającego przyczyny rozwoju firm.**

Czynniki	Rozwiązanie 1	Rozwiązanie 2
Wielkość firmy ( <i>SIZE</i> )	●	○
Zarządzanie strategiczne ( <i>STRATEGIC MANAGEMENT</i> )	●	○
Zarządzanie kapitałem ludzkim ( <i>HC MANAGEMENT</i> )	●	●
Networking	●	
R&D ( <i>R&amp;D</i> )	●	○
Podejmowanie ryzyka ( <i>RISK TAKING</i> )	○	●

Uwaga: ● = główny czynnik przyczynowy (występuje); ○ = główny czynnik przyczynowy (nie występuje); ● = współuczestniczący czynnik przyczynowy (występuje); ○ = współuczestniczący czynnik przyczynowy (nie występuje).

W prezentowanym przykładowym wyjaśnieniu przyczyn rozwoju firm znalazły się dwa rozwiązania pośrednie. Składa się na nie szereg warunków wpływających w różnym stopniu na wynik wnioskowania. Na podstawie logiki wnioskowania można stwierdzić, że niektóre z nich będą występowały zarówno w rozwiązaniu oszczędnym, jak i pośrednim (dokładniej mówiąc, wszystkie warunki/czynniki z rozwiązania oszczędnego znaleźć się muszą w rozwiązaniu pośrednim). Takie warunki można nazwać głównymi (*core*) i oznaczone są w tabeli większymi okręgami. Natomiast inne dochodzą na etapie uzyskiwania rozwiązania pośredniego i współuczestniczą (*contributing*) w tak

utworzonych ścieżkach przyczynowych, przez co oznaczone zostały mniejszymi okręgami<sup>90</sup>. Na podstawie dwóch uzyskanych rozwiązań można stwierdzić, że pierwsze z nich dotyczy sytuacji dużych firm (obecność warunku), do których rozwoju przyczynia się przede wszystkim posiadanie strategii zarządzania kapitałem ludzkim i budowanie sieci powiązań z klientami i innymi podmiotami. Mniejszy udział ma natomiast wizja strategicznego zarządzania i posiadanie działu B+R. Co ważne, w tym rozwiązaniu w rozwoju dużych firm niewskazana jest tendencja do podejmowania ryzyka. Z kolei element ten występuje w drugim rozwiązaniu, właściwym dla małych podmiotów (brak warunku), które aby się rozwijać, nie muszą posiadać działu B+R ani wykazywać się myśleniem strategicznym, zarządzają natomiast kapitałem ludzkim. Można zatem powiedzieć, że główną różnicą w przyczynach rozwoju firm byłyby ich wielkość, która determinuje dwie ścieżki: obie oparte na zarządzaniu kapitałem ludzkim, ale w przypadku małych firm – opartą na gotowości do ryzyka, a w przypadku dużych – na posiadaniu bardziej formalnych strategii działania.

## Podsumowanie

Jakościowa analiza porównawcza jest jedną z metod mieszanych (*mixed methods*), czyli łączących podejście jakościowe i ilościowe. Może ona znaleźć szereg zastosowań w analizach przyczynowych, tak ważnych w ewaluacji wpływu. Dlatego bez wątplenia warto ją polecić do ewaluacji interwencji publicznych, zwłaszcza jeśli przeprowadzamy ewaluację wpływu i chcemy sprawdzić, jak w konkretnych warunkach sprawdza się teoria programu lub teoria wdrożenia. Położenie nacisku na teorię – czy to teorię programu, czy teorię wdrożenia – jest punktem wspólnym TBIE i QCA, a zastosowanie obu podejść wymaga od ewaluatora opracowania modelu wiążącego czynniki przyczynowe z ich skutkami. QCA możemy stosować wszędzie tam, gdzie chcemy zweryfikować trafność przyjętej teorii programu, zweryfikować jej adekwatność do danego kontekstu, przetestować, jak na skuteczność interwencji wpływa sposób jej wdrożenia. W odróżnieniu od klasycznych technik statystycznych, pozwalających określić efekty netto interwencji, QCA umożliwia zrekonstruowanie modelu złożonych układów przyczynowych, prowadzących do określonych rezultatów. Istotne jest również, że w wyniku zastosowania QCA możemy zidentyfikować różne ścieżki przyczynowe prowadzące do wystąpienia tych samych efektów – przykładowo odmienne mogą być ścieżki prowadzące do rozwoju przedsiębiorstw działających w różnych sektorach gospodarki czy w różnych krajach.

Warto podkreślić, że analizy przyczyn sukcesów, ale i porażek działań publicznych prowadzone za pomocą QCA nie muszą wiązać się z dużymi nakładami finansowymi ani czasowymi. Analizy przeprowadza się bowiem na małych i średnich próbach, liczących od 5 do 50 przypadków. Nie wymaga to więc prowadzenia szerokich i kosz-

---

<sup>90</sup> Szczegółowy opis tej metody prezentacji wyników QCA można znaleźć w: C. C. Ragin, P. C. Fiss, „Net effects analysis versus configurational analysis: An empirical demonstration”, w C. C. Ragin (red.), *Re-designing social inquiry: Fuzzy sets and beyond*, University of Chicago Press, Chicago 2008, s. 190-212.

townych badań na dużych, reprezentatywnych próbach. Do analizy możemy ponadto wykorzystać dane o różnorodnym charakterze – mogą to być informacje pochodzące ze statystyk publicznych, dane jakościowe uzyskane w wywiadach pogłębionych czy dane zebrane za pomocą technik sondażowych. Również oprogramowanie komputerowe stosowane do modelowania QCA jest powszechnie dostępne i stosunkowo proste w obsłudze. To wszystko sprawia, że metoda ta może być bardzo użytecznym narzędziem dla ewaluatorów oceniających wpływ danej interwencji nawet wówczas, gdy nie dysponują oni dużym budżetem czy wyspecjalizowanym zespołem badawczym.

Jakościowa analiza porównawcza mimo swoich licznych zalet stawia również przed badaczami wyzwania, które warto mieć na uwadze, decydując się na jej wykorzystanie. Przede wszystkim mimo względnej prostoty posługiwania się tym narzędziem wymaga ono dość sporego doświadczenia. Dotyczy to szczególnie doboru przypadków do analizy, budowania samego modelu przyczynowego oraz kodowania (kalibrowania) czynników uwzględnionych w modelu. Ponadto mimo precyzyjnych reguł wnioskowania, opartych na zasadach logiki i algorytmach matematycznych, wyniki tej analizy nie mają charakteru uniwersalnego. W przeciwieństwie do klasycznych modeli statystycznych nie można ich uogólniać na szersze populacje. Niemniej w wielu przypadkach poznanie układu przyczyn prowadzących do określonego rezultatu może się okazać bardziej wartościowe z perspektywy oceny logiki interwencji niż oszacowanie efektów netto określonych czynników.

## O autorach

### **Barbara Worek**

---

Adiunkt w Zakładzie Socjologii Gospodarki, Edukacji i Metod Badań Społecznych Instytutu Socjologii Uniwersytetu Jagiellońskiego w Krakowie oraz ekspert Centrum Ewaluacji i Analiz Polityk Publicznych UJ. Posiada wieloletnie doświadczenie zarówno w prowadzeniu badań ewaluacyjnych jak i szkoleń dotyczących ewaluacji. Od kilku lat prowadzi na Uniwersytecie Jagiellońskim zajęcia poświęcone ewaluacji i seminarium dotyczące zastosowania metod jakościowych w ewaluacji i analizie polityk publicznych. Realizując badania ewaluacyjne współpracowała z wieloma instytucjami administracji publicznej (centralnej i regionalnej). Doświadczenia badawcze zdobywała uczestnicząc w polskich i międzynarodowych projektach, dotyczących wykluczenia społecznego, rynku pracy i uczenia się przez całe życie. Autorka i współautorka publikacji dotyczących rynku pracy, kapitału ludzkiego i ewaluacji, oraz licznych raportów badawczych. Jej zainteresowania naukowe koncentrują się wokół metodologii badań jakościowych, zagadnień rozwoju kapitału ludzkiego i analizy polityk publicznych. Od 2009 r. zaangażowana w badania realizowane w ramach projektu PARP „Bilans Kapitału Ludzkiego”, gdzie odpowiada za problematykę związaną z uczeniem się osób dorosłych, działalnością sektora szkoleniowo-rozwojowego i inwestycjami pracodawców w rozwój kadr.

### **Marcin Kocór**

---

Adiunkt w Zakładzie Socjologii Gospodarki, Edukacji i Metod Badań Społecznych Instytutu Socjologii Uniwersytetu Jagiellońskiego w Krakowie oraz ekspert Centrum Ewaluacji i Analiz Publicznych UJ. Z badaniami marketingowymi i społecznymi jest związany od 18 lat. W tym okresie zdobył rozległe doświadczenie badawcze, prowadząc projekty m.in. dla Polskiej Agencji Rozwoju Przedsiębiorczości (ekspert ds. badania pracodawców w projekcie PARP „Bilans Kapitału Ludzkiego”), Wojewódzkiego Urzędu Pracy w Krakowie (pilotaż Zintegrowanego Systemu Kwalifikacji). Swoim doświadczeniem dzieli się prowadząc od 16 lat zajęcia dydaktyczne, obejmujące metody badań społecznych i marketingowych, projektowanie badań oraz techniki analizy danych. Zainteresowania naukowe skupia na metodologii badawczej nauk społecznych, technikach statystycznej analizy danych oraz zagadnieniach rynku pracy i kapitału ludzkiego.



## Bibliografia:

1. Babbie, E., *Badania społeczne w praktyce*, PWN, Warszawa 2004.
2. Basurto, X., Speer, J., „Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA)”, *Field Methods* 2012, 24.
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.
4. Cronqvist L., *Tosmana: Tool for small-n analysis, Version 1.522* [program komputerowy], Trier, Germany: University of Trier 2016, <https://www.tosmana.net/> (dostęp 25.05.2017 r.).
5. Donaldson S.I., Christie C.A., Mark M.M. (red.), *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](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf) (dostęp 24.02.2017 r.).
7. Gates E., Dyson L., „Implications of the Changing Conversation About Causality for Evaluators”, *American Journal of Evaluation*, 2016, volume: 38, issue: 1.
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”, w: J. Lichtarski, S. Nowisielski, G. Osbert-Pociecha, E. Tabaszewska-Zajbert (red.), „Nowe kierunki w zarządzaniu przedsiębiorstwem – wiodące tendencje”, *Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 2014, nr 340.
10. Leeuw F., „Reconstructing Program Theories: Methods Available and Problems to be Solved”, *American Journal of Evaluation* 2003, 24 (1).
11. Mayne J., „Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly”, *Canadian Journal of Program Evaluation* 2001, 16 (1).
12. McCluskey E.J., „Minimization of Boolean functions”, *Bell Systems Technical Journal* 1956, 35.
13. Quine W.V., „A way to simplify truth functions”, *American Mathematical Monthly* 1955, 62.
14. Pawson R., Tilley N., *Realist evaluation*, 2004, [http://www.communitymatters.com.au/RE\\_chapter.pdf](http://www.communitymatters.com.au/RE_chapter.pdf) (dostęp 25.02.2017 r.).
15. Pawson R., Tilley N., *Realistic evaluation*, Sage, London 1997.
16. Ragin Ch. C., Becker H. S. (red.), *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” w: C.C. Ragin (red.), *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* [program komputerowy], Department of Sociology, University of Arizona, Tucson 2009, <http://www.u.arizona.edu/~cragin/fsQCA/software.shtml> (dostęp 25.05.2017 r.).
21. Rihoux B., Ragin Ch. C. (red.), *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, Sage, Thousand Oaks, CA 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.
24. Stern E., Andersen O.W., Hansen H., „Editorial: Special issue: What can case studies do?”, *Evaluation*, 2013, 19 (3).
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, <http://www.oecd.org/derec/50399683.pdf> (dostęp 25.02.2017 r.).
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”, w: Berger E. S. C., Kuckertz A. (red.), *Complexity in Entrepreneurship, Innovation and Technology Research*, Springer International Publishing 2016.
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.

## **Badanie teorii interwencji przy użyciu analizy sieci bayesowskiej: wyniki pilotażowego działania dotyczącego wsparcia MŚP<sup>91</sup>**

Niniejszy rozdział jest poświęcony działaniu pilotażowemu, które zostało przeprowadzone w kontekście oceny *ex post* programów Europejskiego Funduszu Rozwoju Regionalnego (EFRR) w latach 2007-2013. Skupia się on na ocenie jednego z instrumentów Regionalnego Programu Operacyjnego 2007-2013 w Apulii. Instrument ten był skierowany do dużej liczby mikro- i małych przedsiębiorstw, realizował szeroko zakrojone cele i finansował wiele różnych rodzajów działań. W celu zbadania trudnych do zidentyfikowania zmian zachodzących w wyniku wdrażania instrumentu, zastało wykorzystane podejście ewaluacji wpływu opartej na teorii (*Theory Based Impact Evaluation, TBIE*) i zastosowano analizę sieci bayesowskiej (*Bayesian Network Analysis, BNA*). Uważamy, że ze względu na szeroki zakres celów, instrument ten nie wydaje się być szczególnie wydajny pod względem inicjowania długofalowych zmian strukturalnych, chociaż ma kluczowe znaczenie we wspieraniu przetrwania wielu mikro- i małych przedsiębiorstw w czasie kryzysu gospodarczego. Dzięki mechanizmom wykrytym za pomocą analizy sieci bayesowskiej możliwe było zidentyfikowanie poszczególnych wzorców zmian na poziomie przedsiębiorstwa. W naszej opinii możliwa jest poprawa efektywności instrumentu poprzez nałożenie większych ograniczeń w zakresie wyboru sektorów i dokonywanych inwestycji, oraz poprzez ograniczenie docelowych przedsiębiorstw do tych o najwyższym potencjale wzrostu.

<sup>91</sup> Niniejszy rozdział w dużej mierze czerpie z opracowania *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*. (Numer umowy: 2014CE16BAT002) realizowanego przez CSIL, na zlecenie Jednostki Ewaluacyjnej Dyrekcji Generalnej ds. Polityki Regionalnej i Miejskiej, Komisji Europejskiej, w drodze przetargu nieograniczonego. Jest to jedno z szesnastu badań w ramach całej oceny *ex post* EFRR i Funduszu Spójności, badających kluczowe wyniki polityki spójności w latach 2007-2013 oraz jedno z szesnastu badań poświęconych wsparciu przedsiębiorstw (pozostałe dwa dotyczą instrumentów finansowych - pakiet roboczy WP3 oraz dużych przedsiębiorstw - WP4).

Autorki dziękują personelowi KE i innym członkom Grupy Sterującej za nader przydatne informacje. Autorki ponoszą odpowiedzialność za wszelkie pozostałe błędy i pominięcia. Raporty okresowe i końcowe z badania są dostępne pod poniższym linkiem: [http://ec.europa.eu/regional\\_policy/en/policy/evaluations/ec/2007-2013/#3](http://ec.europa.eu/regional_policy/en/policy/evaluations/ec/2007-2013/#3).

## Wprowadzenie

Omawiamy zastosowanie podejścia ewaluacji wpływu opartej na teorii w celu dokonania oceny ex post skuteczności wsparcia skierowanego do MŚP w ramach programów polityki spójności na lata 2007-2013. Badamy w szczególności wykorzystanie innowacyjnej metodologii, analizy sieci bayesowskiej, w celu przeprowadzenia oceny zasadności logiki interwencji i ujawnienia mechanizmów zmiany oraz rezultatów generowanych przez badany instrument polityki.

Znaczna część literatury przedmiotu, badająca efekty interwencji publicznych wspierających przedsiębiorstwa opiera się na analizie ekonometrycznej lub metodach kontrfaktycznych<sup>92</sup>. Logika leżąca u podstaw tych metod polega na ustalaniu wpływu netto, jaki można przypisać działaniu w ramach wdrażanej polityki. Jednakże metody te nie pozwalają zrozumieć dlaczego dany instrument sprawdził się lub nie i które mechanizmy przyczynowe mogą wyjaśnić zaobserwowany wpływ. Mechanizmy te są zazwyczaj powiązane z behawioralnymi reakcjami, stymulowanymi przez politykę publiczną oraz z charakterystykami zewnętrznego kontekstu.

Dla decydentów, którzy chcieliby korzystać z wcześniejszych doświadczeń, aby móc projektować lepsze instrumenty, zdobywanie wiedzy o mechanizmach, wyjaśniających dlaczego dany instrument osiągnął swoje cele, jest równie istotne, co znalezienie ilościowych dowodów jego skuteczności. W tym celu konieczne jest przyjrzenie się zawartości „czarnej skrzynki” MŚP, zbadanie, w jaki sposób dany instrument przyczynia się do niwelowania problemów ograniczających zdolność MŚP do innowacji i wzrostu oraz do stymulowania zmiany zachowań<sup>93</sup>.

Ewaluacja wpływu oparta na teorii została opracowana w odpowiedzi na niektóre ograniczenia metod kontrfaktycznych. TBIE jest powszechnie uznaną metodą<sup>94</sup>, która odzwierciedla logikę (lub teorię) leżącą u podstaw ocenianej interwencji, a także bada założenia i zależności przyczynowo-skutkowe, które decydują o powstawaniu określonych efektów, zarówno pożądanym, jak i niepożądanym, oczekiwanych i nieoczekiwanych. Oferuje ona wartościową możliwość zbadania dlaczego i w jaki sposób określona interwencja wytworzyła lub nie dany efekt. Metoda polega, po pierwsze, na rekonstruowaniu logiki interwencji instrumentu wsparcia opracowanego przez decydentów, a następnie na te-

---

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

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

<sup>94</sup> Astbury B., Leeuw F.L., "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.

stowaniu teorii w celu odrzucenia lub potwierdzenia oczekiwań co do mechanizmów zmiany i rezultatów, jakie zostaną wygenerowane przez ten instrument.

Przedstawiamy zastosowanie podejścia TBIE w ocenie *ex post* Regionalnego Programu Operacyjnego na lata 2007-2013 w Apulii. Instrument, zwany „TITLE II”, który dostarczył wsparcia na inwestycje dla mikro- i małych przedsiębiorstw, polegał na połączeniu dotacji na spłatę odsetek oraz dotacji na pokrycie zaciągniętego zadłużenia bankowego na rozpoczęcie różnego rodzaju projektów inwestycyjnych. Jest to jedno z trzech objętych ewaluacją działań pilotażowych w zakresie indywidualnych instrumentów, przy czym pozostałe dwa, to „Kredyt Technologiczny” wdrażany w Polsce w ramach Programu Operacyjnego Innowacyjna Gospodarka 2007-2013 oraz „Wsparcie Badań, Rozwoju oraz Innowacji w Przemysle” w ramach Regionalnego Programu Operacyjnego 2007-2013 dla regionu Kastylia i León w Hiszpanii<sup>95</sup>. O ile polskie i hiszpańskie instrumenty były bardziej ambitne pod względem realizowanych celów i, do pewnego stopnia, bardziej innowacyjne w sposobie, w jaki zostały zaprojektowane i wdrażane, instrument apulijski stanowi bardziej tradycyjny system wsparcia i jest reprezentatywny dla szeregu instrumentów polityki, które były finansowane w okresie programowania 2007-2013, mających na celu wspieranie MŚP w przewyżczeniu negatywnych skutków kryzysu gospodarczego.

Nowatorstwo naszego podejścia w ewaluacji polega na wykorzystaniu analizy statystycznej i sieci bayesowskich do badania teorii interwencji, ujawnienia mechanizmów zmiany jak również rezultatów generowanych przez omawiany instrument wsparcia. Analizę sieci bayesowskiej uznano za szczególnie przydatną przy przeprowadzaniu analizy złożoności przedmiotu badania. Wykorzystano ją w połączeniu z modelami regresji do kontrolowania poziomu statystycznej istotności korelacji wśród zmiennych. Miała ona kluczowe znaczenie dla prawidłowego przetestowania teorii i znalezienia ukrytych lub nieoczekiwanych mechanizmów zmiany. W połączeniu z innymi metodami, BNA może zagwarantować uzyskanie wiarygodnych rezultatów i doprowadzić do wyklarowania poglądu na temat tego czy i w jaki sposób dany instrument jest efektywny<sup>96</sup>.

Rozdział został podzielony na pięć części. W części 2. omawiamy działanie TITLE II. W części 3. przedstawiamy metodologię sieci bayesowskiej i jej przydatność w prze-

---

<sup>95</sup> Ocena wybranych instrumentów została zamówiona w ramach ewaluacji, która obejmowała wiele innych działań (w szczególności: przegląd literatury przedmiotu, przegląd dokumentacji 50 wybranych programów operacyjnych i ich rocznych sprawozdań z wdrażania, 8 gruntownych studiów przypadku z szeroko zakrojonymi pracami w terenie dotyczącymi wybranych programów operacyjnych oraz seminarium z podmiotami zainteresowanymi, ekspertami i przedstawicielami instytucji zarządzających). Wszystkie trzy instrumenty wybrano na podstawie sprawozdań ze studiów przypadku pod względem reprezentatywności (realizujących różne i typowe cele polityki w dziedzinie wsparcia MŚP) oraz wykonalności (liczba beneficjentów i dostępność danych, w szczególności w przypadku danych kontaktowych użytkowników końcowych).

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

prowadzaniu wybranego studium przypadku. W części 4. omawiamy rezultaty ewaluacji instrumentu, z podziałem na rezultaty w zakresie kondycji ekonomicznej, zmiany zachowania w przedsiębiorstwach oraz mechanizmów leżących u podstaw zmian zidentyfikowanych za pomocą analizy sieci bayesowskiej. Część 5. zawiera wnioski.

## Kontekst

Region Apulii w południowych Włoszech, klasyfikowany jako region konwergencji, charakteryzuje się przewagą najmniejszych firm działających w tradycyjnych sektorach, zatrudniających od 0 do 2 osób. Kryzys finansowy i gospodarczy spowodował znaczny spadek poziomu wzrostu gospodarczego i zatrudnienia. Z końcem 2013 r. liczba osób bezrobotnych w regionie była o ponad 70% wyższa niż w 2007 r.<sup>97</sup>

W wyniku globalnej recesji i w celu ograniczenia krajowego długu publicznego, wkład środków krajowych w ramach polityki spójności, a także krajowe instrumenty wsparcia przedsiębiorstw, zostały drastycznie zredukowane, szczególnie w południowych regionach kraju. Dla zniwelowania tej tendencji, wykorzystano regionalne fundusze. Celem Regionalnego Programu Operacyjnego 2007-2013 było, z jednej strony, promowanie badań i rozwoju oraz innowacji za pomocą selektywnych instrumentów, ukierunkowanym na zwiększanie potencjału innowacyjnego oraz wspieranie bardziej ambitnych planów inwestycyjnych i zmian strukturalnych, z drugiej zaś strony, wspieranie konkurencyjności systemu produkcji i zatrudnienia, w oparciu o zazwyczaj bardziej ogólne i małe projekty inwestycyjne.

W tej ostatniej grupie mieści się instrument będący przedmiotem naszego opracowania, czyli TITLE II, który oferował pomoc na inwestycje dla mikro- i małych przedsiębiorstw (zatrudniających poniżej 50 pracowników, zgodnie z definicją obowiązującą w KE). Instrument polegał na połączeniu dotacji na spłatę odsetek z dotacją dla mikro- i małych przedsiębiorstw, które miały zaciągnięte kredyty bankowe, z przeznaczeniem na rozpoczęcie różnego rodzaju projektów inwestycyjnych<sup>98</sup>. TITLE II otrzymał największą część zadeklarowanych funduszy publicznych z zestawu instrumentów adresowanych do MŚP w ramach PO (niemal 120 milionów EUR). Instrument zapewniający kompleksowe podejście, został uruchomiony w kwietniu 2009 r. i trwał do czerwca 2014 r., a kwalifikowane sektory z biegiem lat rozszerzono<sup>99</sup>. W ramach

---

<sup>97</sup> Komisja Europejska „Support to SMEs – Increasing Research and Innovation in SMEs and SME Development.” Trzecie sprawozdanie okresowe - Pakiet roboczy 2. Przedłożone przez CSIL w partnerstwie z CSES i ZEW. Bruksela 2015.

<sup>98</sup> Tamże.

<sup>99</sup> Z początku jako kwalifikujące się wskazano następujące sektory: przedsiębiorstwa rzemieślnicze, przedsiębiorstwa handlowe, w tym detalistów i przedsiębiorstwa usług spożywczych (np. bary i restauracje); inne firmy w sektorach wytwórstwa, budownictwa, informacji i łączności. W kolejnych miesiącach tego samego roku dodano nowe kategorie działalności w ramach sektora handlowego i pewne rodzaje działalności w dziedzinie zdrowia i pracy socjalnej (np. usługi dla przedszkoli). Zakres kwalifikowalnych przedsiębiorstw stopniowo rozszerzał się w kolejnych latach, ostatecznie obejmując również rodzaje działalności w sektorze zielonym (np. recykling plastiku) (Komisja Europejska 2015).

instrumentu mogło być finansowane sześć rodzajów wydatków: zakup gruntów, koszt budowy i prac remontowych, zakup infrastruktury, zakup maszyn, wyposażenia i pojazdów, zakup programów komputerowych, zakup patentów i licencji. W ogólnym ujęciu, TITLE II miał promować działania związane z modernizacją działalności gospodarczej, w tym zarówno w oparciu o ogólne rodzaje wydatków (np. prace remontowe lub zakup nowego sprzętu, wyposażenia przestrzeni handlowych lub administracyjnych) jak i inne wydatki w większym stopniu zorientowane na innowacje (np. zakup patentów w celu wprowadzenia innowacyjnych procesów produkcji). Kryteria definiujące progi pomocy były wielokrotnie poddawane przeglądowi w trakcie okresu programowania, co miało na celu zwiększenie atrakcyjności instrumentu dla przedsiębiorstw. Pomoc polegała na przyznaniu dotacji przeznaczonej na spłatę części odsetek od zadłużenia zaciągniętego na sfinansowanie inwestycji. Kwota pomocy była wyliczana jako procent odsetek należnych od zaciągniętego kredytu. Ponadto mikroprzedsiębiorstwa mogły skorzystać z dotacji na częściowe pokrycie kosztów zakupu maszyn i wyposażenia. Siedem miesięcy po uruchomieniu instrumentu dotacja ta została rozszerzona również na małe przedsiębiorstwa. Maksymalna intensywność pomocy (udział całkowitego wkładu publicznego w wielkości inwestycji) wzrosła z biegiem lat, z 40% i 30% w kwietniu 2009 r. odpowiednio dla mikro- i małych przedsiębiorstw, do 45% dla obydwu rodzajów firm na początku 2011 r.

Lokalna agencja pełniąca rolę Instytucji Pośredniczącej odpowiadała za zarządzanie dystrybucją środków publicznych dla przedsiębiorstw. Odpowiadała ona za ocenę otrzymanych wniosków, głównie pod względem spójności finansowej. Przedsiębiorstwa składające wnioski musiały przedstawić prognozowane wskaźniki dotyczące zatrudnienia (w podziale na kobiety i mężczyzn), które zostaną osiągnięte w wyniku inwestycji. Jednakże wskaźniki te nie były wiążące i nie były wykorzystywane do rankingowania projektów. Projekty inwestycyjne były wybierane lub odrzucane indywidualnie, bez porównywania między sobą.

## **Analiza sieci bayesowskiej**

W tej części omawiamy narzędzie analizy sieci bayesowskiej i objaśniamy powody jego przydatności w kontekście przeprowadzonej ewaluacji. BNA jest podejściem łączącym w sobie analizę mapy graficznej z analizą statystyczną, w celu pokazania powiązań pomiędzy zmiennymi. Sieć bayesowska ilustruje prawdopodobieństwo relacji pomiędzy zbiorem zmiennych a ich niezależnością warunkową. Zapewnia również kompaktowy obraz wspólnego rozkładu prawdopodobieństwa<sup>100</sup>. Sieci bayesowskie są definiowane przez strukturę sieciową, zwaną skierowanym grafem acyklicznym (*the directed acyclic graph, DAG*) oraz zbiór warunkowych rozkładów prawdopodobieństwa, powiązanych ze zmiennymi wchodzącymi do DAG. Skierowany graf acykliczny

---

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

jest zbiorem zmiennych losowych przedstawianych jako węzły. Rolą sieci jest wyrażenie relacji warunkowej niezależności wśród zmiennych w modelu, poprzez graficzne rozdzielanie. Hierarchiczne pozycjonowanie zmiennych (które zakłada, że A jest powiązane z B, które jest powiązane z C, które jest powiązane z D) można interpretować jako domniemanie przyczynowości pomiędzy tymi zmiennymi. Jednak relacje przyczynowe muszą zostać potwierdzone przez analityka w oparciu o wcześniej nabytą wiedzę o zmiennych. W przypadkach, w których relacja przyczynowa nie jest znana, węzły sieci są połączone ze sobą bez określonego kierunku.

Przedstawiamy kilka zalet zastosowania analizy sieci bayesowskiej w kontekście niniejszego badania<sup>101</sup>. Przede wszystkim BNA jest pomocna wówczas, gdy celem badania jest zidentyfikowanie wielu, wzajemnie ze sobą połączonych czynników determinujących, na przykład szczególnego zachowania, oraz zrozumienie w jaki sposób są ze sobą połączone. Ze względu na złożony charakter MŚP, trudno jest poznać *a priori* mechanizmy przyczynowe, które łączą wszystkie możliwe zmienne objaśniające i zależne. BNA pomaga rozwikłać złożone relacje i wykryć leżący u ich podstaw, przynajmniej częściowo, nieznan system przyczynowy. W oparciu o dostępne dane możliwe jest sprawdzenie istnienia czynników zakłócających, czyli zmiennych, które są zarazem zmiennymi zależnymi i niezależnymi w modelu statystycznym, a także oszacowanie warunkowych prawdopodobieństw dla wszystkich zmiennych w modelu. W tym kontekście należy podkreślić, że sieci bayesowskie niekoniecznie pokazują te same korelacje, które powstają w wyniku zastosowania modeli regresji. Analiza regresji identyfikuje zmienne, które w bardziej znaczący sposób wpływają na uprzednio zdefiniowaną zmienną zależną. Natomiast BNA pokazuje, które zmienne są względem siebie niezależne w oparciu o ich warunkowe prawdopodobieństwo wystąpienia. Na przykład, w przypadku, gdy analiza regresji pokazuje, że wyniki ekonomiczne MŚP zależą od wysokości otrzymanego wsparcia ze środków publicznych, BNA może pokazać, że wyniki ekonomiczne zależą od rodzaju zmiany, która nastąpiła w funkcji produkcyjnej przedsiębiorstwa (na przykład nabycie nowych technologii produkcyjnych), co z kolei będzie powiązane ze wsparciem ze środków publicznych. W związku z powyższym wyniki ekonomiczne i wsparcie publiczne są niezależne względem siebie w sieci, gdy jest ona kontrolowana w oparciu o inne zmienne.

Po drugie BNA może być stosowana do wsparcia wnioskowania, zarówno predykcyjnego jak i diagnostycznego<sup>102</sup>. Wsparcie predykcyjne („od ogółu do szczegółu”) opiera się na wydobywaniu danych ze zmiennej nadrzędnej i wykorzystaniu jej do przewidywania wzoru zmiennych podrzędnych, a mianowicie jej warunkowego rozkładu praw-

---

<sup>101</sup> Daly R., Shen Q., Aitken S., „Learning Bayesian networks: approaches and issues”, *The Knowledge Engineering Review* 2011, vol. 26, no. 2, pp. 99-157; Komisja Europejska „Support to SMEs – Increasing Research and Innovation in SMEs and SME Development.” Trzecie sprawozdanie okresowe - Pakiet roboczy 2. Przedłożone przez CSIL w partnerstwie z CSES i ZEW. Bruksela 2015.

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



dopodobieństwa. Wsparcie diagnostyczne („od szczegółu do ogółu”) działa w kierunku odwrotnym: zmienne podrzędne w BNA można wykorzystać do analizy rozkładu zmiennej nadrzędnej. Na przykład badając potencjalną relację pomiędzy nabywaniem nowych technologii produkcyjnych (zmienna nadrzędna) a wynikami ekonomicznymi (zmienne podrzędne) można wykorzystać wsparcie predykcyjne (tj. na podstawie danych dotyczących zmiany technologii produkcyjnej pozyskujemy wiedzę na temat wpływu tej zmiany na wyniki ekonomiczne) lub diagnostyczne (tj. na podstawie obserwowanej zmiany wyników ekonomicznych pozyskujemy wiedzę o potencjalnej przyczynie, analizując rozkład zmiennej technologii produkcyjnej oraz wszelkiej innej zmiennej).

Trzecią korzyścią ze stosowania BNA jest ich zdolność funkcjonowania w roli potencjalnych prognostyków interwencji. W istocie, jeśli relacje pomiędzy zmiennymi nadrzędnymi a podrzędnymi w sieci są stabilnym i autonomicznym mechanizmem zmian, wiedza na temat danego instrumentu wsparcia w takich modularnych konfiguracjach, zwiększa możliwość racjonalnego przewidywania wpływu zewnętrznych interwencji. Dzięki tej modułowości możliwe jest przeprowadzanie symulacji i ocena scenariuszy, poprzez proste zmiany lub ustalenie wartości poszczególnych zmiennych oraz sprawdzenie w jaki sposób zmienia się odpowiednia zmienna nadrzędna lub podrzędna. Na przykład jeśli wiemy, że poziom eksportu jest powiązany z rozkładem wielkości przedsiębiorstwa, można przeprowadzić symulację pokazującą, co stałoby się z rozkładem zmiennej wynikowej, gdyby rozkład zmiennej wielkości przedsiębiorstwa zmienił się, na przykład poprzez zwiększenie liczby średnich przedsiębiorstw i zredukowanie liczby małych przedsiębiorstw.

Biorąc pod uwagę wszystkie powyższe argumenty BNA jest innowacyjnym narzędziem wspierającym strategiczne decyzje<sup>103</sup> oraz prowadzenie ewaluacji *ex post*.

## Zastosowanie i wyniki

Logika interwencji programu TITLE II została zbadana empirycznie. Celem było sprawdzenie, czy spodziewane rezultaty zostały osiągnięte zgodnie z oczekiwaniami decydenta, a także jakie mechanizmy lub czynniki przyczyniły się do zaobserwowanych zmian. Przy przeprowadzaniu analizy kierowano się następującymi pytaniami badawczymi<sup>104</sup>:

- Czy instrument przyczynił się do utrzymania zatrudnienia w mikro- i małych przedsiębiorstwach, które otrzymały wsparcie? Jakie zmiany w podstawowej działalności firm lub inne czynniki mogą uzasadniać zaobserwowane osiągnięcia?

---

<sup>103</sup> 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.

<sup>104</sup> Zgodnie ze specyfikacją przetargową.

- Czy instrument był skuteczny w stymulowaniu wzrostu skłonności przedsiębiorstw do inwestowania? Jakie zmiany w podstawowej działalności firm lub inne czynniki mogą uzasadniać zaobserwowane osiągnięcia?
- Czy zaobserwowane zmiany są zgodne z teorią interwencji? Jeśli zarejestrowano odchylenia od teorii, dlaczego wystąpiły?
- Czy instrument wygenerował inne rodzaje zmian zachowań w mikro- i małych przedsiębiorstwach objętych wsparciem? Jakie czynniki mogą tłumaczyć zaobserwowane zmiany?

Aby odpowiedzieć na te pytania przesłano ankietę do 2 441 małych i mikroprzedsiębiorstw, które zakończyły inwestycje i otrzymały już środki publiczne. Ogółem zostało wypełnionych 399 kwestionariuszy, co stanowi 25% przedsiębiorstw, z którymi byliśmy faktycznie w kontakcie. Przedstawiając rezultaty tego badania wyodrębniliśmy sferę wyników ekonomicznych, sferę zmian zachowań w przedsiębiorstwach oraz mechanizmy leżące u podstaw rezultatów, które zostały wykryte za pomocą BNA.

## **Wyniki ekonomiczne**

Przeprowadzona analiza wykazała, że instrument był ogólnie skuteczny pod względem generowania pozytywnych rezultatów ekonomicznych, w szczególności zwiększenie sprzedaży, zdolności przeciwdziałania skutkom kryzysu oraz obniżenie ryzyka bezrobocia. W swoich odpowiedziach na temat osiągniętych wyników ekonomicznych, uzyskanych dzięki realizacji i wsparciu projektu inwestycyjnego, 80% przedsiębiorstw twierdzi, że projekt faktycznie pozwolił im lepiej przeciwdziałać skutkom kryzysu gospodarczego i ograniczyć jego wpływ. W odniesieniu do poziomu zatrudnienia, 12% przedsiębiorstw twierdzi, że wśród najistotniejszych zaobserwowanych zmian, jakie zaszły w ich produkcji po zrealizowaniu inwestycji, należy wymienić utrzymanie istniejącego poziomu zatrudnienia, podczas gdy 22% przedsiębiorstw zadeklarowało wzrost wskaźnika zatrudnienia. Ponadto można zaobserwować, że wysoki odsetek przedsiębiorstw (85%) odnotował wzrost sprzedaży, pomimo, że wzrost ten był zazwyczaj umiarkowany.

W celu zrozumienia mechanizmów uzasadniających takie wyniki ekonomiczne, należy uwzględnić rodzaj realizowanych inwestycji, ponieważ TITLE II wspierał bardzo zróżnicowane spektrum inwestycji. Pozyskane dane wskazują, że większość objętych wsparciem projektów miała na celu ogólne rozszerzenie działalności gospodarczej, nie zaś inwestowanie w innowację produktową lub procesową. W badaniu podjęto próbę wykroczenia poza tę szeroką klasyfikację na rzecz lepszego sprecyzowania rodzaju i celu poniesionych wydatków. Badane przedsiębiorstwa wykorzystały pozyskane z działania TITLE II wsparcie głównie na zakup nowych aktywów do celów produkcyjnych lub działalności handlowej. W mniejszym stopniu wsparcie zostało wykorzystane do sfinansowania prac restrukturyzacyjnych lub budowy nowych budynków i biur. 29% przedsiębiorstw nabyło dobra o wyższej wartości dodanej, takie

jak systemy informatyczne, sprzęt lub roboty do automatyzacji, patenty i licencje. 204 przedsiębiorstwa dzięki inwestycji ulepszyły istniejące produkty lub oferowane usługi.

Instrument wsparcia pozostawiał beneficjentowi stosunkowo dużą dowolność wyboru ścieżki planów inwestycyjnych. Dlatego, co nie jest zaskakujące, analiza statystyczna wykazała, że rodzaj i poziom osiągniętych wyników ekonomicznych w dużej mierze zależał od rodzaju zmian, jakie nastąpiły w MŚP w ramach realizowanych projektów inwestycyjnych. Ścisłej, analiza danych zidentyfikowała następujące korelacje:

- po pierwsze, wzrost odporności przedsiębiorstwa na kryzys jest pozytywnie i znacząco skorelowany z inwestycjami, które zwiększają rozpoznawalność/znajomość marki przedsiębiorstwa oraz z liczbą pracowników i obniżonym zużyciem energii;
- po drugie, wzrost sprzedaży jest pozytywnie skorelowany z rozszerzeniem gamy produktów i oferowanych usług, a także ze wzrostem wydajności i zatrudniania. Jednakże kontrola innych zmiennych, na przykład sektora, wielkości i kategorii (rzemiosło, handel lub inne) przedsiębiorstwa, wykazała negatywny związek z powiększaniem lub udoskonaleniem przestrzeni roboczej zakładu pracy;
- po trzecie, przedsiębiorcy, którzy rozpoczęli nową działalność, ulepszyli istniejące produkty lub usługi, rozszerzyli gamę produktów, zaoferowali nowe produkty, rozwinęli nowe kanały sprzedaży lub doprowadzili do większej rozpoznawalności przedsiębiorstwa, uzyskali wyższy od pozostałych wzrost dochodów;
- po czwarte, bardziej gwałtowne spadki kosztów całkowitych osiągają przedsiębiorstwa, które obniżyły swoje zużycie energii, ulepszyły organizację pracy lub w inny sposób poprawiły efektywność;
- po piąte, przedsiębiorstwa, które zwiększyły lub utrzymały poziom zatrudnienia, rozszerzyły gamę oferowanych produktów i zwiększyły swoją rozpoznawalność. Rezultat ten jest również pozytywnie i znacząco skorelowany z jednoczesnym wdrażaniem innych projektów inwestycyjnych.

Analiza sieci bayesowskiej potwierdza te ustalenia i dodaje kilka nowych, dotyczących mechanizmów, za pomocą których generowane są wyniki ekonomiczne. W szczególności pokazuje, że poprawa postrzegania przez przedsiębiorstwo jego zdolności do opierania się kryzysowi gospodarczemu zależy od zyskania większej liczby klientów, co z kolei jest pozytywnie powiązane ze wzrostem sprzedaży oraz realizacją innych projektów inwestycyjnych w tym samym czasie. Ponadto, okazuje się, że wzrost sprzedaży bezpośrednio przekłada się na wzrost dochodów przedsiębiorców i jest ściśle powiązany ze wzrostem liczby i rodzaju klientów. Wreszcie zatrudnienie nowych pracowników, a zatem zwiększenie wielkości przedsiębiorstwa, jest najistotniejszym czynnikiem decydującym o wzroście sprzedaży. Natomiast wzrost liczby klientów jest powiązany ze wzrostem rozpoznawalności firmy. Stąd BNA umożliwia nie tylko empiryczne zbadanie zasadności teorii zmiany leżącej u podstaw opracowanych interwencji, lecz także objaśnienia i precyzyjniej określa niektóre z zaproponowanych mechanizmów.

Oprócz rodzajów zmian powstałych w wyniku realizacji projektu inwestycyjnego, kluczowymi zmiennymi objaśniającymi wyniki ekonomiczne są cechy przedsiębiorstwa. Ścisłej rzecz biorąc, podczas gdy średnie wyniki ekonomiczne wydają się generalnie pozytywne, po uwzględnieniu pełnej próby respondentów, zindywidualizowana analiza, rozróżniająca poszczególne typy przedsiębiorstw, umożliwia poznanie nowych i interesujących wniosków. Po pierwsze, przedsiębiorstwa produkcyjne wdrażały średnio bardziej kosztowne inwestycje, które umożliwiły im osiągnięcie istotnych zmian w procesie produkcji. Przedsiębiorstwa te, chociaż nie stanowią większości beneficjentów, osiągnęły istotne korzyści, które doprowadziły do wyższej odporności na kryzys i wyższego wzrostu zatrudnienia. W konsekwencji wydaje się, że ta kategoria przedsiębiorstw przechodzi konkretną ścieżkę rozwoju, na co również wskazuje ich wysoka skłonność do dalszych inwestycji. Po drugie, przedsiębiorstwa handlowe częściej wykorzystywały środki pozyskane w ramach TITL II do nabywania mebli i nowych aktywów dla działalności handlowej i duża część beneficjentów sądzi, że zmiany te przyniosły wyraźny wzrost obrotów. Niemniej, jak wspomniano powyżej, ulepszenie przestrzeni roboczej nie jest równie efektywne, co rozszerzanie gamy produktów i zwiększanie wydajności. Po trzecie, przedsiębiorstwa rzemieślnicze, które stanowią większość objętych wsparciem podmiotów (z których niemal wszystkie prowadzą jednoosobową działalność gospodarczą lub są mikroprzedsiębiorstwami), osiągnęły niejednorodne rezultaty, jednak ogółem niższe od tych zanotowanych przez inne kategorie podmiotów. Rzemieślnicy działający w sektorze budowlanym ucierpieli w kryzysie i wsparcie uzyskane w ramach TITL II zapewniło im ograniczoną pomoc. Przedsiębiorstwa działające w sektorze produkcyjnym uzyskały więcej korzyści, jednak nie tak wysokich, jak te uzyskane przez najmniejsze przedsiębiorstwa.

### **Zmiana behawioralna w odniesieniu do skłonności do inwestowania**

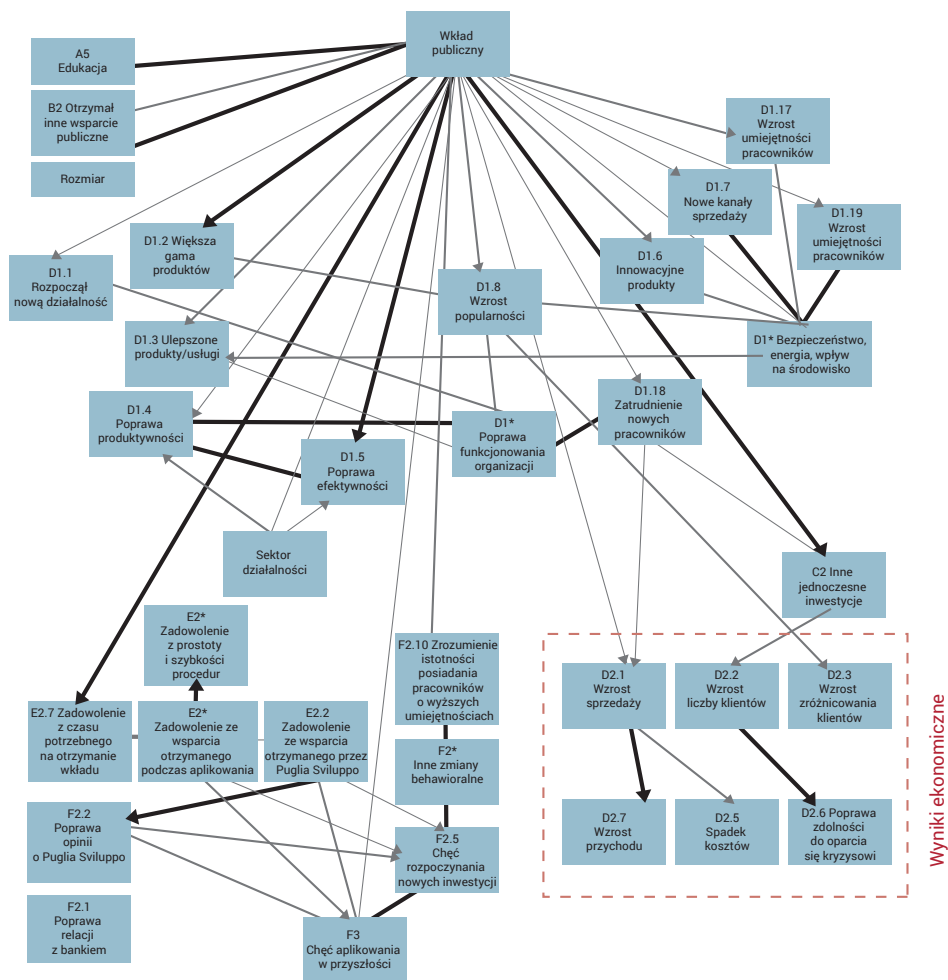
Drugim głównym celem instrumentu było stymulowanie zmian zachowań przedsiębiorstw, polegających na zwiększeniu skłonności do realizowania projektów inwestycyjnych. 73% wszystkich badanych przedsiębiorców twierdzi, że po skorzystaniu z TITL II zaczęli rozważać realizację nowych projektów inwestycyjnych, których nigdy wcześniej nie brali pod uwagę. Odpowiedzi na to pytanie są ściśle skorelowane z chęcią przedsiębiorstw do ubiegania się w przyszłości po raz kolejny o dotacje ze środków publicznych. Należy zaznaczyć, że 70 % przedsiębiorstw, które nigdy nie korzystały ze wsparcia publicznego przed otrzymaniem TITL II, deklaruje, że prawdopodobnie ponownie będą ubiegać się w przyszłości o dofinansowanie ich potencjalnych planów inwestycyjnych. O ile może to otworzyć drogę do oportunistycznych zachowań w poszukiwaniu wsparcia publicznego, może być też pozytywną szansą dla rzeczywistych inwestycji realizowanych przez MŚP, które borykają się z poważnymi ograniczeniami kredytowymi. Z kolei analiza statystyczna pokazuje, że zamiar ubiegania się w przyszłości o dofinansowanie jest silniejszy, jeśli przedsiębiorstwa pozytywnie oceniły wsparcie otrzymane przez konsultantów i lokalną Instytucję Pośredniczącą w trakcie całej procedury ubiegania się o środki TITL II. Pozytywne opinie dotyczą-

ce tych aspektów rekompensują zazwyczaj niższy poziom zadowolenia z obciążeń administracyjnych, trudności proceduralnych i czasu potrzebnego na przygotowanie wniosku oraz uzyskiwania publicznego wkładu po zakończeniu projektu. W sieci bayesowskiej możliwe jest prześledzenie, w przypadku których ścieżek skłonność przedsiębiorstw do myślenia o potencjalnych projektach inwestycyjnych jest silnie powiązana z ich opinią o całym procesie wdrażania instrumentu. Sieci te pokazują również, że aspekty zmiany zachowań nie są tak ściśle powiązane z wynikami ekonomicznymi osiąganymi dzięki inwestycji, jak można by oczekiwać.

Przy uwzględnianiu cech beneficjentów dostępne dane pokazują, że „czysty” efekt dodatkowości (inwestycja, która nie została by rozpoczęta bez przyznanej pomocy), który jest szacowany w oparciu o odpowiedzi przedsiębiorstw, wydaje się silniejszy w przypadku mikroprzedsiębiorstw rzemieślniczych w porównaniu z przedsiębiorstwami sektora handlowego. Zarazem mniej prawdopodobne jest, że przedsiębiorstwa rzemieślnicze i handlowe (w porównaniu do innych rodzajów przedsiębiorstw) będą rozważały nowe potencjalne projekty inwestycyjne, a w przyszłości miałyby ubiegać się ponownie o inne formy wsparcia publicznego. Chociaż prawdą jest, że wszystkie trzy rodzaje przedsiębiorstw deklarują, że planują rozpoczęcie nowych inwestycji, przedsiębiorstwa inne niż rzemieślnicze i handlowe są relatywnie bardziej nastawione na inwestycje. Ponadto, nawet w tych samych latach, w których realizowano projekt finansowany z TITTLE II, połowa z nich wdrażała inne projekty inwestycyjne, co stanowiło wyższy odsetek niż w przypadku firm z sektorów rzemiosła i handlu. Różnice są wyraźne, nawet w przypadku porównania mikro- i małych przedsiębiorstw, przy czym te ostatnie cechują się zazwyczaj wyższą skłonnością do inwestowania. Analiza wykazała również inne zmiany behawioralne wśród niektórych beneficjentów, na przykład dotyczące preferencji co do rodzajów potrzebnych pracowników (np. bardziej wykwalifikowanych lub młodszych), jednak dotyczą one zazwyczaj jedynie połowy przedsiębiorstw.

Rysunek 1 przedstawia sieć bayesowską związaną z TITTLE II. Pokazuje on powiązania łączące poszczególne rodzaje działalności realizowanej przez beneficjentów, zmiany w ich modelu produkcji powstałe w wyniku wsparcia uzyskanego z TITTLE II, cechy przedsiębiorstw (np. rozmiar, sektor, fakt wdrażania innych projektów inwestycji w tych samych latach, itd.), ich wyniki ekonomiczne (zmienne od D2.1 do D2.7) oraz ich zmianę w zachowaniu (od F2.1 i F2.10).

**Rysunek 1. Zmiany wywołane przez TITLE II zgodnie z analizą sieci bayesowskiej**



Plik źródłowy: DB\_JT  
Wiersze z danymi: 399

Algorytm uczyć się: Bayesian  
Czas wyszukiwania: 129.562 sekund  
Najlepszy wynik w 4 iteracji: -13978.6

Parametry algorytmu:

Iteracje: 20  
Wielkość próby: 50  
Prawdopodobieństwo połączenia: 0.001  
Maksymalny czas wyszukiwania: 0  
Używać dokładności jako punktacji: nie

A3  
Rok założenia

Region  
NUTS3

Uwaga: skierowane strzałki odpowiadają relacji przyczynowej; zwykłe połączenia między zmiennymi wskazują na korelację, bez żadnego pewnego kierunku przyczynowego. Im grubsza jest strzałka, tym silniejsza korelacja między zmiennymi, zgodnie z szacunkami GeNIe. Graf obejmuje pewne zmienne (u dołu po prawej), które pomimo że były kontrolowane podczas budowy modelu, nie są ostatecznie silnie powiązane z żadną konkretną zmienną. Gwiazdki wskazują główne zmienne składowe.

Źródło: Komisja Europejska (2015).

## **Mechanizmy i warunki stojące za generowaniem efektów wykrytych za pomocą BNA**

Strategia instrumentu publicznego nie opisuje wyraźnie oczekiwanych mechanizmów stojących za generowaniem pożądaných efektów, głównie z powodu dużej różnorodności typów wydatków kwalifikowalnych i rodzajów wspieranych inwestycji. Inwestycja mogła w rzeczywistości determinować inny rodzaj zmian w MŚP oraz efekty ekonomiczne, w zależności od tego czy polegała na przykład na renowacji lokalu handlowego i wyposażenia, innowacjach produktowych i procesowych, rozpoczęciu nowej działalności gospodarczej czy na czymś innym. Decydenci opracowują logikę instrumentu raczej w sposób ogólny, bez faktycznego opisywania sposobu, w jaki różne rezultaty powinny występować.

Analiza sieci bayesowskiej potwierdza istnienie zróżnicowanych efektów i pomaga określić w bardziej precyzyjny sposób mechanizmy i warunki generowania poszczególnych rezultatów, zarówno tych prognozowanych, jak i tych nieoczekiwanych. Przeprowadzona BNA pozwala nam przedstawić kilka wniosków.

Po pierwsze, pomimo różnorodności wydatków poczynionych przez przedsiębiorstwa kwalifikujące się do otrzymania wsparcia TITLE II, jedynie niektóre z nich przyniosły bezpośrednie i pozytywne skutki ekonomiczne. W szczególności przyjrzenie się różnym cechom przedsiębiorstw, pozwoliło nam zauważyć, że większy wpływ na obroty miały inwestycje, które umożliwiły przedsiębiorstwu poszerzenie zakresu oferowanych produktów, poprawę wydajności i zatrudnienie nowych pracowników. Inwestycje, które stymulują rozpoznawalność przedsiębiorstwa, podnoszą jego odporność na kryzys.

Po drugie, rodzaj i intensywność efektów różniły się istotnie w zależności od cech beneficjentów. Małe przedsiębiorstwa sektora produkcyjnego realizowały przeciętnie bardziej kosztowne inwestycje zorientowane na pozyskanie nowych aktywów produkcyjnych. Dzięki inwestycji osiągały wzrost produktywności i wyższą jakość oferowanych produktów, dzięki czemu mogły czerpać rozległe korzyści gospodarcze. Sektor handlu, w szczególności przedsiębiorstwa działające w branży usług spożywczych, zazwyczaj korzystały z TITLE II w celu dokonania renowacji umeblowania i przestrzeni w swoich lokalach użytkowych. Ten rodzaj wydatków, pomimo że wpływający na obroty w niewielkim stopniu, w porównaniu z innymi rodzajami inwestycji, umożliwiał beneficjentowi zwiększenie sprzedaży. W pewnym stopniu pozytywne wyniki tego sektora można również tłumaczyć pozytywnym trendem w sektorze turystyki w regionie. TITLE II nie mógł przynieść żadnych znaczących korzyści przedsiębiorstwom sektora budowlanego, który ucierpiał bardziej od innych z powodu złych trendów makroekonomicznych i bezrobocia.

Po trzecie, osiągnięcie przez przedsiębiorstwa pozytywnych efektów zależy również od jednoczesnego realizowania innych projektów inwestycyjnych. Częściej były one inicjowane przez małe przedsiębiorstwa produkcyjne (nie rzemieślnicze), które okazały

się bardziej dynamiczne od innych rodzajów przedsiębiorstw pod względem ubiegania się o środki publiczne i wykazywały większą chęć do rozpoczynania innych inwestycji w przyszłości. Niemniej ten typ przedsiębiorstw stanowi niewielką część beneficjentów instrumentu.

Jedynym powiązaniem przyczynowo-skutkowym, które BNA obala, jest spodziewany związek pomiędzy chęcią ubiegania się przez przedsiębiorstwa o dofinansowanie ze środków publicznych, a ich zadowoleniem z prostoty aplikowania i procesu wdrażania instrumentu oraz osiąganymi wynikami ekonomicznymi. Pomoc oferowana przez lokalną Instytucję Pośredniczącą i prywatnych konsultantów biznesowych ułatwiła składanie wniosku i wydatkowanie funduszy oraz zachęciło przedsiębiorstwa do ponownego aplikowania.

## **Wnioski**

Apulijski instrument wsparcia TITLE II skierowany do bardzo dużej liczby mikro- i małych przedsiębiorstw, obejmował szeroko zakrojone cele i finansował wiele różnego rodzaju działań. Pomimo że tradycyjne techniki ewaluacji wykryły ograniczoną poprawę wyników ekonomicznych, nasze badanie ewaluacyjne wykazało, że instrument odegrał ważną rolę we wspieraniu przedsiębiorstw w przezwyciężaniu poważnych skutków kryzysu gospodarczego. Dowody pokazują, że za pomocą TITLE II udało się osiągnąć oczekiwane rezultaty przynajmniej w ujęciu ogólnym, jednak biorąc pod uwagę rodzaje realizowanych projektów i korzystające ze wsparcia przedsiębiorstwa, istnieje wiele różnych interpretacji warunków wyjaśniających wybrane historie sukcesu.. Z uwagi na ograniczoną selektywność tego instrumentu pod względem wspieranych inwestycji oraz wywołanych zmian, ocena jego skuteczności jest jeszcze trudniejsza.

BNA została użyta do zbadania zmian, które zaszły w wyniku wdrożenia instrumentu publicznego, których identyfikacja jest zazwyczaj bardziej skomplikowana. W rzeczywistości zmiany występują w dynamicznym procesie i są trudne do wykrycia i kwantyfikacji. Dzięki BNA wyłaniają się interesujące zmiany w zachowaniu na poziomie przedsiębiorstwa obejmujące skłonność do inwestowania, strategię zatrudniania i zdolność do przyciągania nowych klientów. Ta innowacyjna technika umożliwiła zidentyfikowanie mechanizmów i warunków stojących za powstawaniem rezultatów projektu. Ogółem, badanie przeprowadzone na apulijskim instrumencie TITLE II pokazuje, że w przypadkach, w których początkowa teoria nie jest szczegółowo określona, a instrument wsparcia nie wskazuje konkretnego mechanizmu zorientowanego na osiągnięcie pożądaných efektów, wówczas droga, jaką może podążać ten instrument dla osiągnięcia celów, nie jest drogą prostą. Z uwagi na szeroki zakres celów możemy powiedzieć, że instrument ten nie jest szczególnie skuteczny pod względem wywoływania długofalowych zmian strukturalnych, chociaż z ekonomicznego i społecznego punktu widzenia był odpowiedni dla wsparcia przetrwania wielu mikro- i małych przedsiębiorstw w kluczowym okresie. Niemniej, myśląc w kategoriach portfela instrumentów, można



zaobserwować, że inne instrumenty wsparcia mogą odegrać o wiele ważniejszą rolę w generowaniu głębokich zmian w zachowaniu firm. Taki instrument jak TITILE II może stworzyć przestrzeń, która umożliwi wyłonienie się przedsiębiorstw, chętniej reagujących na inne rodzaje polityki i w sposób bardziej ambitny skupiające się na zmianach strukturalnych. Dzięki mechanizmom wykrytym za pomocą analizy sieci bayesowskiej można założyć, że w przypadku wprowadzenia większej liczby ograniczeń w odniesieniu do wspieranych sektorów i rodzajów inwestycji, w celu zawężenia dostępności instrumentu do przedsiębiorstw i działań o najwyższym potencjale wzrostu i wkładu w konkurencyjność regionalnej gospodarki, skuteczność TITILE II byłaby wyższa.

## O autorkach

### Emanuela Sirtori

---

Emanuela Sirtori posiada tytuł magistra ekonomii Uniwersytetu w Mediolanie i obecnie jest doktorantką z dziedziny ekonomii technologii, innowacji i rozwoju w Szkole Zarządzania UNU-MERIT (Maastricht). W 2009 r. uczestniczyła w Szkole Letniej Metod Ilościowych ewaluacji wpływu funduszy UE z wykorzystaniem podejścia kontryfaktycznego. Obecnie jest partnerem i kierownikiem projektów w CSIL w dziedzinie rozwoju regionalnego, polityki przemysłowej i innowacji, ekonomiki małych przedsiębiorstw. Przez lata uczestniczyła w różnych badaniach ewaluacyjnych na zlecenie Parlamentu Europejskiego i Komisji Europejskiej. Odgrywała główną rolę w ewaluacji *ex post* programów Polityki Spójności wspierających innowacje i rozwój MŚP w latach 2007-2013. W ramach tego badania ewaluacyjnego testowała analizę sieci bayesowskich jako innowacyjnego narzędzia do przetwarzania danych na poziomie przedsiębiorstwa.

### Elena Vallino

---

Elena Vallino posiada tytuł magistra nauk politycznych Uniwersytetu Turyńskiego oraz tytuł magistra w dziedzinie ekonomii międzynarodowej i rozwoju Uniwersytetu Nauk Stosowanych w Berlinie. Ukończyła studia doktoranckie w dziedzinie ekonomii systemów złożonych na Uniwersytecie Turyńskim, prowadząc niektóre badania w Belgii i w Arizonie (USA). Obecnie jest starszym pracownikiem naukowym (Senior Researcher) w CSIL, specjalizującym się w obszarze ewaluacji polityk publicznych i Polityki Spójności UE. W przeszłości pracowała dla włoskich organizacji pozarządowych i niemieckiej Agencji Współpracy Technicznej w Indonezji. Prowadziła zajęcia dydaktyczne i uzyskała stypendia badawcze, a także publikowała artykuły naukowe w dziedzinie ekonomii krajów rozwijających się, ekonomii środowiska, ekonomii transportu i modelowania agentowego.

### Silvia Vignetti

---

Silvia Vignetti posiada tytuł magistra ekonomii Uniwersytetu Bocconi. W przeszłości wykładała na Uniwersytecie w Mediolanie w dziedzinie makroekonomii, ekonomii integracji europejskiej i analizy kosztów i korzyści. Obecnie jest partnerem CSIL i dyrektorem jednostki ewaluacyjnej. Jest ekspertem w dziedzinie rozwoju regionalnego, Polityki Spójności oraz ewaluacji projektów i programów. Zdobyła odpowiednie doświadczenie w tej dziedzinie i zajmowała się kilkoma pakietami roboczymi ewaluacji *ex post* programów Polityki Spójności UE w latach 2007-2013 i 2000-2006, w zakresie wsparcia biznesu, rozwoju infrastruktury, kultury i turystyki, zmian demograficznych. Brała udział w kilku badaniach na zlecenie Parlamentu Europejskiego, dotyczących mechanizmów skuteczności i realizacji Polityki Spójności, a obecnie zajmuje się badaniem stanowiącym wkład w debatę na temat reformy Polityki Spójności UE po 2020.

## Bibliografia:

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?", w: Donaldson S., Christie C., Mark M. (red.) *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. Komisja Europejska „Support to SMEs – Increasing Research and Innovation in SMEs and SME Development.” Trzecie sprawozdanie okresowe - Pakiet roboczy 2. Przedłożone przez CSIL w partnerstwie z CSES i ZEW. Bruksela 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> (dostęp 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.



## Zastosowanie analizy kontrybucji jako podejścia ewaluacji opartej na teorii w ewaluacji ex post wsparcia dla dużych przedsiębiorstw w ramach programów Polityki Spójności UE 2007-2013

### Wstęp

Ewaluacja oparta na teorii (*theory-based evaluation, TBE*) staje się coraz bardziej popularnym podejściem stosowanym przy badaniu europejskich funduszy strukturalnych i inwestycyjnych (*European Fund for Strategic Investments, EFSI*). Choć koncepcja TBE nie jest nowa – w literaturze pojawiła się już na początku lat 70-tych<sup>105</sup> i była szeroko wykorzystywana np. w ewaluacji pomocy rozwojowej (*development evaluation*) – to jej zastosowanie w praktyce, w kontekście ewaluacji europejskich instrumentów wsparcia, wciąż stanowi duże wyzwanie.

Niniejszy rozdział ma na celu przedstawienie przykładu zastosowania TBE w ewaluacji ex post programów Polityki Spójności Unii Europejskiej (UE) na lata 2007-2013, wykonanej w ramach tzw. Pakietu Roboczego nr 4 (*Work Package 4, WP4*) – *Wsparcie dla dużych przedsiębiorstw*<sup>106</sup>. Ewaluacją objęto osiem Państw Członkowskich UE, w których wsparcie dla przedsiębiorstw stanowiło 75% łącznych środków wydatkowanych na ten cel w ramach Polityki Spójności. W siedmiu z analizowanych krajów (Polska, Niemcy, Portugalia, Hiszpania, Węgry, Włochy i Czechy) odnotowano największe wydatki bezwzględne na wsparcie dużych przedsiębiorstw spośród wszystkich Państw Członkowskich. Natomiast w Austrii (ósmym badanym kraju) zarejestrowano relatywnie najwyższe proporcjonalne wydatki w tym obszarze wsparcia.

Przeprowadzona ewaluacja bazowała na koncepcji ewaluacji opartej na teorii, kładąc szczególny nacisk na zagadnienie rekonstrukcji oraz weryfikacji zidentyfikowanych teorii zmiany. Ściślej rzecz biorąc, w ewaluacji zastosowano analizę kontrybucji (*contribution analysis*)<sup>107</sup> – jeden z możliwych wariantów ewaluacji opartej na teorii. Jest ona szczególnie przydatna w przypadku złożonych pytań przyczynowych.

<sup>105</sup> F. Leeuw, *Theory-Based Evaluation - Guidance produced for DG Regional Policy*, Komisja Europejska 2012.

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

<sup>107</sup> J. Mayne, *Contribution analysis: addressing cause and effect*, w: R. Schwartz, K. Forss and M. Marra (red.) *Evaluating the Complex*, Transaction Publishers, New Brunswick, NJ 2011, pp. 53–96.

Przed przedmiotową ewaluacją, przeprowadzone zostały ewaluacje kontrfaktyczne dotyczące efektów wsparcia dla dużych przedsiębiorstw. W wielu przypadkach wynikało z nich, że wsparcie dla tej grupy firm nie miało żadnego wpływu lub był on relatywnie niewielki<sup>108</sup>. Wyniki te miały istotne konsekwencje biorąc pod uwagę np. ustalenie skuteczności wsparcia, jednak na ich podstawie nie było możliwym otworzenie tzw. „czarnej skrzynki” i wyjaśnienie zmiany w zachowaniu dużych przedsiębiorstw. W związku z tym, celem podjętej ewaluacji ex post WP4 była ocena zasadności i dowodów na skuteczność Polityki Spójności w tym obszarze wsparcia. Przesłanki stojące u podstaw polityk publicznych oraz dostępna literatura, dotycząca przedmiotu wsparcia dużych przedsiębiorstw, posłużyły do opracowania typowych teorii zmiany, ilustrujących dlaczego i jak powinno działać tego typu wsparcie. Ustalone założenia zostały następnie krytycznie ocenione w oparciu o dowody wynikające ze studiów przypadku, przeprowadzonych w ramach ewaluacji. Dowody te skonfrontowano również z wynikami dostępnych ewaluacji kontrfaktycznych<sup>109</sup>.

Pozostała część rozdziału ma następującą strukturę: Części 2 i 3 przedstawiają uzasadnienie dla stosowania podejścia TBE. Część 4 opisuje proces rekonstrukcji teorii zmiany. Część 5 przedstawia proces weryfikacji teorii zmiany (z polskim przykładem). Część 6 zawiera wybrane wnioski z ewaluacji WP4, które pomagają zaprezentować istotne z punktu widzenia polityk publicznych ustalenia, wynikające z wykonanej weryfikacji teorii zmiany. Ostatnia część podsumowuje nasze wnioski w zakresie stosowania TBE i analizy kontrybucji w praktyce ewaluacji.

## **Pułapka „czarnej skrzynki” w ewaluacji**

Kontrfaktyczna ewaluacja wpływu (*Counterfactual Impact Evaluation, CIE*) pozwala odpowiedzieć na pytanie czy „dana interwencja działa”<sup>110</sup>. Innymi słowy, pozwala ocenić czy pomysł na rozwiązanie danego problemu społeczno-gospodarczego przyniósł oczekiwane skutki czy też nie. Choć taka wiedza często jest wystarczająca, by podjąć decyzję o kontynuacji lub zawieszeniu interwencji w dotychczasowej formie, to w większości przypadków nie pozwala ona na wyjaśnienie „dlaczego dana interwencja działa”, „w jaki sposób działa” i wreszcie „jak ją udoskonalić”. Jest to problem tzw. „czarnej skrzynki”: wyniki CIE pozwalają prześledzić przejście między nakładami i skutkami interwencji, jednak szczegółowe zależności między nimi pozostają nieznane. Aby znaleźć odpowiedzi na wskazane wyżej pytania, ewaluacje, oprócz ustalenia wpływu

<sup>108</sup> D. Mouqué, *What are counterfactual impact evaluations teaching us about enterprise and innovation support?*, DG for Regional and Urban Policy, Bruksela 2012 [http://ec.europa.eu/regional\\_policy/sources/docgener/focus/2012\\_02\\_counterfactual.pdf](http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_02_counterfactual.pdf).

<sup>109</sup> 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 - Komisja Europejska, DG ds. Polityki Regionalnej i Miejskiej 2012.

<sup>110</sup> A. Martini, Wprowadzenie do podstawowych pytań stawianych w ewaluacji oddziaływania interwencji: Czy interwencja coś zmieniła? A jeśli tak, to dla kogo?, w: A. Haber., R. Trzcirski (red.), *Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2011, ss. 23-31.

interwencji, muszą skoncentrować się na mechanizmach przyczynowo-skutkowych, dzięki którym oczekiwane efekty faktycznie się materializują. Jest to istotne, ponieważ w większości przypadków interwencje (programy, projekty, itp.) nie są w procesie projektowania polityk publicznych ani całkowicie odrzucane, ani nie są też kontynuowane w niezmienionej formie. Wyzwania społeczno-gospodarcze, którym interwencje publiczne starają się stawić czoła (bezrobocie, ubóstwo, wykluczenie społeczne, niska innowacyjność i konkurencyjność przedsiębiorstw, itp.) mają w większości przypadków trwały charakter. Z tego powodu kolejne programy publiczne stanowią zwykle ewolucję a nie rewolucję, względem uprzednio zrealizowanych interwencji. Ewaluacja może więc stanowić ważne źródło dowodów w tym procesie i tym samym musi udzielać odpowiedzi na wszystkie wskazane powyżej złożone pytania.

W celu pełnego wykorzystania wyników CIE oraz zrozumienia, co stoi za powodzeniem lub porażką interwencji publicznych, konieczne jest równoległe wykonanie analizy procesów i mechanizmów odpowiadających za wystąpienie zdefiniowanych zmian. Istotnym zwolennikiem takiego podejścia jest np. Howard White<sup>111</sup>, który propaguje wykorzystanie ewaluacji łączących podejście CIE i TBE - tzw. ewaluacji wpływu opartych na teorii (*Theory-based Impact Evaluation, TBIE*). Podejście sugerowane przez White'a rekomendowane jest także przez Komisję Europejską dla ewaluacji programów operacyjnych realizowanych w ramach perspektywy finansowej 2014-2020<sup>112</sup>.

## Ewaluacja oparta na teorii

Koncepcja prowadzenia badań z zastosowaniem podejścia opartego na teorii była rozwijana przez ostatnie 40 lat. W tym czasie wypracowano kilka podejść metodycznych, takich jak np. ewaluacja wspierana teorią<sup>113</sup>, ewaluacja oparta na teorii<sup>114</sup> czy też ewaluacja realistyczna<sup>115</sup>. Podejścia te łączy główny element wspólny: wszystkie traktują interwencje publiczne jako swoistą „teorię”, możliwego rozwiązania zidentyfikowanego problemu społeczno-gospodarczego<sup>116</sup>.

Podczas projektowania polityk publicznych, często spotykane jest ogólne założenie, że dana interwencja jest konieczna (a w niektórych przypadkach również wystarczająca), aby rozwiązać określony problem społeczno-gospodarczy. Wiele z tych polityk skier-

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

<sup>112</sup> Komisja Europejska, *Guidance document on monitoring and evaluation. European Cohesion Fund European Regional Development Fund, Concepts and Recommendations*, Belgia 2014.

<sup>113</sup> 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, ss. 106-122.

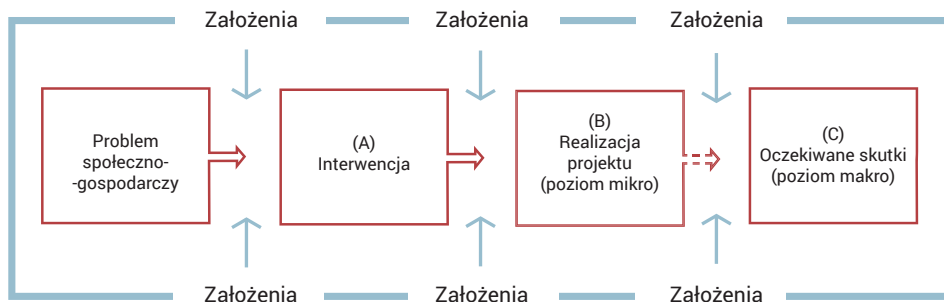
<sup>114</sup> C. H. Weiss, Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families, w: J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (red.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts*, Nowy Jork, NY: Aspen Institute 1995, ss. 65-92.

<sup>115</sup> R. Pawson, N. Tilley, *Realistic Evaluation*, Londyn 1997.

<sup>116</sup> J. Górniak, Ewaluacja w cyklu polityk publicznych, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*. Uniwersytet Ekonomiczny w Krakowie, Małopolska Szkoła Administracji Publicznej 2007.

rowanych jest do osób (np. w formie szkoleń) lub podmiotów (np. wsparcie dla MŚP, organizacji pozarządowych itp.). Zakłada się, że zmiany na poziomie mikro (tj. w grupie beneficjentów interwencji) są w stanie osiągnąć wystarczającą „masę krytyczną”, aby przynieść wymierne i widoczne efekty również na poziomie makro. Z kolei zmiany na poziomie makro są zazwyczaj utożsamiane z oczekiwanymi skutkami interwencji.

### Schemat 1. Ogólna logika realizacji programu



Biorąc za przykład program wsparcia działalności badawczo-rozwojowej w dużych przedsiębiorstwach, w podejściu opartym na teorii interwencji definiuje się jako pewien impuls (zdarzenie A) mający na celu zachęcanie przedsiębiorstw do większej aktywności w dziedzinie badań i rozwoju. Projekty realizowane przez przedsiębiorstwa (inwestycje w badania i rozwój) powinny zwiększać ich potencjał innowacyjny, ale również skutkować powstaniem innowacyjnych produktów lub usług, które, po wprowadzeniu na rynek, powinny mieć pozytywny wpływ na konkurencyjność, ogólne wyniki finansowe i dalszy rozwój działalności gospodarczej firm (zdarzenie B). Jednocześnie oczekuje się, że udzielone wsparcie poprawi kondycję nie tylko beneficjentów pomocy publicznej, ale przełoży się również na gospodarkę regionalną lub krajową oraz przyniesie efekty dodatkowe – np. w grupie działających lokalnie firm sektora MŚP czy też wreszcie wzmocni gospodarkę w wymiarze handlu zagranicznego (zdarzenie C).

Przedstawioną powyżej uproszczoną ogólną logikę interwencji można oczywiście dalej uszczegóławiać. Z jednej strony osiągnięcie założonych w interwencji skutków zależy bowiem od wielu innych (często powiązanych) czynników zewnętrznych, a przez to wiąże się z koniecznością przyjęcia dodatkowych założeń, bez których interwencja nie będzie działać w praktyce. Większość z tych założeń nie jest najczęściej wyrażona wprost w realizowanych programach publicznych, co utrudnia ocenę ich znaczenia w osiągnięciu zamierzonych skutków.

Z drugiej strony interwencje publiczne są znacznie bardziej skomplikowane, gdyż obejmują wiele pośrednich „mikro-etapów”, które wchodzą ze sobą w związki przyczynowe. Łańcuch przyczynowy tych „mikro-etapów”, poczynwszy od dofinansowania publicznego, w założeniu prowadzić ma do osiągnięcia zamierzonych rezultatów. In-



terwencje często również wywołują skutki pośrednie - planowane lub nieplanowane - które dodatkowo komplikują strukturę łańcuchów przyczynowych interwencji.

Celem TBE jest odtworzenie szczegółowej, przyczynowej „teorii zmiany” (*Theory of Change*) interwencji oraz jej weryfikacja z wykorzystaniem dostępnych dowodów. Opracowanie Leeuwa dotyczące „Ewaluacji opartej na teorii”<sup>117</sup> definiuje (za Carol Weiss) „teorię zmiany” jako: „Sposób opisanie zbioru założeń, które wyjaśniają zarówno mini-etapy prowadzące do osiągnięcia długoterminowego celu, jak i połączenia między działaniami interwencji lub programu a rezultatami, które występują na każdym z analizowanych etapów”.

Spośród wielu wariantów ewaluacji opartej na teorii, w ewaluacji WP4 w celu oceny wsparcia dla dużych przedsiębiorstw wybrano analizę kontrybucji, ze względu na jej zdolność do rozkładania złożonych łańcuchów przyczynowych właśnie na „mikro-etapy”, internalizacji założeń i czynników zewnętrznych w jednolitych ramach („Teorii zmiany”) i ich rygorystycznej weryfikacji z wykorzystaniem zróżnicowanych źródeł informacji. Pod względem technicznym analiza kontrybucji może:

- wyjaśnić, częścią jakiego „pakietu przyczynowego”<sup>118</sup> są poszczególne programy,
- określić interakcje między składnikami poszczególnych pakietów przyczynowych,
- ustalić, w jakim stopniu programy były faktycznie konieczne i/lub wystarczające w procesie rozwiązywania danego problemu,
- ocenić prawdopodobny wkład programów.

Analiza kontrybucji, jako specyficzny wariant ewaluacji opartej na teorii, koncentruje się na „ustaleniu powiązań przyczynowych między zaobserwowanymi zmianami a konkretnymi interwencjami oraz przedstawieniu wniosków, które wyjaśniają ich wystąpienie”<sup>119</sup>. Podejście to bazuje na założeniach, które należy uwidocznic w ewaluacji zarówno jako warunki konieczne jak i swoiste ograniczenia. Jak pisze Leeuw<sup>120</sup>, „można w uzasadniony sposób postawić wniosek o wystąpieniu kontrybucji interwencji, jeśli:

- istnieje uzasadniona teoria zmiany dla danego programu: kluczowe założenia stojące za tym dające interwencja ma zadziałać, mają sens, są wiarygodne, mogą zostać poparte dowodami i zostały uzgodnione przez co najmniej część kluczowych aktorów,
- zaplanowane w programie działania zostały wdrożone zgodnie z przyjętą teorią zmiany,

<sup>117</sup> DG REGIO, *EVALSED Sourcebook: Method and Techniques, Regional and Urban policies*, Bruksela 2013.

<sup>118</sup> Wiele interwencji nie działa w pojedynkę, a pożądane rezultaty będą stanowić wynik połączenia czynników przyczynowych, obejmujących wiele powiązanych interwencji oraz wydarzeń i warunków zewnętrznych dla interwencji. W istocie wiele interwencji ma stanowić część takiego „pakietu przyczynowego” i nawet jeśli nie jest to planowane, ich ewaluacja musi wziąć te inne czynniki pod uwagę (J. Mayne, Contribution analysis: Coming of age?, *Evaluation* 2012 18: 270).

<sup>119</sup> *Ibid.*, ss. 55.

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

- teoria zmiany – lub jej kluczowe elementy – jest poparta i potwierdzona dowodami na temat zaobserwowanych skutków i podstawowych założeń, zarówno przez ekspertów, jak i fakty: zdefiniowane w łańcuchu przyczynowym rezultaty wystąpiły a teoria zmiany nie została obalona,
- przeanalizowano inne czynniki zewnętrzne i albo okazało się, że nie miały one żadnego znaczącego wpływu na skutki interwencji albo ich względny wpływ na jej skuteczność został rozpoznany.<sup>121</sup>

Podejście metodologiczne sugerowane przez Mayne'a niekoniecznie przestrzega ścisłej, realizowanej krok po kroku logiki. Stanowi raczej wieloetapowy, iteracyjny proces prowadzenia prac analitycznych.

Na poziomie ogólnym występuje jednak kilka generalnych kroków, których przestrzegano również podczas ewaluacji WP4:

1. *Wskazanie problemu przyczynowo-skutkowego, będącego przedmiotem weryfikacji*  
Potwierdź problem atrybucji, określ konkretne pytanie o relację przyczynowo-skutkową, na które należy udzielić odpowiedzi, określ wymagany poziom ufności, przeanalizuj rodzaj oczekiwanego wpływu, ustal inne kluczowe czynniki i oceń prawdopodobieństwo wystąpienia oczekiwanego wpływu w odniesieniu do wielkości programu.
2. *Opracowanie teorii zmiany*  
Opracuj teorię zmiany i łańcuch rezultatów interwencji, określ poziom szczegółowości, określ oczekiwany wpływ programu, wskaż założenia dla teorii zmiany, uwzględnij inne czynniki mogące wpływać na rezultaty interwencji, określ w jakim stopniu teoria zmiany jest kwestionowana, oceń schemat według jakiego ma oddziaływać interwencja.
3. *Zgromadzenie istniejących dowodów, służących do weryfikacji teorii zmiany*  
Zgromadź dowody (np. wyniki ewaluacji kontrfaktycznych, innych ewaluacji wpływu, badań).
4. *Ocena schematu oddziaływania interwencji*  
Oceń logikę powiązań występujących w teorii zmiany, wiarygodność ogólnego sposobu jej działania, możliwe braki w teorii. Sprawdź również czy główni aktorzy związani z realizacją interwencji zgadzają się co do sposobu w jaki powinna ona działać.
5. *Poszukiwanie dodatkowych dowodów empirycznych*  
Określ, jakie nowe dane są potrzebne, zgromadź dowody (np. za pomocą studiów przypadku, wywiadów z zainteresowanymi podmiotami).

---

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

## 6. *Przegląd i wzmocnienie schematu oddziaływania interwencji*

Przejrzyj ponownie schemat działania interwencji, wróć do etapu 4, jeśli to konieczne.

### **Rekonstrukcja teorii programu**

Właściwa teoria programu musi spełniać pewne kryteria: musi być wiarygodna, osiągalna i możliwa do weryfikacji. Proces rekonstrukcji teorii programu ma charakter głównie koncepcyjny. Proces ten jest zwykle iteracyjny, tzn. teoria zmiany jest rekonstruowana i udoskonalana przez cały czas trwania ewaluacji. Rekonstrukcja teorii programu musi obejmować:

1. opis łańcucha przyczynowo-skutkowego interwencji, w tym skutki pośrednie;
2. identyfikację kluczowych założeń (i czynników zewnętrznych) interwencji.

#### *Tworzenie łańcucha przyczynowo-skutkowego interwencji*

Teoria programu powinna definiować problem społeczno-gospodarczy, który wymaga rozwiązania (wskazując pożądaną zmianę) oraz powinna określać sekwencję zdarzeń (łańcuch przyczynowy), których realizacja prowadzi do rozwiązania tego problemu. Teoria powinna również określać skutki pośrednie, czy też szersze korzyści wynikające z realizacji interwencji. Na każdym etapie łańcuch przyczynowo-skutkowy powinien określać wszystkie kluczowe czynniki i przyczyny oczekiwanych zmian, zwracając uwagę na sposób, w jaki dana zmiana materializuje się u wspieranych jednostek lub podmiotów. Według Carol Weiss, kluczową kwestią jest wyszczególnienie wszystkich istotnych mini-etapów tworzących charakterystyczny łańcuch przyczynowo-skutkowy, w którym poszczególne działania stanowią rezultat wcześniejszych działań podjętych w ramach danej interwencji<sup>122</sup>.

Rekonstrukcja teorii zmiany w ewaluacji WP4 stanowiła intensywną fazę prac, obejmującą wiele iteracji. Każde źródło informacji dodawało kolejne warstwy do rekonstruowanych teorii. W tym procesie wykorzystano następujące, kluczowe źródła informacji:

- Dokumenty programowe (Programy Operacyjne, dokumenty konkursowe itp.) zawierały informacje na temat ogólnych celów interwencji, chociaż często brakowało w nich szczegółowych informacji, dotyczących wsparcia dużych przedsiębiorstw (uzasadnienie, łańcuch przyczynowy, założenia, itp.). Stanowiły one główne źródła przy identyfikacji nakładów, zrealizowanych działań i podstawowych rezultatów.
- Autorzy programów i inni przedstawiciele Instytucji Zarządzających w Państwach Członkowskich okazali się ważnym źródłem informacji dla udzielenia odpowiedzi na pytanie: „Dlaczego duże przedsiębiorstwa są wspierane z Europejskiego Funduszu Rozwoju Regionalnego (EFRR)?”. Ogólnie rzecz biorąc, mogli oni przywołać istotne informacje (często wcześniej nigdzie nie wyrażone wprost) na temat „pla-

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

nowanej czy też oczekiwanej zmiany”, które stanowiły czynnik napędowy procesu planowania i ewentualnych modyfikacji programów.

- Przegląd literatury był pomocny w ponownej ocenie wszystkich planowanych w interwencjach zmian. Jego najważniejszą rolą było jednak ujawnienie założeń, czynników zewnętrznych, skutków pośrednich i większych korzyści, które można było następnie włączyć – w stosownych przypadkach – do analizowanych teorii zmiany.

W sumie w procesie ewaluacji zrekonstruowano 27 teorii zmiany na poziomie programów. W wyniku wieloetapowego procesu klasyfikacji ustalono cztery ogólne teorie wsparcia dla dużych przedsiębiorstw w ośmiu analizowanych krajach. Teorie te obejmują najczęściej występujące elementy teorii na poziomie poszczególnych krajów, a także bazują na rezultatach przeglądu literatury, a zwłaszcza tych dotyczących kluczowych założeń, decydujących czynników oraz pośrednich i szerszych skutków interwencji. Poniżej przedstawiono te cztery ogólne teorie zmiany.

**Tabela 1. Przegląd czterech ogólnych teorii zmiany**

Teoria zmiany	Opis
„LE1: Inwestycje przedsiębiorstw na dużą skalę”	Wsparcie finansowe dla inwestycji dużych przedsiębiorstw (w tym zagranicznych), których podstawowym celem jest zwiększenie zatrudnienia w obszarze objętym programem w długiej perspektywie.
„LE2: Modernizacja technologiczna”	Wsparcie finansowe dla dużych przedsiębiorstw, aby pomóc im we wdrażaniu usprawnień technologicznych w celu wzmocnienia ich konkurencyjności i tym samym potencjału do wzrostu regionalnej gospodarki.
„LE3: Wsparcie innowacji”	Wsparcie finansowe dla dużych przedsiębiorstw, aby pomóc im we wdrażaniu innowacyjnych inwestycji dla potrzeb komercjalizacji nowych produktów i procesów, w celu wzmocnienia potencjału regionalnych innowacji i długotrwałego wzrostu PKB i zatrudnienia (wysokiej jakości miejsca pracy).
„LE4: Inwestycje w potencjał badawczo-rozwojowy”	Wsparcie finansowe dla dużych przedsiębiorstw, aby pomóc im w ustanawianiu, rozbudowie lub doskonaleniu obiektów badawczo-rozwojowych i/lub prowadzeniu działalności badawczo-rozwojowej, w celu rozbudowania regionalnej bazy wiedzy i długotrwałych zdolności badawczo-rozwojowych i innowacyjnych regionu (w tym tworzenie miejsc pracy w dziedzinie badań).

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

Każdą z 27 szczegółowych teorii zmiany na poziomie programu i 4 ogólne teorie opisano za pomocą schematu blokowego. Poniżej prezentujemy przykład takiego sche-

matu dla jednej ze zrekonstruowanych teorii – „LE1: Inwestycje przedsiębiorstw na dużą skalę”. Tego typu ogólną teorię zmiany zidentyfikowano w pięciu krajach (Węgry, Włochy, Polska, Niemcy i Hiszpania). W pierwszych trzech, teoria zmiany obejmowała interwencje realizowane w ramach programów krajowych. W przypadku Niemiec i Hiszpanii ramy analizy stanowiły interwencje na poziomie regionalnym. W przypadku Polski teoria ta odnosi się do inwestycji podjętych w ramach 4. Osi Priorytetowej (np. Działanie 4.5) Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013. Schemat obejmuje następujące kluczowe elementy teorii zmiany:

1. Planowana zmiana: ostatnie pole na końcu łańcucha przyczynowego;
2. Etapy pośrednie służące do osiągnięcia zaplanowanej zmiany: opis nakładów i etapów pośrednich;
3. Relacje przyczynowe: reprezentowane przez trzy rodzaje strzałek:
  - a. przyczyna: zdarzenie *A* to jedna z głównych, podstawowych przyczyn zdarzenia *B* (“musi wystąpić”)
  - b. warunek wstępny: zdarzenie *A* jest ważnym warunkiem wstępnym dla zdarzenia *B*, ale nie jest jego główną przyczyną (jego brak nie wyklucza zdarzenia *B*);
  - c. czynnik wspomagający: zdarzenie *A* wspiera wystąpienie zdarzenia *B*, ale nie jest ani jego przyczyną ani warunkiem wstępnym (“dodatek”).
4. Kluczowe założenia i czynniki zewnętrzne: przedstawione za pomocą wyjaśnionych poniżej zielonych oznaczeń;
5. Pośrednie i szersze skutki interwencji: zmiany będące szerszym następstwem zrealizowanej interwencji, choć bezpośrednio z nią nie powiązane (przedstawione na schemacie za pomocą czerwonych kółek).

Biorąc pod uwagę typowe założenia interwencji bazujących na teorii zmiany „LE1”, można wskazać, że oczekiwanym skutkiem wsparcia dla przedsiębiorstw w formie dotacji i instrumentów finansowych miały być inwestycje rozwojowe w grupie tych firm, w następstwie których w pierwszej kolejności miało wyraźnie wzrosnąć zatrudnienie. Inwestycje na dużą skalę dokonywane były poprzez zakup środków trwałych, w tym maszyn i urządzeń do produkcji. Takie działania miały na celu zwiększenie ogólnego poziomu i zdolności produkcyjnych przedsiębiorstw, jak również poziomu ich rozwoju technologicznego. Poprawa konkurencyjności przedsiębiorstw powinna z kolei przekładać się na poprawę ich wyników finansowych, które umożliwią im stały rozwój i dalsze zwiększanie zatrudnienia. Oczekiwano, że programy wspierające teorię zmian „LE1” wpłyną na zmianę wskaźników zatrudnienia, mierzonych na poziomie krajowym.



## Identyfikacja kluczowych założeń interwencji

Model przedstawiony na rys. 1 przedstawia idealną sytuację, gdzie mikro-etapy teorii zmiany i oczekiwane rezultaty występują po kolei i tworzą bezpośrednie powiązania przyczynowe. Niemniej, przedstawiona powyżej teoria wystąpi w rzeczywistości tylko w określonych warunkach i gdy zostaną spełnione dodatkowe założenia. Dlatego teoria zmiany musi identyfikować założenia, których spełnienie jest istotne dla materializacji tej teorii. W powyższym przykładzie założenia te oznaczono na schemacie kolorem zielonym. W procesie rekonstrukcji teorii zmiany „LE1: Inwestycje przedsiębiorstw na dużą skalę”, w ewaluacji WP4, wyróżniono następujące kluczowe założenia:

1. Zachęty podatkowe są konkurencyjne (międzynarodowo);
2. Strategia firmy wspiera długoterminowy pobyt w kraju;
3. Rozbudowana podstawowa infrastruktura (autostrady, dostęp do lotniska, infrastruktura ICT);
4. Na obszarze znajduje się „dziedzictwo” biznesowe/przemysłowe;
5. Pomocny samorząd (zezwolenia, procedury);
6. Rynek pracy zapewnia pracowników w wymaganej liczbie i o odpowiednim poziomie kwalifikacji;
7. Inwestycje są na tyle duże, aby wpływać na rynek pracy;
8. Ogólne warunki ekonomiczne umożliwiają wzrost.

Jak pokazano na rys. 1, założenia mogą odgrywać różne role w łańcuchu przyczynowym. Niektóre z nich okazały się stanowić ważne warunki wstępne, natomiast inne to tylko czynniki wspomagające. Na przykład założenie nr 8 (*Ogólne warunki ekonomiczne umożliwiają wzrost*) odnosi się do zewnętrznego otoczenia gospodarczego, które może mieć decydujący wpływ na urzeczywistnienie się teorii zmiany i tym samym powinno być traktowane jako istotny warunek wstępny. Jeśli to założenie nie zostanie spełnione – co może mieć miejsce na przykład w sytuacji pogorszenia koniunktury gospodarczej – nawet najlepsze inwestycje nie doprowadzą do oczekiwanych rezultatów. Takie założenia i czynniki zewnętrzne są często wzajemnie ze sobą powiązane, dlatego charakter tych powiązań należy zweryfikować w trakcie ewaluacji.

## Weryfikacja teorii zmiany

Po opisanej powyżej fazie badań koncepcyjnych następuje faza empiryczna, która obejmuje weryfikację zrekonstruowanych teorii zmiany. Według prof. Górniaka, ewaluacja ma na celu sprawdzenie czy w świetle uzyskanego wpływu teorię programu można utrzymać czy też należy ją odrzucić<sup>123</sup>. Etapy metodyczne analizy teorii zmiany za pomocą analizy kontrybucji, zgodnie z sugestią Mayne’a<sup>124</sup>, niekoniecznie

---

<sup>123</sup> J. Górniak, „Ewaluacja w cyklu polityk publicznych”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych - perspektywa regionalna*, Kraków 2007.

<sup>124</sup> J. Mayne, „Contribution analysis: addressing cause and effect”, w: R. Schwartz, K. Forss, M. Marra (red.) *Evaluating the Complex*, Transaction Publishers, New Brunswick, NJ 2011, str. 53-96.

przestrzegają ścisłej logiki, ale raczej sugerują zastosowanie iteracyjnego procesu, obejmującego wiele cykli weryfikacji wcześniej wykonanych zadań. W istocie takie podejście zastosowano zarówno podczas rekonstrukcji jak i badania teorii zmiany w ewaluacji WP4.

Weryfikacja teorii zmiany w trakcie TBE musi być zasilana dwoma rodzajami dowodów.

1. Pierwszą grupę dowodów można określić jako faktyczne<sup>125</sup>. Informują one „w jaki sposób przebiega(ła) interwencja” pod względem rodzaju i funkcjonowania systemu wyboru projektów, rzeczywistego poziomu innowacyjności zrealizowanych projektów lub zakresu wdrożonych inwestycji, itp. W znacznym stopniu dowody faktyczne mają dostarczyć dane, które umożliwią zweryfikowanie czy przyjęte założenia interwencji zostały faktycznie spełnione. Takie dane zostały w ewaluacji częściowo dostarczone przez systemy monitorowania, wcześniej przeprowadzone badania i dodatkowe badania, zrealizowane z wykorzystaniem mieszanego podejścia badawczego.
2. Drugi rodzaj dowodów (czasami określanych jako kontrfaktyczne) odnosi się do problemu ustalenia wpływu interwencji. Ponieważ dokładne wymierne dowody na skuteczność wsparcia dla dużych przedsiębiorstw nie były dostępne w większości analizowanych interwencji, zidentyfikowane teorie zmiany zbadano na podstawie wielu wymiarów jakościowych i w drodze konfrontacji tych źródeł informacji (triangulacja źródeł danych). Badanie przeprowadzono z zastosowaniem wystandaryzowanego schematu, którego główne części objęły: (A) Historię przedsiębiorstwa (kontekst), (B) Nakłady i zrealizowane działania, (C) Bezpośrednie skutki (i ich trwałość), (D) Pośrednie skutki i większe korzyści (i ich trwałość) oraz (E) Podsumowanie i wnioski (w tym możliwe alternatywne wyjaśnienia dla zaobserwowanych zmian). Zgromadzono dowody dla każdego istotnego mikro-etapu, rezultatu, kluczowego założenia i decydującego czynnika w zrekonstruowanych teoriach zmiany, po czym oceniono czy etapy te (skutki) uległy materializacji oraz czy były one w związku przyczynowo-skutkowym z wcześniejszymi etapami/rezultatami w ramach ustalonego łańcucha przyczynowego. Jednocześnie ustalono poziom wiarygodności dostępnych dowodów.

Analizę kontrybucji przeprowadzono za pomocą porównawczych studiów przypadku. Obejmowało to osiem studiów przypadku na poziomie poszczególnych Programów Operacyjnych, w ośmiu Państwach Członkowskich. Punktem centralnym tej struktury było 45 studiów przypadku zrealizowanych w przedsiębiorstwach. W każdej firmie przeprowadzono kilka wywiadów jakościowych – łącznie przeprowadzono ponad 130

---

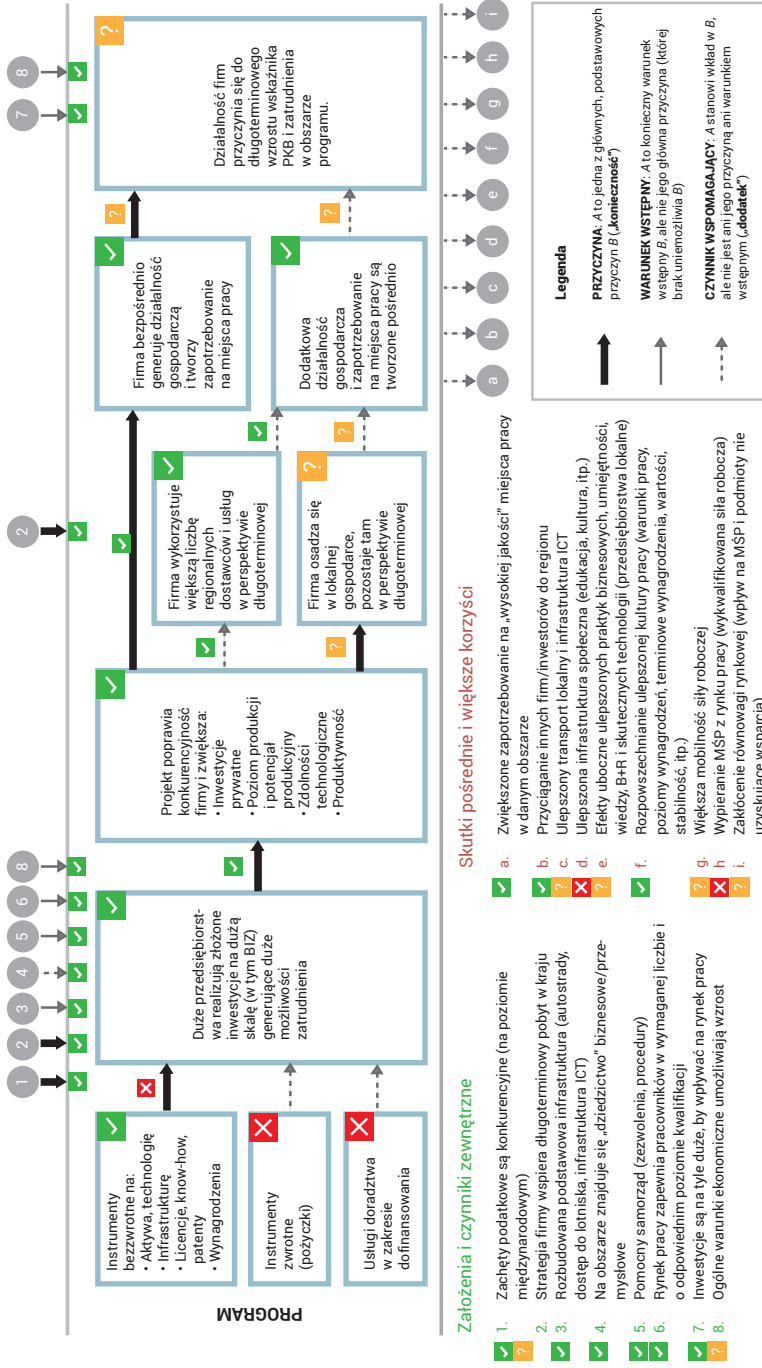
<sup>125</sup> H. White, *Theory-Based Impact Evaluation: Principles And Practice*, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi 2009.



wywiadów. Studia przypadków w przedsiębiorstwach wykorzystywały wystandaryzowane ramy do oceny wkładu wsparcia dla dużych przedsiębiorstw w realizację projektu i w zaobserwowane rezultaty. Ponadto, do analizy włączono dane z monitorowania oraz wnioski z kompleksowego przeglądu literatury obejmującej ponad 100 artykułów naukowych i ewaluacji wsparcia dla dużych przedsiębiorstw. Taką bazę empiryczną stosowano na różnych etapach zastosowanej w ewaluacji triangulacji. Choć pierwsze wywiady z instytucjami zarządzającymi i instytucjami pośredniczącymi pomogły w odkryciu uzasadnień dla udzielonego dofinansowania i ogólnej strategii wsparcia dużych przedsiębiorstw, to wywiady z pracownikami firm umożliwiły zbadanie związków przyczynowych, analizę zmian zachowania dużych przedsiębiorstw, jak również stopnia ich wkładu w rozwój społeczno-gospodarczy regionu. Aby zweryfikować ustalenia z tego etapu, przeprowadzono dodatkowe wywiady z ekspertami-akademikami w celu zrównoważenia potencjalnych obciążeń na poziomie danych uzyskanych od firm. Takie połączone podejście analityczne umożliwiło sformułowanie wniosków bazujących na solidnym fundamencie dowodowym.

Proces badania teorii zmiany udokumentowano ponownie za pomocą schematów blokowych. Poniższy rysunek przedstawia zweryfikowaną teorię zmiany „LE1: Inwestycje przedsiębiorstw na dużą skalę” w przypadku Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013 (Działanie 4.5).

Rysunek 2. Ocena teorii zmiany „LEI: Inwestycje przedsiębiorstw na dużą skalę”, przypadek Polski



Źródło: KPMG/Prognos, Wsparcie dla dużych przedsiębiorstw. 2. Raport okresowy – Pakiet Roboczy 4, Ewaluacja ex-post programów Polityki Spójności 2007-2013, Bruksela 2016

\*Znak zapytania dotyczy sytuacji, gdy nie istniały wystarczające dowody na odrzucenie lub potwierdzenie istnienia związku przyczynowego lub wpływu

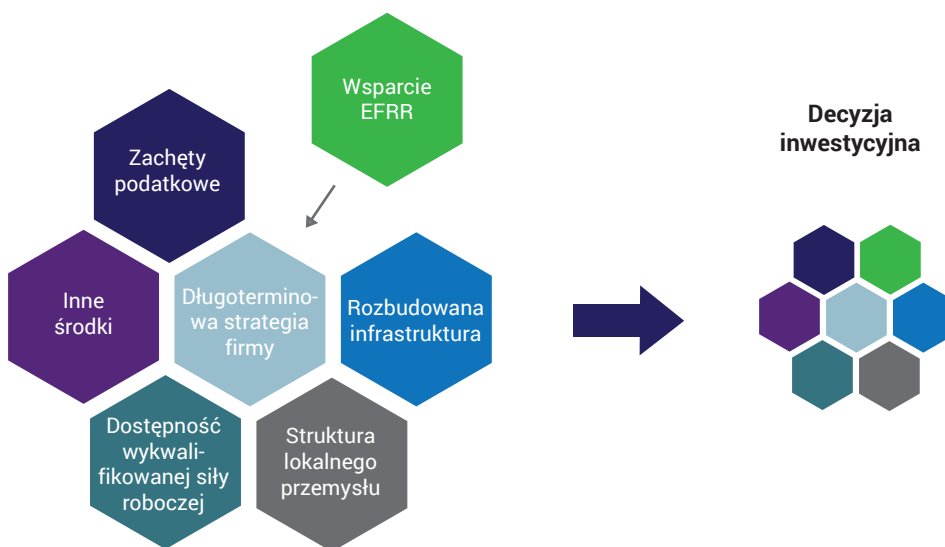
## Wnioski dla polityk publicznych

Wykorzystanie ewaluacji opartej na teorii w badaniu WP4, w połączeniu z dostępnymi ewaluacjami kontrfaktycznymi, umożliwiło wyciągnięcie następujących wniosków dotyczących zidentyfikowanych teorii zmiany w obszarze wsparcia EFRR dla dużych przedsiębiorstw:

1. Występuje szereg czynników determinujących decyzje inwestycyjne dużych przedsiębiorstw w kontekście wsparcia Polityki Spójności UE.

Przeprowadzona analiza wykazała, że w przypadku dużych przedsiębiorstw wsparcie EFRR zazwyczaj pełniło rolę warunku wstępnego dla zrealizowanych inwestycji – ważnego, lecz nie wystarczającego do realizacji danego projektu w firmie. Jak pokazano na rys. 3, wsparcie UE często stanowiło tylko jeden z wielu czynników dla podejmowania decyzji inwestycyjnych, z których najważniejsze to długoterminowa strategia przedsiębiorstwa, dostępność infrastruktury transportowej i struktura lokalnego przemysłu.

**Rysunek 3. Czynniki determinujące decyzje inwestycyjne dużych przedsiębiorstw**



Źródło: KPMG/Prognos, na podstawie dowodów empirycznych z 45 studiów przypadku zrealizowanych w przedsiębiorstwach, 2016.

2. W 20% z analizowanych przypadków zaszła pożądana zmiana zachowania firm, a wsparcie EFRR było jedną z głównych przyczyn wdrożenia projektu.

Wywołana zmiana zachowań w tych przypadkach wynikała ze zmiany zakresu i harmonogramu realizacji projektów strategicznych, które w przeciwnym razie nie zostałyby wdrożone. Ten wzorzec został zaobserwowany najczęściej w dużych przedsiębiorstwach stojących w obliczu problemów finansowych, wynikających z kryzysu gospodarczego.

3. W 50% przypadków wsparcie EFRR wywołało zmiany w zachowaniu przedsiębiorstw, wpływając w szczególności na harmonogram i zakres realizowanych projektów.

W 50% przypadków wsparcie EFRR stanowiło warunek konieczny dla realizacji projektu oraz spowodowało, że wsparcie miało wpływ na kształt projektu. Zazwyczaj wsparcie pełniło rolę katalizatora realizacji projektów, doprowadzając do tego, że duże firmy inwestowały wcześniej niż planowały. Wsparcie oddziaływało także na zakres projektu, zwłaszcza w postaci włączenia przez firmy do projektu nowych zadań projektowych (np. podjęcie współpracy z innymi podmiotami gospodarczymi) lub zdarzeń powiązanych z projektem (np. zatrudnienie większej liczby osób, wykorzystanie technologii przyjaznych środowisku).

4. W 30% przypadków wsparcie EFRR miało niewielki wpływ na zachowanie dużych przedsiębiorstw.

Podczas ewaluacji stwierdzono, że w pozostałych 30% przypadków dofinansowanie UE odgrywało tylko drugoplanową rolę w realizacji projektu i nie stanowiło niezbędnego elementu prowadzącego do dokonania inwestycji. Wpływ wsparcia EFRR na zachowanie dużych przedsiębiorstw był w tych przypadkach niewielki, często dlatego, że realizowane projekty były mocno osadzone w długoterminowych planach przedsiębiorstw. Szczególnie niski przyrost dodatkowości wynikający z otrzymanego wsparcia uzyskano w przypadku projektów koncentrujących się na podstawowych ulepszeniach technologicznych. Był on natomiast znacznie wyższy w przypadku inwestycji przedsiębiorstw na dużą skalę, w projektach wsparcia innowacji oraz wsparcia aktywności badawczo-rozwojowej.

#### Rysunek 4. Wpływ wsparcia EFRR na zachowanie dużych przedsiębiorstw



Źródło: KPMG/Prognos, na podstawie dowodów empirycznych z 45 studiów przypadku, zrealizowanych w przedsiębiorstwach, 2016.

5. W 90% wspieranych projektów odnotowano pozytywne skutki bezpośrednie związane z realizacją projektów, ale – jak opisano powyżej – nie wszystkie zaobserwowane rezultaty zostały w pełni wyjaśnione otrzymanym wsparciem z UE.

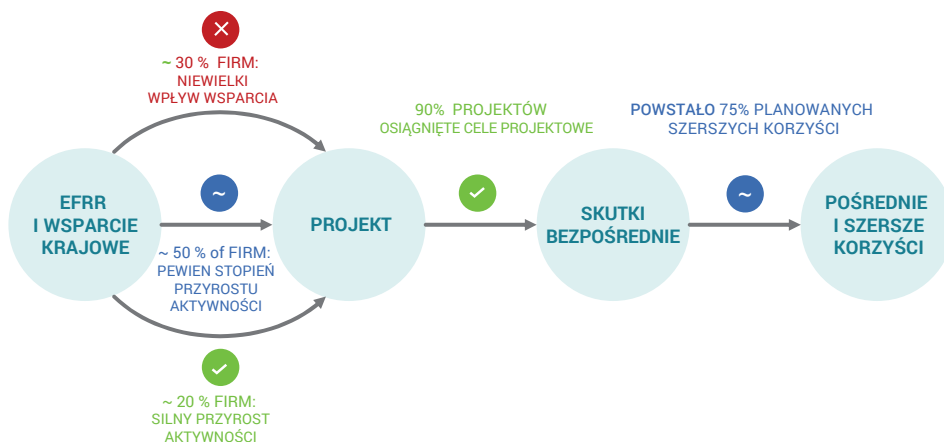
W przypadku większości zrealizowanych projektów wzrosła zarówno zdolność produkcyjna jak i produktywność dużych przedsiębiorstw. Dodatkowo projekty najczęściej wiązały się z istotnym przyrostem nowych miejsc pracy. W przypadku większości firm istniały silne powiązania przyczynowe między projektem i zaobserwowanymi rezultatami. Jednak biorąc pod uwagę fakt, że wsparcie EFRR miało jedynie umiarkowaną zdolność wpływania na zachowanie dużych przedsiębiorstw, należy uznać, że wpływ wsparcia na powyższe aspekty funkcjonowania firm nie miał bezpośredniego charakteru.

6. Pośrednie i szersze korzyści powstały w 75% przypadków, w których spodziewano się ich wystąpienia. Jednocześnie w wielu przypadkach, w których pośrednie i szersze korzyści nie były planowane, skoncentrowanie na nakładach i wynikach bezpośrednich, uniemożliwiło wystąpienie takich efektów.

W 75% projektów, w których przewidywano wystąpienie pośrednich i szerszych korzyści, przynajmniej częściowo zostały one zaobserwowane. W sytuacji, gdzie oczekiwano efektów pośrednich dla MŚP, w większości projektów udało się osiągnąć pozytywne skutki. Podobne wyniki uzyskano w zakresie poprawy mobilności siły roboczej, kultury pracy, rozwoju biznesowej i społecznej infrastruktury, a także w zakresie skutecznego przyciągania innych firm do regionu. Aby mogły wystąpić szersze korzyści z realizacji projektów, spełnione musiały być dodatkowe warunki, takie jak np. wystarczająca zdolność do absorpcji nowej wiedzy wśród podmiotów pośrednio korzystających z pomocy (np. MŚP) oraz uprzednio występujące więzy współpracy.

Powyższe wnioski zilustrowano na poniższym rysunku.

**Rysunek 5. Uproszczony łańcuch przyczynowy i skuteczność wsparcia.**



Źródło: KPMG/Prognos, na podstawie dowodów empirycznych z 45 zakładowych studiów przypadku, 2016.

## Wnioski

Wymogi Komisji Europejskiej, opisane w wytycznych dla planów ewaluacji na lata 2014-2020, z pewnością doprowadzą do rozwoju badań gromadzących wiarygodne dowody na temat faktycznego oddziaływania wdrożonych polityk publicznych.

Doświadczenia ostatnich lat wskazują jednak, że decyzje w obszarze polityk publicznych rzadko kiedy mogą polegać wyłącznie na wynikach ewaluacji wpływu, bazujących na podejściu kontrfaktycznym. Choć badania te umożliwiają w sposób rygorystyczny udzielenie odpowiedzi na pytanie czy dany program faktycznie przyniósł oczekiwane skutki, to jednocześnie mają one ograniczoną zdolność do identyfikacji podstawowych przyczyn powodzenia lub niepowodzenia interwencji. Ewaluacje opar-

te na teorii mogą przewyciężyć to ograniczenie, poprzez rozpatrywanie poszczególnych programów w kategoriach „teorii zmiany”, definiującej sposób w jaki oczekiwane skutki interwencji faktycznie się materializują. Podejście oparte na teorii może wspierać decyzje w obszarze polityk publicznych dzięki możliwości odkrywania złożonych pakietów przyczynowych, które sprawiają, że poszczególne interwencje w rzeczywistości działają. Wykorzystanie podejść opartych na teorii może okazać się szczególnie pożyteczne, w przypadku ich połączenia z ewaluacjami wpływu, bazującymi na podejściu kontrfaktycznym.

Przed przeprowadzeniem ewaluacji WP4, dostępne były wyniki badań kontrfaktycznych poświęconych efektom wsparcia dla dużych przedsiębiorstw. Przeprowadzone analizy w wielu przypadkach wskazywały, że udzielona dużym firmom pomoc publiczna nie miała żadnego wpływu lub był on co najwyżej niewielki. Jednocześnie badania te nie mogły otworzyć „czarnej skrzynki”, tj. wyjaśnić wkładu udzielonego wsparcia w zmianę zachowania dużych przedsiębiorstw.

Dla potrzeb analizy i oceny tego wkładu, podczas ewaluacji WP4 skorzystano z analizy kontrybucji – jednego z wariantów ewaluacji opartej na teorii. W badaniu w pierwszym kroku zrekonstruowano różne teorie zmiany przypisane do poszczególnych schematów wsparcia dużych przedsiębiorstw. Proces ich odtwarzania był niezbędny, gdyż często nie były one *explicite* wyrażone w dokumentach programowych. Zrekonstruowane teorie zmiany opisały planowane rezultaty interwencji, poszczególne „mikro-etapy” realizacji interwencji, kluczowe założenia, czynniki zewnętrzne, możliwe pośrednie i szersze korzyści interwencji, jak również charakter związków przyczynowych między tymi elementami. W drugim kroku przeprowadzone zostały systematyczne, szeroko zakrojone badania jakościowe, w celu zebrania dowodów użytecznych z punktu widzenia weryfikacji i oceny odtworzonych teorii zmiany. Prace empiryczne bazowały w znacznym stopniu na częściowo ustrukturyzowanych wywiadach z pracownikami firm (menadżerowie wysokiego szczebla, osoby zarządzające środkami UE, kierownicy projektów, pracownicy obsługujący zakupione maszyny) i instytucjami wdrażającymi interwencje (instytucje zarządzające, instytucje pośredniczące). Dla zachowania wiarygodności stawianych wniosków, zastosowano triangulację źródeł gromadzonych danych oraz krytyczne podejście w stosunku do zebranych informacji, wspierane przez opinie niezależnych ekspertów, naukowców, przedstawicieli lokalnych instytucji, itp.

Wreszcie w ewaluacji wskazano czy oczekiwane skutki interwencji zostały osiągnięte, czy teoria zmiany jako całość zapewniła uzasadnione wyjaśnienie wystąpienia tych efektów, i czy istniały jakiegokolwiek inne alternatywne wyjaśnienia dla ich wystąpienia, które nie zostały uwzględnione w oryginalnej teorii zmiany. Ostatecznie na podstawie wykonanej analizy kontrybucji możliwe było przedstawienie wniosków w zakresie wkładu wsparcia UE w zaobserwowane zmiany.

Ewaluacja oparta na teorii wykazała, że kompleksowe studium przypadku, zakorzenione w empirycznych badaniach jakościowych, ale także bazujące na dostępnych ewaluacjach kontrfaktycznych, mogło stworzyć spójną i wiarygodną narrację na temat wkładu wsparcia UE w zaobserwowane zmiany – zarówno w zachowaniu dużych przedsiębiorstw jak również w środowisku społeczno-gospodarczym wspieranych regionów. Dogłębne zrozumienie łańcucha przyczynowego wsparcia dla dużych przedsiębiorstw, może następnie stanowić podstawę rekomendacji istotnych dla polityk publicznych, które z kolei, mogą pomóc decydującym w opracowaniu podobnych interwencji w przyszłości.



## O autorach

### András Kaszap

---

András Kaszap kieruje unijną grupą ds. konsultacji i ewaluacji KPMG na Węgrzech jako Starszy Menedżer (*Senior Manager*). Posiada doświadczenie w planowaniu, prowadzeniu i zarządzaniu ewaluacjami (*ex ante*, okresowe, *ex post*, tematyczne i meta-ewaluacje) oraz ocenami wpływu na przedsiębiorstwa. Jako praktyk w zakresie ewaluacji stosował najnowocześniejsze narzędzia i podejścia ilościowe i jakościowe, w tym ewaluacje kontrfaktyczne oraz oparte na teorii, w celu stworzenia wartości dla klientów w sektorze publicznym i prywatnym. Zarządzał i przeprowadził wiele ewaluacji w kontekście międzynarodowym, obejmującym 13 krajów europejskich. Specjalizuje się w ewaluacji Polityki Spójności UE oraz Wspólnej Polityki Rolnej UE.

### Jan-Philip Kramer

---

Posiada tytuł doktora geografii ekonomicznej Uniwersytetu Leibniza w Hanowerze i UC Berkeley. Starszy Kierownik Projektów w Wydziale ds. Gospodarki / Innowacji / Regionu oraz kierownik Biura UE w Brukseli, zajmującego się pełnym cyklem polityki, od badań prognostycznych, strategii politycznych oraz programowania, monitorowania i ewaluacji (*ex ante* i *ex post*, łączących podejście TBE i CIE) do ocen wpływu. Zainteresowania tematyczne koncentruje w obszarze funduszy strukturalnych i inwestycyjnych (zwłaszcza EFRR), strategii inteligentnych specjalizacji (S3) i klastrów, a także scenariuszowych strategiach rozwoju (*scenario-based strategy development*) dla branż i regionów. Posiadający ponad 10-letnie doświadczenie zawodowe Jan-Philip Kramer pracował dla kilku klientów międzynarodowych i niemieckich, w tym dla Komisji Europejskiej (DG REGIO, DG RTD), Sekretariatu EUREKA, niemieckiego Federalnego Ministerstwa Edukacji i Badań, jak również dla wielu klientów regionalnych.

### Rafał Trzciński

---

Absolwent Instytutu Socjologii Uniwersytetu Warszawskiego. Od marca 2017 r. dyrektor Centrum Badań i Edukacji Statystycznej Głównego Urzędu Statystycznego. W latach 2016-2017 dyrektor do spraw badań i analiz ilościowych w Fundacji IDEA Rozwoju. Ma wieloletnie doświadczenie w prowadzeniu badań i ewaluacji programów publicznych, zdobyte podczas pracy w sektorze publicznym (Polska Agencja Rozwoju Przedsiębiorczości w latach: 2005-2014; Instytut Badań Edukacyjnych w latach: 2014-2015), w organizacjach pozarządowych (Fundacja IDEA Rozwoju w latach: 2014-2017; Fundacja ISEE w latach: 2013-2016) oraz współpracy z podmiotami prywatnymi (EGO S.C., KPMG i inne). W ramach prowadzonych projektów badawczych współpracował również z Komisją Europejską oraz Bankiem Światowym. Specjalizuje się w metodologii statystycznych badań przyczynowych. Autor oraz współredaktor publikacji i artykułów w tym zakresie. Koordynował, jak również osobiście w roli anali-

tyka przeprowadził kilkanaście badań ewaluacyjnych, w których zastosowane zostało podejście kontrfaktyczne. Prowadzi szkolenia i wykłady na temat metodologii badań ewaluacyjnych wykorzystujących podejście kontrfaktyczne. Członek Polskiego Towarzystwa Ewaluacyjnego.

## Bibliografia:

1. Bienias S., Herman-Pawłowska K. (red.), *Ocena wpływu regulacji. Poradnik dla administracji publicznej*, Warszawa 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, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Uniwersytet Ekonomiczny w Krakowie, Małopolska Szkoła Administracji Publicznej, Kraków 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?, w: A. Haber., R. Trzciński (red.), *Ocena wpływu i prognozowanie efektów w badaniach ewaluacyjnych*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 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 ds. Polityki Regionalnej i Miejskiej, Bruksela 2012.
10. Olejniczak K., *Mechanizmy wykorzystania ewaluacji. Studium ewaluacji śródkresowych INTERREG III*, Centrum Europejskich Studiów Regionalnych i Lokalnych, Uniwersytet Warszawski, Wydawnictwo SCHOLAR, Warszawa 2008.
11. Pawson, R., Tilley, N., *Realistic Evaluation*, Londyn 1997.
12. Weiss, C. H., Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families, w: J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (red.), *New approaches to evaluating community initiatives: Volume 1, concepts, methods, and contexts*, Aspen Institute, Nowy Jork, NY 1995.
13. White H., *Theory-Based Impact Evaluation: Principles And Practice*, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi 2009.

## Inne:

1. GUS, *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych*, Warszawa 2015.
2. ISR, *Ewaluacja zmian w systemie wyboru projektów w ramach wybranych działań PO IG 2007-2013*, raport zlecony przez Ministerstwo Infrastruktury i Rozwoju, Warszawa 2013.
3. KPMG/Prognos, *Wsparcie dla dużych przedsiębiorstw. Raport końcowy – Pakiet Roboczy 4, Ewaluacja ex-post programów Polityki Spójności 2007-2013*, Bruksela 2016.

4. Komisja Europejska, *Okres programowania 2014-2020. Dokument wytyczny w sprawie planów ewaluacji*, 2015.
5. PAG Uniconsult, *Wpływ polityki spójności na poziom i jakość zatrudnienia w Polsce*, raport zlecony przez Ministerstwo Infrastruktury i Rozwoju, Warszawa 2010.

## **Jak lepiej zrozumieć wyniki analiz kontrfaktycznych? Dobór beneficjentów do badań jakościowych, wykorzystujący dopasowanie statystyczne, na przykładzie ewaluacji wpływu Programu Operacyjnego Innowacyjna Gospodarka 2007-2013**

### **Wprowadzenie**

Celem rozdziału jest przedstawienie techniki doboru beneficjentów do badań jakościowych, opartej na statystycznym dopasowaniu (dalej nazywanej „techniką”) i pokazanie korzyści wynikających z jej zastosowania. Technika zostanie opisana na tle analiz kontrfaktycznych i w oparciu o doświadczenia związane z jej wykorzystaniem w ramach ewaluacji *ex post* wybranych działań Programu Operacyjnego Innowacyjna Gospodarka (dalej nazywanej „badaniem” lub „badaniem PARPu”). Głównym celem badania było oszacowanie jaki poziom obserwowanych zmian w przedsiębiorstwach, które zrealizowały projekty w ramach POIG, stanowi bezpośredni efekt POIG<sup>126</sup>. W ramach badania analizowano efekty działań 4.1, 4.2, 4.4 i 8.2 Programu. Szacowano zarówno efekty netto (odniesienie sytuacji beneficjenta po zakończeniu projektu do hipotetycznej sytuacji, w której nie otrzymałby wsparcia) jak i efekty względne (odniesienie sytuacji beneficjenta po zakończeniu projektu do hipotetycznej sytuacji, w której skorzystałby z alternatywnego wsparcia w ramach innego działania). Komponent jakościowy uzupełniał analizy ilościowe, a jego celem była analiza czynników sukcesu i porażki m.in. w ramach działania 4.4.

Prezentowana tu technika wpisuje się w aktualne, opisane w dalszej części rozdziału, teoretyczne i metodologiczne rozważania dotyczące pomiaru efektów interwencji. Przede wszystkim pokazuje, w jaki sposób można łączyć analizy kontrfaktyczne z podejściem ewaluacji opartej na teorii programu (*theory-based evaluation*). Ponadto, ilustruje ograniczenia związane z wykorzystaniem analiz kontrfaktycznych oraz pokazuje w jakich sposób można je przewyciężyć.

<sup>126</sup> Analiza efektów netto wybranych działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013, z wykorzystaniem podejścia counterfactual impact evaluation, w ramach projektu ewaluacyjnego PARP „Barometr Innowacyjności”, konsorcjum Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego i EGO – Evaluation for Government Organizations s.c., na zamówienie Polskiej Agencji Rozwoju Przedsiębiorczości, Kraków 2015 [http://badania.parp.gov.pl/images/badania/Raport\\_Barometr\\_netto\\_POIG.pdf](http://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf) (dostęp 24.04.2017 r.)

W dalszej części rozdziału przedstawiono kontekst wykorzystania techniki. W opisie odniesiono się zarówno do tła teoretycznego, jak i do praktyki ewaluacji *ex post*. Następnie przedstawiono samą technikę wraz z przykładem jej zastosowania w opisywanym badaniu. Następnie opisane zostały kluczowe wyniki przeprowadzonych analiz oraz wnioski z badania. Z kolei w podsumowaniu przedstawiono wnioski dotyczące wykorzystania techniki.

## Kontekst wykorzystania techniki

Analizy kontrfaktyczne na świecie stosowane są od kilku dekad, choć w ewaluacjach *ex post* w Polsce pierwsze zastosowania miały miejsce stosunkowo niedawno<sup>127</sup>. W sytuacji, gdy przeprowadzenie eksperymentu jest w praktyce niemożliwe (lub istotnie ograniczone)<sup>128</sup>, wydają się one wystarczająco dobrym przybliżeniem oceny efektów interwencji.<sup>129</sup> David Storey opisywał analizy kontrfaktyczne jako „raj ewaluatora”, do którego powinniśmy dążyć w ramach każdej ewaluacji *ex post*.<sup>130</sup>

Po zakończeniu udzielania największej jak dotąd pomocy strukturalnej w Polsce, zrealizowanej w ramach unijnej polityki spójności na lata 2007-2013, w polskiej praktyce badań ewaluacyjnych coraz częściej zamawiający ewaluację *ex post* chcą sięgać po analizy kontrfaktyczne. Rosnąca popularność metod kontrfaktycznych wynika m.in. z zaleceń Komisji Europejskiej, która stwierdza, że analizy kontrfaktyczne i ewaluacja oparta na teorii programu powinny być kluczowymi podejściami w ewaluacji *ex post*<sup>131</sup>. Istnieje już w Europie dość pokaźna liczba badań wykorzystujących metody kontrfaktyczne w ewaluacji<sup>132</sup>. Podobnie, można już mówić o pewnej tradycji badań w tym obszarze dla interwencji realizowanych w Polsce<sup>133</sup>.

---

<sup>127</sup> Pierwsze przykłady tego typu badań w Polsce, pojawiły się w 2006 r. w ewaluacjach PARP dotyczących efektów przedakcesyjnych programów Phare 2002 Spójność Społeczno-Gospodarcza (komponenty wsparcia Małych i Średnich Przedsiębiorstw i Rozwoju Zasobów Ludzkich).

<sup>128</sup> Np. ze względów etycznych (podmioty niewybrane do objęcia wsparciem mogą czuć się pokrzywdzone), merytorycznych (gdym celem interwencji jest udzielenie wsparcia konkretnym podmiotom) czy finansowych.

<sup>129</sup> P. Rosenbaum, *Observational studies*, Springer, New York 2002.

<sup>130</sup> D. Storey, „Six steps to heaven: Evaluating the impact of public policies to support small businesses in developed economies”, w: D. Sexton, H. Landstrom (red.), *The Blackwell Handbook of Entrepreneurship*, Blackwell, Oxford 2000, s. 176-193.

<sup>131</sup> Komisja Europejska, *Guidance document on monitoring and evaluation. Concepts and Recommendations*, Bruksela 2014.

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

<sup>133</sup> Por. m.in. ewaluacje PARP (np. ewaluacje *ex post* wsparcia Małych i Średnich Przedsiębiorstw i Rozwoju Zasobów Ludzkich w ramach programów Phare 2002 i 2003 SSG, badania efektów netto projektów SPOWKP 2004-2006 wspierających wzrost konkurencyjności przedsiębiorstw, analizy efektów netto wybranych działań POIG 2007-2013 skierowanych bezpośrednio do przedsiębiorstw, oceny efektów dofinansowania POKL 2007-2013 studiów podyplomowych oraz kształcenia w grupie wiekowej 50 plus czy wpływu PORPW 2007-2013 na rozwój przedsiębiorczości w Polsce Wschodniej).

Jednocześnie wśród teoretyków i metodologów badań ewaluacyjnych nadal trwa dyskusja dotycząca właściwych strategii planowania i prowadzenia ewaluacji *ex post*. Często podkreśla się w niej potrzebę szerszego spojrzenia na podejścia badawcze stosowane w ewaluacji i każdorazowe dopasowanie podejścia do pytań badawczych i dostępnych zasobów<sup>134</sup>. Trwają także poszukiwania ram koncepcyjnych, nowych podejść i technik, pozwalających na optymalne łączenie różnych paradygmatów ewaluacyjnych<sup>135</sup>. Na przykład Howard White proponuje opartą na teorii ewaluację wpływu (*theory-based impact evaluation*), która łączy zalety analiz kontrfaktycznych i ewaluacji opartej na teorii programu<sup>136</sup>.

Postulat łączenia analiz kontrfaktycznych z ewaluacją opartą na teorii wynika z faktu, że analizy kontrfaktyczne umożliwiają jedynie oszacowanie uśrednionych efektów interwencji. Nie pozwalają natomiast na prześledzenie mechanizmów, które do nich doprowadziły (nie odpowiadają na pytania o przyczyny). Innymi słowy, podczas ewaluacji opartej na teorii programu dokonuje się rozkładu interwencji na szereg hipotez dotyczących tego, jak zasoby w programie zmieniają się poprzez działania w efekty interwencji.

Ewaluacja oparta na teorii może uzupełniać analizy kontrfaktyczne. Pozwala wówczas lepiej zrozumieć oszacowane efekty programu. Wielkości efektów netto bowiem rzadko bywają zgodne z obserwowanymi zmianami (brutto) w obszarze objętym interwencją, również ich kierunki mogą być odmienne od pierwotnych oczekiwań (dodanie, ujemne), a to z kolei nastrocza wielu trudności interpretacyjnych. Może się także okazać, że nawet jeśli średni wynik interwencji wynosi zero, to pewna (nawet stosunkowo nieliczna) grupa beneficjentów osiągnęła znakomite efekty. Co więcej, efekty te mogą stanowić wystarczający argument za skutecznością interwencji. Uzupełnianie wyników analiz kontrfaktycznych o wnioski z ewaluacji opartej na teorii, pomaga także w skutecznym komunikowaniu ewaluacji programu. Po przeprowadzeniu analiz, ich właściwa interpretacja (przejście od wyników do wniosków i rekomendacji) wiąże się często z szeregiem zastrzeżeń i nie zawsze jest oczywista dla interesariuszy. Z kolei wyniki ewaluacji, które nie są zgodne z oczekiwaniami (w tym także intuicyjnymi przekonaniem) odbiorców badania, mogą skłaniać do ich negocjowania. Natomiast sformułowanie precyzyjnych hipotez dotyczących spodziewanego kierunku efektów, oraz kontekstowa interpretacja odnotowanych efektów, pomaga tworzyć narrację zrozumiałą dla odbiorców badania.

---

<sup>134</sup> 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

<sup>135</sup> Różne podejścia ewaluacyjne oparte na teorii programu zostały szczegółowo opisane w niniejszej publikacji w rozdziale pt. „Przegląd systematyczny koncepcji ewaluacji opartej na teorii” autorstwa Katarzyny Hermann-Pawłowskiej oraz Pauliny Skórskiej.

<sup>136</sup> H. White, „Theory-based impact evaluation: principles and practice”, *Journal of Development Effectiveness* 2009, vol. 3, s. 271–284.

Podsumowując, jak wynika z teorii i dotychczasowych praktyk ewaluacyjnych, analizy kontrfaktyczne powinny nadal być brane pod uwagę jako jedno z najlepszych podejść w ramach ewaluacji *ex post*. Jednocześnie także tam, gdzie ich stosowanie jest możliwe i celowe, warto rozważyć wykorzystanie innych technik, które wzbogacą rozumienie oszacowanych efektów interwencji. Ostateczny kształt projektu badawczego powinien oczywiście jednak zależeć od potrzeb informacyjnych (wyrażonych w postaci określonych pytań badawczych) i dostępnych zasobów na jego realizację. Przykład takiego podejścia w ewaluacji tj. wzbogacającego rozumienie efektów interwencji oszacowanych metodami kontrfaktycznymi oraz adekwatnego do pytań badawczych i zasobów, opisano w dalszej części niniejszego rozdziału.

Ważnym kryterium użyteczności prezentowanej tu techniki jest jej wkład w lepsze zrozumienie wyników analiz kontrfaktycznych. Dlatego też, w niniejszym rozdziale opis techniki bardzo silnie został osadzony w kontekście badania, w którym została ona wykorzystana. Oprócz komplementarności techniki względem zastosowanych w ewaluacji analiz kontrfaktycznych, odwołanie się do konkretnych doświadczeń z realizacji całego badania, pozwala na ilustrację opisanych wcześniej wyzwań, związanych w ogóle z wykorzystaniem analiz kontrfaktycznych w ewaluacji.

Opisywaną technikę przedstawiono na przykładzie jej wykorzystania w ramach interpretacji efektów działania 4.4 POIG – „Nowe inwestycje o wysokim potencjale innowacyjnym”. Celem działania było wsparcie przedsiębiorstw dokonujących nowych inwestycji (oraz niezbędnych dla ich realizacji projektów doradczych i szkoleniowych), obejmujących nabycie innowacyjnych rozwiązań technologicznych. Kluczowe zakładane efekty to wyższa innowacyjność (w tym większy potencjał do prowadzenia działalności badawczo-rozwojowej) przedsiębiorstw oraz ich wyższa konkurencyjność. Działanie zostało wytypowane do tego typu analiz m.in. ze względu na jego stosunkowo duży budżet (w tym duże oczekiwania co do uzyskanych efektów), w porównaniu do innych działań wdrażanych w POIG.

## **Zastosowana technika na tle metodologii badania**

### **Procedura szacowania efektów netto**

Ponieważ ze względu na ograniczenia prawne, urzędy statystyczne co do zasady nie udostępniają podmiotom zewnętrznym danych jednostkowych badanych przedsiębiorstw (podobnie postępuje polski GUS<sup>137</sup>), w realizowanych dotychczas kontrfaktycznych ewaluacjach wpływu wsparcia przedsiębiorstw w Polsce, wypracowano dotychczas dwie procedury wykorzystania danych ze sprawozdawczości firm do GUS<sup>138</sup>.

---

<sup>137</sup> Por. ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (Dz.U. 1995 nr 88 poz. 439 z późn. zm.).

<sup>138</sup> Dane na potrzeby analiz kontrfaktycznych można także pozyskiwać bezpośrednio od przedsiębiorstw w toku badań ankietowych. Takie rozwiązanie jest jednak zdecydowanie bardziej kosztowne i stanowi dodatkowe obciążenie dla firm.



Pierwsza została zastosowana w omawianym tu badaniu. W uproszczeniu przebiegała ona następująco:

1. Na początku do beneficjentów zostali statystycznie dopasowani nieskuteczni wnioskodawcy, na podstawie danych gromadzonych podczas aplikowania o dofinansowanie. Do każdego beneficjenta – o ile było to możliwe – starano się dobrać co najmniej dwie podobne firmy, co zwiększało szansę na uzyskanie później kompletu danych dla dopasowanej pary (beneficjenta i jednostki kontrolnej).
2. Lista beneficjentów i ich „statystycznie bliźniaczych firm”, dobranych spośród nieskutecznych wnioskodawców, została przekazana do GUS z prośbą o obliczenie średnich wartości przyjętych wskaźników efektu w obu grupach dla wskazanych lat (tj. wyznaczonych na podstawie indywidualnych danych beneficjentów, lat następujących po zakończeniu dofinansowanych przedsięwzięć oraz roku bezpośrednio je poprzedzającego; analogiczne lata kalendarzowe przyjęto dla jednostek kontrolnych).
3. GUS zidentyfikował firmy z przekazanej listy na podstawie numerów REGON i NIP, dokonał stosownych obliczeń, a następnie przekazał odpowiednie wartości średnie dla wskazanych grup przedsiębiorstw (tj. grupy beneficjentów i grupy kontrolnej), jednocześnie zachowując tajemnicę statystyczną (tj. wymogi w zakresie anonimowości i bezpieczeństwa danych jednostkowych).
4. Na koniec, w oparciu o uzyskane dane z GUS, techniką podwójnej różnicy (*difference in differences*, DID) dokonano obliczeń efektów interwencji oraz ich interpretacji<sup>139</sup>.

Ze względu na małą próbkę firm uzyskanych z GUS, na podstawie której możliwe było oszacowanie efektów oraz ograniczone możliwości kontroli podobieństwa porównywanych grup w opisanej wyżej procedurze, w niniejszym rozdziale prezentujemy oszacowania efektów netto dla działania 4.4 POIG uzyskane w toku innego badania, w którym uczestniczył jeden z autorów tego rozdziału<sup>140</sup>.

Efekty netto dla działania 4.4 POIG tu zaprezentowane, ustalono za pomocą drugiej (alternatywnej) procedury szacowania efektów netto, w ramach analiz kontrfaktycznych prowadzonych w oparciu o dane ze sprawozdawczości do GUS (badanie GUSu). Przebiegała ona następująco:

1. Na początku do GUS przekazywana jest lista beneficjentów interwencji, będących przedmiotem zainteresowania ewaluatora.
2. Na podstawie numerów REGON i NIP beneficjentów, GUS identyfikuje te przedsiębiorstwa i oznacza je w swoich bazach.

---

<sup>139</sup> W wyniku opisanej procedury, dla 99 przekazanych par beneficjentów działania 4.4 i nieskutecznych wnioskodawców, w bazie danych ze sprawozdań SP („Roczna ankieta przedsiębiorstw”), GUS udało się zidentyfikować jedynie 31 par, dla których dysponował kompletem danych.

<sup>140</sup> „Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych”, GUS, Warszawa 2015.

3. Następnie ewaluator opracowuje odpowiednie kody do analizy danych (skrypt) i przekazuje je GUS.
4. GUS, korzystając z tego samego oprogramowania do analiz statystycznych, uruchamia kod, który w pierwszym kroku dobiera pary „statystycznych bliźniaków” do beneficjentów (spośród populacji pozostałych firm, których danymi dysponuje), a w drugim kroku oblicza określone wskaźniki efektów interwencji.
5. W końcu GUS przekazuje wyniki analiz ewaluatorowi, który ocenia jakość uzyskanego dopasowania i stabilność oszacowanych efektów (procedura może być powtarzana do uzyskania satysfakcjonującego dopasowania i stabilnych wyników oszacowań).

W tym przypadku, analizy są wykonywane bezpośrednio na danych jednostkowych GUS. Tajemnica statystyczna jest nadal zachowana tj. podmiot zewnętrzny nie ma wglądu w dane jednostkowe, a jedynie planuje i zdalnie wspomaga proces prowadzenia analiz przez pracowników GUS.

### **Procedura szacowania efektów względnych**

Oprócz efektów netto, przydatne w ocenie interwencji mogą być oszacowania efektów względnych. Na przykład, pozwalają one określić skuteczność interwencji w porównaniu z inną, o podobnych celach i kierowaną do podobnej grupy docelowej. W zakresie przedmiotowym omawianego badania, względnie podobne do działania 4.4 cele miało działanie 4.2 POIG<sup>141</sup>. Do przedsiębiorstw z grupy docelowej (beneficjenci działania 4.4) zostały dobrane (dopasowane statystycznie) podobne firmy z grupy odniesienia (beneficjenci działania 4.2). W doborze porównywanych firm wykorzystano zmienne kontrolne z tych samych źródeł, co przy szacowaniu efektów netto, tj. m.in. dane z Lokalnych Systemów Informatycznych, gromadzących dane z etapu aplikowania o dofinansowanie. Efekty obliczono techniką DID dla roku ukończenia dofinansowanego projektu (Z) i dla roku następującego po roku ukończenia (Z+1). Informacje o wartościach wskaźników efektów tu zaprezentowanych pozyskano z badania ewaluacyjnego POIG – „Barometr Innowacyjności” (badanie PARPu)<sup>142</sup>.

---

<sup>141</sup> Celem działania 4.2 było wspieranie działalności badawczo-rozwojowej przedsiębiorstw, wspieranie tworzenia wzorów przemysłowych i użytkowych oraz ich wdrożeń. „Zarówno zakres możliwego finansowania (oparty przede wszystkim na ogólnym rozporządzeniu w sprawie wyłączeń blokowych (...) nr 800/2008 / GBER), grupa docelowa, która mogła korzystać ze wsparcia (przedsiębiorstwa wysokiego wzrostu – MSP jak również duże) oraz wspólne cele na poziomie osi priorytetowej były zbliżone (...)” pomiędzy działaniem 4.4 i 4.2 POIG (Por. J. Pokorski, „Budowa potencjału innowacyjnego firm w Polsce – Wyniki Programu Operacyjnego Innowacyjna Gospodarka”, w: P. Zadura-Lichota (red.), *Innowacyjna przedsiębiorczość w Polsce. Odkryty i ukryty potencjał polskiej innowacyjności*, PARP, Warszawa 2015).

<sup>142</sup> Była to ewaluacja *on-going* realizowana w PARP latach 2011-2015 wśród beneficjentów (przedsiębiorców) wybranych działań POIG (Por. 2. „Barometr Innowacyjności. Ewaluacja *on-going* działań Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013, skierowanych do przedsiębiorstw”, konsorcjum ARC Rynek i Opinia sp. z o.o. i Exacto sp. z o.o., na zamówienie PARP, Warszawa 2015 [www.badania.parp.gov.pl/barometr-innowacyjnosci](http://www.badania.parp.gov.pl/barometr-innowacyjnosci) (dostęp 24.04.2017 r.)

## Procedura doboru beneficjentów do wywiadów jakościowych

Ilościowe oszacowania efektów (netto i względnych) pozwalają ocenić czy interwencja zakończyła się sukcesem czy też nie. Jednak do zrozumienia przyczyn sukcesu lub porażki potrzebne są badania jakościowe. W tym rozdziale przedstawiamy procedurę doboru beneficjentów do wywiadów jakościowych z wykorzystaniem dopasowania statystycznego. W omawianym badaniu, wywiady pogłębione prowadzono wśród par kontrastowych beneficjentów działania 4.4 POIG, z których jeden otrzymane wsparcie przełożył na sukces rynkowy, a drugi nie. Miało to umożliwić identyfikację kluczowych czynników sukcesu i porażki analizowanej interwencji, odwołując się do określonej sytuacji danego beneficjenta. Procedura doboru kontrastowych par beneficjentów przebiegała następująco:

1. Na początku zdefiniowano wskaźnik sukcesu. Był nim zmodyfikowany<sup>143</sup> wskaźnik zwrotu ze sprzedaży (*return on sales*, ROS) obliczony jako iloraz zysku przed opodatkowaniem (*earnings before taxes*, EBT) i przychodów netto ze sprzedaży. Zwrot ze sprzedaży jest jednym z podstawowych wskaźników oceny sytuacji finansowej przedsiębiorstwa i jego konkurencyjności, stąd decyzja o jego wyborze jako wskaźnika sukcesu interwencji.
2. Sformułowano model regresji, w którym przewidywano wartość ROS w rok po zakończeniu dofinansowanej inwestycji (Z+1) za pomocą wartości ROS w roku rozpoczęcia inwestycji (R) oraz zestawu zmiennych charakteryzujących przedsiębiorstwo i szeregu szczegółowych informacji dotyczących otrzymanego dofinansowania, w tym: rok rozpoczęcia i zakończenia inwestycji, typ firmy (spółka z o.o., spółka akcyjna, pozostałe), wielkość firmy (do 49 zatrudnionych, 50-249, 250 i więcej zatrudnionych), branża głównej działalności (produkcja, pozostałe), rok założenia firmy, procent wkładu własnego w koszty inwestycji, kwota otrzymanego dofinansowania, województwo, w którym jest siedziba firmy. Dobór zmiennych kontrolnych w modelu, poza kryteriami merytorycznymi, podyktowany był kryteriami pragmatycznymi (dostępność danych) i statystycznymi (istotność zmiennych w modelu). Rezultaty diagnostyki modelu okazały się satysfakcjonujące, dzięki czemu uzyskano pewność co do prawidłowości oszacowania reszt. Zastosowanie metody regresji dało możliwość kontroli potencjału firmy w modelu, przez co wartość wskaźnika sukcesu była bardziej obiektywna. Reszty z modelu regresji stanowią indywidualny wskaźnik sukcesu firmy, dzięki temu dokładniejszy niż np. odniesienie sytuacji firmy do średniej sytuacji innych firm.
3. W następnym kroku zapisano reszty z modelu regresji, tj. różnice między rzeczywistym a przewidywanym poziomem ROS w rok po zakończeniu dofinansowanej inwestycji (Z+1). W tym przypadku reszty mogą być interpretowane jako miara suk-

<sup>143</sup> Modyfikacja wskaźnika polegała na zastąpieniu zysku netto, który standardowo znajduje się w liczniku wskaźnika, zyskiem brutto (czyli zyskiem przed odprowadzeniem podatku dochodowego). Było to podyktowane dostępnością danych. W ramach „Barometru Innowacyjności” bowiem, badano zysk brutto a nie zysk netto, co wynikało z chęci identyfikacji wpływu interwencji na zysk bez uwzględnienia zmian w zakresie opodatkowania.

cesu – przedsiębiorstwo otrzymane wsparcie przekuło na sukces rynkowy (osiągnęło ROS wyższe niż przewidywane w modelu w oparciu o potencjał przedsiębiorstwa „na starcie”) lub nie (osiągnęło ROS poniżej oczekiwań). Ta logika interpretacji reszt z regresji została zaczerpnięta z modelowania edukacyjnej wartości dodanej (EWD), wykorzystywanego do informowania o efektywności szkół i nauczycieli<sup>144</sup>. W Polsce metodę EWD dla szkół rozwijano w latach 2005-15 w ramach projektu Instytutu Badań Edukacyjnych<sup>145</sup>.

4. Beneficjenci działania 4.4 zostali uszeregowani wg wielkości reszty. W ten sposób ukonstytuowano dwie grupy: firmy sukcesu (powyżej trzeciego kwartyła) i firmy, którym nie udało się przekuć otrzymanego dofinansowania we wzrost ROS (poniżej mediany). Przedsiębiorstwa powyżej trzeciego kwartyła i poniżej mediany pozostawiono do dalszych analiz (co podyktowane było chęcią utrzymania proporcji 1:2 w przygotowanej próbie do badań jakościowych, tj. jednej firmy z historią sukcesu do dwóch firm z historią porażki), pozostałe usunięto z roboczego zbioru danych.
5. W kolejnym kroku przeprowadzono procedurę dopasowania statystycznego polegającą na znalezieniu (dopasowaniu) firm podobnych do siebie („statystycznych bliźniaków”) a różniących się jedynie faktem zaklasyfikowania do grupy firm sukcesu i firm, które poniosły porażkę. Zasadniczo, w dopasowaniu statystycznym można wykorzystać zestaw tych samych kowariantów co w modelu regresji, z którego pozyskano resztę. W tym przypadku zastosowano dopasowanie oparte na odległości Mahalanobisa (*Mahalanobis distance matching*, MDM) między firmami wg wartości ROS i kwoty uzyskanego dofinansowania (zmiennie ciągłe rekodowano na 5 kategorii wg kwintyli) oraz jednocześnie dopasowanie dokładne wg zmiennych: rok rozpoczęcia i zakończenia inwestycji, wielkość firmy i branży głównej działalności.
6. Na koniec, dla każdej „firmy sukcesu” w miarę możliwości dobrano dwie firmy (bliźniacze pod względem cech statystycznych), które poniosły porażkę. Tak więc w wyniku dopasowania uzyskano nie tylko pary, ale także triplety firm, dzięki czemu jeśli dopasowana do „firmy sukcesu” pierwsza „firma porażki” odmówiła udzielenia wywiadu, można było zwrócić się do tej drugiej.

<sup>144</sup> R. Chetty, J. N. Friedmann, J. E. Rockoff, „Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates”, *American Economic Review* 2014, vol. 104, s. 2593–2632.

C. T. Clotterfeld, H. F. Ladd, „Recognizing and Rewarding Success in Public Schools”, w: H. F. Ladd (red.), *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, s. 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, s. 615–631.

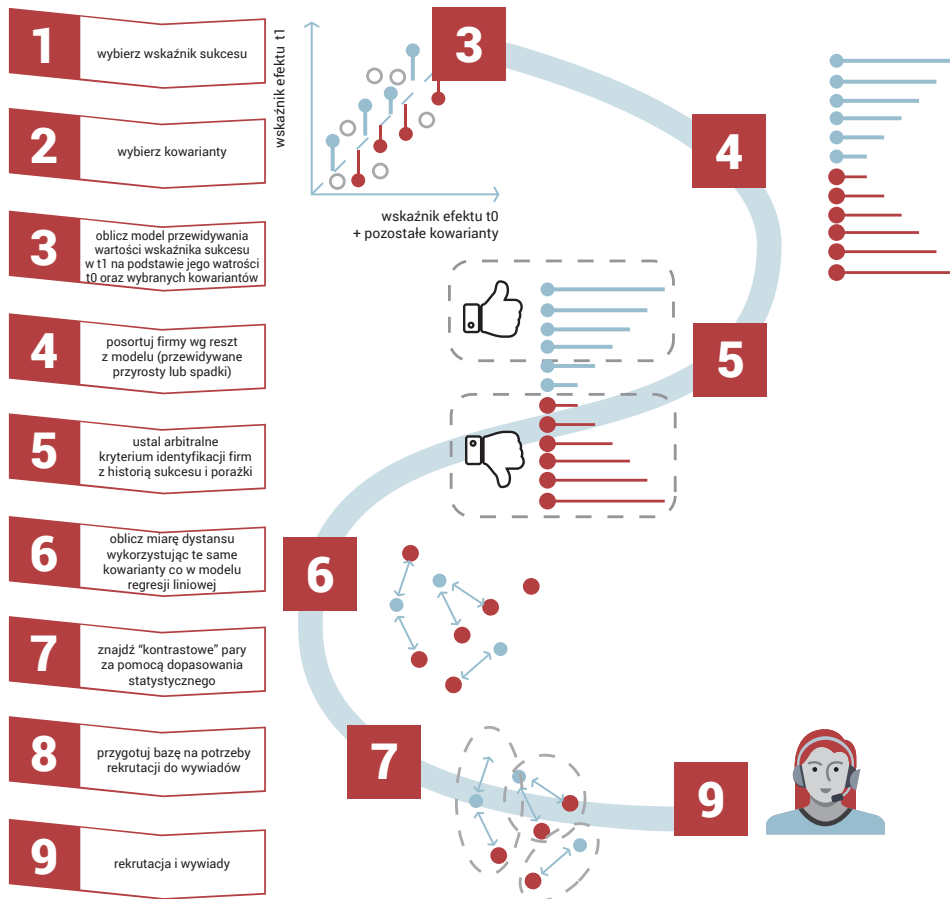
H. F. Ladd, R. P. Walsh, „Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right”, *Economics of Education Review* 2002, vol. 21, s. 1–17.

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

<sup>145</sup> Por. IBE – projekt pt. „Rozwój metody edukacyjnej wartości dodanej na potrzeby wzmocnienia ewaluacyjnej funkcji egzaminów zewnętrznych” [www.ewd.edu.pl](http://www.ewd.edu.pl) (dostęp 24.04.2017 r.).

Wyłonienie beneficjentów działania 4.4 POIG do wywiadów jakościowych, zgodnie z opisaną wyżej procedurą, w stosunkowo niewielu wywiadach pozwoliło naświetlić (tj. wyostrzyć, uwzględniając odpowiedni kontrast) czynniki sukcesu i porażki interwencji, co istotnie obniżyło koszty badania (ograniczono skalę badań pogłębionych na przedsiębiorstwach, konieczną w innych przypadkach – projektach badawczych, pozbawionych tego typu precyzyjnego doboru). Ostatecznie zrealizowano 6 wywiadów pogłębionych (*individual in-depth interviews*, IDI): 3 z firmami sukcesu i 3 z firmami, które poniosły porażkę. Wywiady przeprowadzono z przedstawicielami firm, którzy byli najlepiej poinformowani w zakresie potencjalnego wpływu uzyskanego dofinansowania na konkurencyjność firmy. Najczęściej byli to członkowie zarządu lub osoby bezpośrednio odpowiadające za realizację projektu w przedsiębiorstwie. Rysunek 1. graficznie prezentuje przebieg niniejszej procedury.

## Rysunek 1. Schemat procedury doboru beneficjentów do wywiadów jakościowych z wykorzystaniem dopasowania statystycznego



Źródło: opracowanie własne.

### Dopasowanie statystyczne

W omawianym badaniu dopasowanie statystyczne (*statistical matching*) zastosowane zostało do doboru próby na potrzeby szacowania efektów netto oraz względnych, jak i doboru beneficjentów do wywiadów jakościowych. Ponieważ, dopasowanie statystyczne zostało wykorzystane w każdym z trzech kluczowych elementów badania tu opisanego, metoda ta, a zwłaszcza jej najpopularniejsza technika, tj. *propensity score matching* (PSM), wymaga przybliżenia.

Dopasowanie statystyczne jest grupą technik analitycznych, które można uznać za dobrą praktykę stosowaną w badaniach ewaluacyjnych. „Jest oparte na intuicyjnej

i atrakcyjnej idei kontrastowania efektów osiągniętych przez beneficjentów programu i porównywalnych niebeneficjentów. Obserwowane różnice w poziomie wskaźników efektów między obiema grupami uznaje się za efekt programu<sup>146</sup>. Dopasowanie statystyczne próbuje tym samym przez statystyczną kontrolę obserwowanych zmiennych naśladować warunki randomizowanego eksperymentu, który uważany jest za „złoty standard” w dowodzeniu przyczynowości<sup>147</sup>.

PSM jest najczęściej stosowaną techniką dopasowania statystycznego<sup>148</sup>. Zainteresowanych czytelników odsyłamy do publikacji wprowadzających w tematykę różnych technik opartych na wektorze *propensity score* (PS) jak również dyskusji nad różnicami między tymi technikami a wykorzystaniem regresji wielorakiej do szacowania efektów interwencji<sup>149</sup>, publikacji konsumujących podstawy teoretyczne pierwszych opracowań na temat technik dopasowania statystycznego oraz przykładów ich zastosowania w praktyce badawczej różnych dyscyplin<sup>150</sup>. Dostępne są także praktyczne przewodniki, jak stosować PSM<sup>151</sup>. Godne polecenia są również polskojęzyczne prace<sup>152</sup>. Warto odnotować także głosy krytyki technik opartych na wektorze PS, które powinny wzmagać czujność podczas ich stosowania i interpretacji wywiedzionych z nich rezultatów. Na przykład Elizabeth Wilde i Robinson Hollister wykazali, używając danych z dużego eksperymentu edukacyjnego, że techniki oparte na PS nie dość skutecznie odtwarzają warunki eksperymentalne<sup>153</sup>. Dodatkowo niektórzy autorzy (np. Gary King i Richard Nielsen<sup>154</sup>) ostrzegają, że nieostrożne stosowanie PSM może powodować wzrost obciążenia szacowanych efektów i zwiększać niestabilność modelu. Proponują oni uważny dobór zmiennych kontrolnych i dokładne dopasowanie na wszystkich lub przynajmniej kluczowych z nich. Sugestie te zostały wykorzystane w przeprowadzonych przez nas analizach w ramach omawianego badania.

---

<sup>146</sup>J. J. Heckman, H. Ichimura, P. Todd, "Matching As An Econometric Evaluation Estimator", *Review of Economic Studies* 1998, vol. 65, s. 261.

<sup>147</sup>T. Dunning, "Improving Causal Inference: Strengths and Limitations of Natural Experiments", *Political Research Quarterly* 2008, vol. 61, s. 282–293.

<sup>148</sup>P. Rosenbaum, D. Rubin, "The central role of the propensity score in observational studies for causal effects", *Biometrika* 1983, vol. 70, s. 41–55.

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

<sup>149</sup>P. C. Austin, "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies", *Multivariate Behavioral Research* 2011, vol. 46, s. 399–424.

<sup>150</sup>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, s. 1–21.

<sup>151</sup>M. Caliendo, S. Kopeinig, "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys* 2008, vol. 22, s. 31–72.

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

<sup>153</sup>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, s.455–477.

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

## Wyniki badań

### Analizy kontrfaktyczne

Oszacowania efektów dla beneficjentów działania 4.4 POIG dostarczyły wyników, które okazały się sprzeczne z wyprowadzonymi z logiki interwencji efektami, pierwotnie oczekiwanymi przez zespół ewaluatorów. Otóż analizowani beneficjenci w przypadku większości wskaźników nie osiągnęli wyników lepszych niż nieskuteczni wnioskodawcy, którzy aplikowali o wsparcie w ramach tego działania (efekty netto) czy też beneficjenci działania 4.2 POIG<sup>155</sup> (efekty względne). Do wyjątków należy wpływ interwencji na prowadzenie działalności innowacyjnej i eksportowej. Efekt netto dla działania 4.4 był dodatni dla wskaźników działalności innowacyjnej<sup>156</sup> i eksportowej<sup>157</sup> w odniesieniu do nieskutecznych wnioskodawców. Najbardziej zaskakiwał fakt, iż beneficjenci działania 4.4 znacznie częściej wykazywali stratę niż beneficjenci działania 4.2 (zarówno w roku zakończenia inwestycji jak i w kolejnym). Ze względu na dużo wyższą wartość wsparcia w przypadku działania 4.4 należało bowiem oczekiwać odwrotnego efektu.

**Tabela 2. Istotne statystycznie efekty netto i względne dla działania 4.4 POIG**

	Badanie GUSu		Badanie PARDu			
	Efekt netto 2008-13	Odchylenie standardowe	Efekt względny do 4.2 (t0-t1 <sup>a</sup> )	Odchylenie standardowe	Efekt względny do 4.2 (t0-t2 <sup>b</sup> )	Odchylenie standardowe
<b>Przedsiębiorstwa niewykazujące straty (%)</b>	nd	nd	-26**	13	-33**	13
<b>Przedsiębiorstwa eksportujące (%)</b>	11*	6	-1	3	0	0
<b>Przedsiębiorstwa innowacyjne (%)</b>	20**	8	nd	nd	nd	nd

Nota. nd – nie dotyczy, \* istotność statystyczna na poziomie  $p < 0,01$ ; \*\*  $p < 0,05$ ; <sup>a</sup> rok zakończenia inwestycji w odniesieniu do roku jej rozpoczęcia, <sup>b</sup> rok po zakończeniu inwestycji w odniesieniu do roku jej rozpoczęcia, wartość dodatnia oznacza, że wartość wskaźnika

<sup>155</sup> Wartość wsparcia w ramach działania 4.2 była średnio czterokrotnie mniejsza niż w przypadku 4.4. Należało więc w jego przypadku oczekiwać mniejszych efektów.

<sup>156</sup> Na potrzeby oznaczenia podmiotów prowadzących działalność innowacyjną, przyjęto w ramach badania GUS z 2015 r., że za podmiot prowadzący działalność innowacyjną będzie uznana każda firma, dla której suma kosztów zakończonych prac rozwojowych i wartości praw do wynalazków, patentów i znaków towarowych jest większa od zera.

<sup>157</sup> Prowadzenie działalności eksportowej identyfikowano w badaniu GUS z 2015 r., za pomocą wykazywania przychodów ze sprzedaży na eksport.



u beneficjentów 4.4 jest większa od nieskutecznych wnioskodawców (efekt netto) lub beneficjentów działania 4.2 (efekt względny), wartość ujemna sygnalizuje odwrotną sytuację.

Ewaluator dysponujący takimi wynikami staje przed nie lada wyzwaniem. Powinien bowiem być w stanie je zinterpretować i w oparciu o nie przedstawić rekomendacje dotyczące przyszłych interwencji tego typu. W przypadku omawianego badania, pomocne okazały się wyniki wywiadów jakościowych.

### **Wnioski z wywiadów z beneficjentami**

W ramach jakościowej części badania zidentyfikowano szereg czynników mających potencjalnie wpływ na przełożenie się wyników inwestycji wspieranych w ramach działania 4.4 na zmiany w rentowności przedsiębiorstw-beneficjentów (wartość wskaźnika ROS). Zidentyfikowane czynniki związane są z cechami beneficjentów, specyfiką branży, sposobem wdrożenia inwestycji, a także otoczeniem społeczno-gospodarczego (w tym koniunkturą gospodarczą). Cechy beneficjentów odnoszą się do kompetencji pracowników, szczególnie kadry zarządzającej oraz strategii, którą faktycznie realizuje przedsiębiorstwo. Wyższe kompetencje i spójna strategia, kładąca nacisk na wprowadzenie na rynek innowacyjnych produktów oraz zakładająca przeprowadzanie komplementarnych inwestycji rozwojowych, były czynnikami sprzyjającymi przełożeniu się dotacji POIG na wyższą rentowność przedsiębiorstwa.

Duże znaczenie dla przełożenia efektów inwestycji na sukces firmy miała także specyfika branży, która wpływała na strategię przedsiębiorstwa. W niektórych branżach wprowadzanie na rynek nowych produktów jest bardziej istotne dla konkurencyjności przedsiębiorstwa niż w innych. Niektórzy beneficjenci zwracali uwagę na dominującą rolę innych czynników: wielkości mocy produkcyjnych, ceny i jakości oferowanych dóbr i usług. W określonych branżach, w przypadku mniejszego znaczenia innowacyjności dla konkurencyjności firmy, beneficjenci są bardziej skłonni traktować wspierane w ramach działania 4.4 projekty w kategoriach czysto inwestycyjnych, a nie służących wzrostowi innowacyjności.

W przypadku sposobu wdrożenia nowego produktu na rynek, na uwagę zasługuje odpowiednie zarządzanie całym przedsięwzięciem, wspieranym poprzez udzielaną w ramach działania 4.4 dotację, a także odpowiednie zaplanowanie wprowadzenia wyrobu lub usługi na rynek. Chodzi tutaj o adekwatne działania z zakresu *public relations* i marketingu oraz wybranie adekwatnego sposobu i kanałów dystrybucji nowych wyrobów czy usług. Niektórzy beneficjenci działania 4.4 to firmy produkcyjne, działające w segmencie *business-to-business* (B2B). Wprowadzenie nowego produktu, zwłaszcza dla nowej grupy odbiorców, wymaga odmiennych kanałów komunikacji i dystrybucji. W przypadku części analizowanych przedsięwzięć można na podstawie wypowiedzi respondentów wyciągnąć wniosek, że nierzadko zabrakło odpowiedniej konsekwencji w ich wdrażaniu. Opracowanie nowego produktu nie pociągało za sobą nowego, adekwatnego dla niego sposobu wprowadzenia go na rynek. Na koniec, warto zwrócić

uwagę na zmiany w otoczeniu społeczno-gospodarczym, które wpływały na popyt na innowacyjne produkty. Oprócz ogólnego pogorszenia koniunktury gospodarczej w okresie finalizowania dofinansowanych projektów działania 4.4 istotne znaczenie miała sytuacja konkretnych branż. Pogorszenie się sytuacji w danej branży nieraz miało duży wpływ na sytuację ekonomiczną przedsiębiorstwa i ostateczne efekty projektu.

Jak widać, istnieje szereg czynników, które sprawiły, że część beneficjentów działania 4.4 POIG odnotowała relatywnie niewielkie przełożenie wspartych w ramach działania inwestycji na rentowność przedsiębiorstwa. Najważniejsze z nich, na które beneficjenci mieli wpływ, to brak odpowiedniego monitorowania sytuacji na rynku w kontekście wielkości popytu na nowy produkt i zaplanowania właściwego dotarcia do klienta, oraz brak spójnej i konsekwentnie wdrażanej strategii przedsiębiorstwa, w ramach której duża waga przywiązywana byłaby do wprowadzania na rynek nowych wyrobów i usług.

Ważnym wnioskiem z wywiadów była także jednoznaczna zgodność przyporządkowania beneficjentów do jednej z kategorii (sukcesu lub porażki) z ich subiektywnym postrzeganiem wyniku wykorzystania dotacji. Beneficjenci przyporządkowani w oparciu o dane ilościowe do kategorii sukcesu, w trakcie wywiadów dość wysoko oceniali sposób w jaki wykorzystali uzyskane wsparcie dla rozwoju przedsiębiorstwa. Analogicznie, firmy przypisane w oparciu o dane ilościowe do kategorii porażki, oceniali w trakcie wywiadów, że dotacja mogłaby być przez nich lepiej wykorzystana. Zgodność ta potwierdziła zarówno wybór ROS jako wskaźnika sukcesu, jak i trafność samego podejścia (techniki).

## Podsumowanie

Przedstawiona w niniejszym rozdziale technika doboru beneficjentów do badań jakościowych, wykorzystująca dopasowanie statystyczne, okazała się bardzo użyteczna dla lepszego zrozumienia uzyskanych ilościowych oszacowań efektów interwencji (netto i względnych). Pozwoliła bowiem zidentyfikować możliwe przyczyny sukcesu lub porażki analizowanej interwencji. Posłużyła jej lepszemu zrozumieniu, a w rezultacie jest pomocna w udoskonalaniu teorii programów podobnego typu w przyszłości.

Należy jednocześnie poczynić kilka zastrzeżeń. Zrealizowana liczba wywiadów była efektywna, jednak stosunkowo niewielka. W kolejnych zastosowaniach techniki warto rozważyć zwiększenie liczby wywiadów. Dzięki temu możliwe byłoby zastosowanie dodatkowo komparatywnej analizy jakościowej (*qualitative comparative analysis*, QCA) jako techniki analizy danych<sup>158</sup>. Warto także pamiętać o tym, że za taki a nie

---

<sup>158</sup> Zastosowanie takiej techniki zaplanowano w ramach ewaluacji Programu pomocy publicznej Narodowego Centrum Badań i Rozwoju, realizowanej przez konsorcjum Fundacja Idea Rozwoju, IMAPP sp. z o.o. PAG Uniconsult sp. z o.o. i Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego. Technika QCA opisana została w niniejszej publikacji w rozdziale pt. „Porównawcza analiza jakościowa w ewaluacji wpływu opartej na teorii”, autorstwa Barbary Worek oraz Marcina Kocóra.

inny wynik analiz kontrfaktycznych mogą być odpowiedzialne także inne czynniki nie-uwzględnione w modelu, np. stosunkowo krótki czas pomiaru po uzyskaniu wsparcia (możliwe jest wystąpienie efektów odroczonego w czasie). Uzyskane dzięki omawianej technice wyniki, nie dostarczają więc jednoznacznych rozstrzygnięć odnośnie do uzyskanych wyników analiz kontrfaktycznych, pozwalają jednak je lepiej zrozumieć, a przez to także wzmocnić wynikające z nich wnioski. W tym sensie wykorzystanie techniki a szerzej, komponentu jakościowego, należy traktować jako uzupełnienie a nie alternatywę dla analiz kontrfaktycznych. Ważną cechą omawianej w niniejszym rozdziale techniki jest jej mała kosztocłonność. Warto jednak pamiętać, że realizacja wywiadów z przedstawicielami kontrastowych par beneficjentów oznacza potrzebę uzyskania zgody od obydwu, co z kolei wymaga zaprojektowania niekiedy stosunkowo wyszukanych technik motywacji (w tym przypadku zastosowano m.in. odwołania do naukowego charakteru projektu badawczego i jego społecznej użyteczności, oraz unikalności respondenta w procedurze doboru)<sup>159</sup>. Ponadto, warto podkreślić, że opisywana technika powinna być stosowana tylko w sytuacji, w której chcemy odpowiedzieć na adekwatne dla niej pytania badawcze, tj. odnoszące się do przyczyn różnic w efektach pomiędzy beneficjentami. W przypadku innych pytań badawczych, należy rozważyć inne sposoby łączenia analiz kontrfaktycznych z badaniami jakościowymi lub studiami przypadków.

Omawiane podejście, umożliwiające lepsze zrozumienie wyników analiz kontrfaktycznych, zostało w nieco zmodyfikowanej postaci wykorzystane również w innych ewaluacjach *ex post* w Polsce – przykładowo na potrzeby interpretacji wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 na przedsiębiorczość na poziomie lokalnym<sup>160</sup>. Dowodzi to nie tylko użyteczności samego podejścia, ale także jego elastyczności względem przedmiotu badania i pytań badawczych.

---

<sup>159</sup> Wywiad przeprowadzony tylko z jednym z beneficjentów nie jest wystarczający do wykorzystania w późniejszych analizach.

<sup>160</sup> Por. „Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej”, konsorcjum EVALU sp. z o.o., WYG PSDB sp. z o.o. i EGO – Evaluation for Government Organizations s.c., na zamówienie PARP, Warszawa 2016.

## O autorach

### Maciej Koniewski

---

Specjalista w zakresie metodologii badań i ilościowych metod analizy danych. Pracę zawodową rozpoczął w branży badań rynkowych, realizując ponad 50 projektów krajowych i międzynarodowych (m.in. dla centrali John Deere i Deutsche Telekom). Od 2011 współpracuje z Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego w Krakowie (m.in. przy projektach ewaluacji polskiego Systemu Ewaluacji Oświaty, wsparcia budowy systemu ewaluacji szkolnictwa w Arabii Saudyjskiej, kontrfaktycznej ewaluacji wpływu działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013 (POIG) dla Polskiej Agencji Rozwoju Przedsiębiorczości). Był także członkiem zespołu przygotowującego ekspertyzy dla Głównego Urzędu Statystycznego z zakresu wykorzystania metod analizy stanów kontrfaktycznych do oceny efektywności POIG i Regionalnych Programów Operacyjnych. W latach 2013-2015 pracował w zespole Edukacyjnej Wartości Dodanej w Instytucje Badań Edukacyjnych, gdzie zajmował się m.in. analizą i raportowaniem wyników bieżących badań nad trafnością wskaźników EWD, szkoleniami oraz popularyzowaniem metody. Autor i współautor kilkunastu publikacji z zakresu badań edukacyjnych i pomiaru edukacyjnego.

### Seweryn Krupnik

---

Kierownik Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego w Krakowie. Doktor socjologii (praca doktorska poświęcona ewaluacji dotacji inwestycyjnych dla przedsiębiorstw). Certyfikowany kierownik projektów i trener. Coach. Członek komisji rewizyjnej Polskiego Towarzystwa Ewaluacyjnego. Zainteresowania naukowe związane z politykami publicznymi ukierunkowanymi na wsparcie przedsiębiorstw (zwłaszcza wsparcie bezpośrednie, przedsiębiorstwa społeczne, transfer wiedzy z uczelni do przedsiębiorstw), wykorzystaniem ewaluacji opartej na teorii programu, analizą relacji przyczynowo-skutkowych oraz wykorzystaniem teorii instytucjonalnych w rozumieniu rzeczywistości. Od kilkunastu lat łączy działalność naukową z realizacją projektów z zakresu analiz polityk publicznych.

## Bibliografia:

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", w: Ladd H. F. (red.), *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. Ehler 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](http://www.polmeth.wustl.edu/files/polmeth/psnot4.pdf) (dostęp 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*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2011.
14. Pokorski J. "Building the innovation potential of enterprises in Poland – the outcomes of the Innovative Economy Operational Programme", w: P. Zadura-Lichota (red.), *Innovative entrepreneurship in Poland. Revealed and hidden potential for innovation in Poland*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 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", w: Sexton D., Landstrom H. (red.), *The Blackwell Handbook of Entrepreneurship*, Blackwell, Oxford 2000.

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

### Inne:

1. *Barometr Innowacyjności. Ewaluacja on-going działań Programu Operacyjnego Innowacyjna Gospodarka, na lata 2007-2013, skierowanych do przedsiębiorstw*, konsorcjum ARC Rynek i Opinia sp. z o.o. oraz Exacto sp. z o.o., na zlecenie PARP, Warszawa 2015.  
[http://badania.parp.gov.pl/images/badania/Raport\\_Barometr\\_Innowacyjnosci.pdf](http://badania.parp.gov.pl/images/badania/Raport_Barometr_Innowacyjnosci.pdf) (dostęp 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"*, konsorcjum Centrum Ewaluacji i Analiza Polityk Publicznych Uniwersytetu Jagiellońskiego i EGO – Evaluation for Government Organizations s.c., na zamówienie PARP, Kraków 2015  
[http://badania.parp.gov.pl/images/badania/Raport\\_Barometr\\_netto\\_POIG.pdf](http://badania.parp.gov.pl/images/badania/Raport_Barometr_netto_POIG.pdf) (dostęp 24.04.2017)
3. *Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej*, konsorcjum EVALU sp. z o.o., WYG PSDB sp. z o.o., and EGO s.c., na zamówienie PARP, Warszawa 2016.
4. Komisja Europejska, *Guidance document on monitoring and evaluation. Concepts and Recommendations*, 2014.
5. GUS, *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych*, Warszawa 2015  
[www.stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/5807/1/1/raport\\_rpo\\_poig.pdf](http://www.stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/5807/1/1/raport_rpo_poig.pdf) (dostęp 24.04.2017).
6. Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (Dz.U. 1995 nr 88 poz. 439 z późn. zm.).

## **Stratified Propensity Score Matching – quasi-eksperymentalny schemat badawczy, wykorzystany na potrzeby ewaluacji wpływu opartej na teorii, w sytuacji braku zmiennej zależnej**

### **Wprowadzenie**

Pomimo mnogości badań wpisujących się w logikę ewaluacji wpływu, prowadzonych w zakresie skuteczności funduszy przekazywanych Polsce w ramach polityki Unii Europejskiej, większość z nich nie dostarcza skwantyfikowanych dowodów na to, jak fundusze wpływają na rozwój gospodarczy na poziomie lokalnym. Określenie efektu netto absorpcji funduszy na poziomie gmin i powiatów jest bowiem zadaniem problematycznym ze względów metodycznych. Po pierwsze, na poziomie gospodarek lokalnych ograniczona jest możliwość zastosowania modeli makroekonomicznych, które są często wykorzystywane w szacowaniu wpływu funduszy unijnych na poziomie całego kraju lub poszczególnych województw. Po drugie, fakt, że zasadniczo wszystkie polskie gminy w różnych formach korzystały z funduszy, wyklucza także zastosowanie podejścia quasi-eksperymentalnego opierającego się na metodzie *Propensity Score Matching* (PSM). Wykorzystanie regresji logistycznej, na której bazuje PSM, nie jest bowiem możliwe w sytuacji, gdy każda jednostka – w mniejszym bądź większym stopniu – poddana została interwencji, której efekty chcemy oszacować.

Celem niniejszego artykułu jest przedstawienie założeń i przykładów wykorzystania nowatorskiej metody *Stratified Propensity Score Matching* (SPSM) dającej możliwość oszacowania wpływu netto funduszy unijnych na rozwój gospodarczy gmin i powiatów. W odróżnieniu od szacowania efektów względnych (*relative causal effects*, RCE) w dalszym ciągu mamy tu do czynienia z rzeczywistym szacowaniem efektów netto. Szacowanie RCE wymaga bowiem istnienia przynajmniej dwóch grup funduszy lub interwencji, które następnie porównywane są za sobą, podczas gdy SPSM ma możliwość szacowania efektów w przypadku pojedynczych funduszy lub interwencji.

Artykuł składa się z kilku części. W pierwszej zaprezentowane zostaną podejścia badawcze stosowane w ramach ewaluacji wpływu, których wykorzystanie dostarcza ilościowych danych na temat efektu netto projektów realizowanych przy wsparciu funduszy unijnych. Omówione zostaną ograniczenia wykorzystania tych podejść w bada-

niach rozwoju gospodarczego na poziomie lokalnym. W dalszej części przedstawiona zostanie metoda SPSM, która odpowiada na wyzwania metodyczne, przed jakimi staje ewaluacja wpływu funduszy unijnych na rozwój jednostek terytorialnych na poziomie powiatów i gmin. Ostatnia część zawiera przykłady ewaluacji typu *ex post*, w których z sukcesem zaadaptowano SPSM na potrzeby określenia wpływu netto funduszy na rozwój gospodarczy polskich gmin. W obu badaniach wnioski z analiz kontrfaktycznych uzupełnione zostały pogłębionymi badaniami jakościowymi (częściowo wykorzystującymi zidentyfikowane pary „bliźniaczych” jednostek terytorialnych), które pozwoliły na wyjaśnienie zaobserwowanych różnic w oddziaływaniu zaabsorbowanych funduszy. Zastosowanie metody SPSM w połączeniu z metodami jakościowymi dało więc możliwość kompleksowego ujęcia kwestii skuteczności interwencji wdrażanej na poziomie lokalnym.

## **Ewaluacja wpływu – plany badawcze wykorzystywane w szacowaniu ekonomicznych efektów funduszy unijnych**

W ramach ewaluacji wpływu (*impact evaluation*), która co do zasady ma na celu określenie zakresu oddziaływania interwencji publicznych, dostarczenie skwantyfikowanych dowodów możliwe jest jedynie przy zastosowaniu dwóch planów badawczych (*research designs*): studium statystycznego (inaczej: szacowania strukturalnego) oraz schematu eksperymentalnego i quasi-eksperymentalnego<sup>161</sup>. Potwierdzone zostało to w zaleceniach dotyczących metod badania wpływu pomocy publicznej, w których Komisja Europejska plany te określa jako dające „możliwość określenia przyczynowego wpływu samego programu, niezakłócanego przez inne zmienne, które mogły mieć wpływ na zaobserwowane wyniki, np. ogólne warunki makroekonomiczne”<sup>162</sup>.

Logika pierwszego z planów polega na ocenie efektów przez wykorzystanie zaawansowanych metod statystycznych do określenia relacji (związków i zależności) pomiędzy interwencją a oczekiwanymi lub zaistniałymi zmianami. W szacowaniu efektów interwencji publicznych w ramach tego podejścia najczęściej wykorzystuje się metody analizy statystycznej, w tym przede wszystkim modelowanie ekonometryczne i analizy regresji. Od początku lat 90. opracowano kilka modeli makroekonomicznych (HERMIN, QUEST, EcoMod, GIMF, MaMoR, EUImpact czy RHOMOLO) pozwalających na ocenę wpływu funduszy unijnych na zmiany PKB, poziomu inwestycji, zatrudnienia i bezrobocia czy wartości dodanej produkcji. Istnieje bogata baza badań wykorzystujących modele makroekonomiczne, w szczególności w ramach oceny

<sup>161</sup> Por.: 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 oraz M. Bamberger, J. Vaessen, E. Raimondo, *Dealing With Complexity in Development Evaluation: A Practical Approach*, 1st Edition, SAGE Publications 2015.

<sup>162</sup> Komisja Europejska, *Wspólne metody oceny pomocy państwa, Dokument Roboczy Służb Komisji*, Bruksela 2014, s. 7, [http://ec.europa.eu/competition/state\\_aid/modernisation/state\\_aid\\_evaluation\\_methodology\\_pl.pdf](http://ec.europa.eu/competition/state_aid/modernisation/state_aid_evaluation_methodology_pl.pdf) (dostęp 25.05.2017 r.)



ex ante programów operacyjnych współfinansowanych z funduszy unijnych w ramach polityki spójności oraz wspólnej polityki rolnej<sup>163</sup>. Warto nadmienić, że szacowanie strukturalne jest praktycznie jedynym podejściem stosowanym w badaniach typu ex ante wpływu efektywności funduszy unijnych na rozwój polskich regionów<sup>164</sup>.

W modelowaniu uwzględnione są zazwyczaj szerokie, zarówno popytowe, jak i podażowe efekty absorpcji funduszy przez gospodarki na poziomie krajowym lub regionalnym. W ich ramach porównuje się wyniki, jakie osiągać będzie gospodarka w scenariuszu zakładającym brak dopływu zewnętrznych środków, z wynikami scenariusza je uwzględniającego<sup>165</sup>. Modele najczęściej bazują na koncepcji równowagi będącej odzwierciedleniem funkcjonowania całej gospodarki (modele równowagi ogólnej) lub konkretnego sektora (modele równowagi cząstkowej). Biorąc pod uwagę aspekt czasowości ujęcia modelowanych procesów, modele przyjmują formy statyczne lub dynamiczne. Od lat 80. ubiegłego wieku zainteresowaniem cieszą się także modele dynamicznej stochastycznej równowagi ogólnej (*Dynamic Stochastic General Equilibrium*, DSGE), wykorzystywane do prognozowania i analizy polityk publicznych w wielu krajach, np. przez banki centralne, organizacje międzynarodowe, a także przez Komisję Europejską. Wyjątkowość modeli DSGE polega na tym, że kompleksowo ujmują one zdarzenia gospodarcze, uwzględniając przy tym ich dynamikę i stochastyczny charakter<sup>166</sup>.

---

<sup>163</sup> Początkowo w ewaluacjach ex ante badających potencjalny wpływ funduszy unijnych na gospodarki państw-beneficjentów polityki spójności i wspólnej polityki rolnej UE wykorzystywany był model HERMIN, stworzony przez J. Bradleya z Instytutu Badań Społeczno-Ekonomicznych i zaimplementowany po raz pierwszy w Irlandii w 1989 r. W kolejnych latach w ewaluacjach typu ex ante zaczęto wykorzystywać „krajowe” modele, np. na Węgrzech EcoRET (H. Schalk i A. Vargi), a na Litwie LAFUN (Baltic International Centre for Economic Policy Studies).

<sup>164</sup> Dla przykładu: ewaluacje ex ante wszystkich programów operacyjnych wdrażanych w Polsce w latach 2007-2013 zawierały szacunki dotyczące wpływu funduszy unijnych na polską gospodarkę wyliczone przy użyciu modelu HERMIN.

<sup>165</sup> Pomimo dość powszechnego zastosowania tego typu modeli do przewidywania skutków wdrażania instrumentów polityk publicznych nie brakuje głosów krytycznych, wskazujących na problem prowadzenia niezmiernie skomplikowanej rzeczywistości gospodarczej do matematycznych równań, mających ją odzwierciedlać. W 2010 r. dyskusję dotyczącą ograniczeń wykorzystania modeli DSGE (*Dynamic Stochastic General Equilibrium*) podjęli czołowi amerykańscy ekonomiści w ramach prac Komitetu ds. Nauki i Technologii. Warto przytoczyć opinię R. Solowa, jednego z członków tego Komitetu, który stwierdził m.in.: „[modele DSGE] przyjmują za oczywistość to, że cała gospodarka może być pomyślana, jakby była pojedynczą, spójną osobą lub dynastią realizującą racjonalnie zaprojektowany długofalowy plan, od czasu do czasu zakłócany przez nieoczekiwane szoki, ale adaptującą się do nich w racjonalny i spójny sposób” (por. Solow R., „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-111hhrg57604/pdf/CHRG-111hhrg57604.pdf> (dostęp 25.05.2017 r.)).

<sup>166</sup> O modelach DSGE i ich zastosowaniu w analizie skutków wspólnej polityki rolnej więcej w publikacji A. Bezat, S. Figiel, J. Kufel, *Model dynamicznego, stochastycznego stanu równowagi ogólnej jako narzędzie wspierania formułowania założeń wspólnej polityki rolnej*, Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, Warszawa 2009.

Metody te mają jednak poważne ograniczenia wynikające z kilku kwestii. Po pierwsze, każdy model jest oparty na wielu założeniach teoretycznych, dotyczących sposobu funkcjonowania gospodarki i jej podmiotów oraz zachodzących pomiędzy nimi relacji. Modele odzwierciedlają więc „uproszczoną” rzeczywistość, a wiarygodność wyników modelowania uzależniona jest od rzetelności i kompleksowości założeń leżących u podstaw każdego modelu. Po drugie, w modelach zawsze zakłada się pozytywną zmianę. „Tautologiczność” założeń stosowanych modeli uniemożliwia uchwycenie negatywnych skutków działania funduszy unijnych, które mogą wystąpić w długim okresie, po wyczerpaniu efektów popytowych<sup>167</sup>. Po trzecie, stosowane w ewaluacji polityki spójności modele przeważnie zakładają pełną absorpcję funduszy pomocowych oraz ich efektywne wydatkowanie, co w rzeczywistości może przebiegać zgoła inaczej<sup>168</sup>. Po czwarte w końcu, zastosowanie modelowania możliwe jest tylko w przypadku, gdy wartość pomocy zewnętrznej stanowi znaczący odsetek całkowitego PKB kraju lub regionu – w innym przypadku wpływ funduszy pomocowych okaże się zbyt mały i praktycznie „niewykrywalny” w rezultatach modelowania. Dodatkowo, ze względu na złożoność międzyregionalnych powiązań i „wycieków” związanych z przepływem czynników produkcji, zasadność zastosowania tego typu metodyki praktycznie ogranicza się do analiz prowadzonych na poziomie krajowym<sup>169</sup>. Modelowanie nie odpowiada więc na potrzeby określenia wpływu funduszy unijnych na gospodarki lokalne.

Warto nadmienić, że inną metodyką, stosowaną w ocenie skuteczności wpływu funduszy na rozwój gospodarczy w ramach podejścia badawczego opartego na studium statystycznym, jest analiza relacji pomiędzy zmianami wskaźników odzwierciedlających rozwój społeczno-gospodarczy poszczególnych regionów (województw), powiatów lub gmin a wielkością zaabsorbowanych przez nie środków unijnych<sup>170</sup>. Badania takie w odróżnieniu od modelowania ekonometrycznego nie pozwalają na estymację (prognozowanie wpływu funduszy na wybrane wskaźniki), lecz wykorzystują dane historyczne, na podstawie których ustala się ewentualny kierunek i siłę związku między zaobserwowanymi zmianami a wielkością wykorzystanych funduszy. Istotnie ogranicza to zastosowanie tej metody w ramach ewaluacji typu *ex post*. Dodatkowo wyniki tego typu analiz wskazują jedynie na fakt współwystępowania pewnych zjawisk, nie dając jednak odpowiedzi na pytanie o ewentualną przyczynowość (statystycznie istotny związek między zmianą zmiennej zależnej i niezależnej może być kwestią przypadku lub wpływu trzeciej zmiennej). Dla przykładu: dodatnie

---

<sup>167</sup> G. Gorzelak, „Fakty i mity rozwoju regionalnego”, *Studia Regionalne i Lokalne*, nr 2(36)/2009, s. 5-27.

<sup>168</sup> L. Polverari, J. Bachtler i in., *Balance of Competences Cohesion Review: Literature Review on Cohesion Policy*, Final report to the Department for Business, Innovation and Skills 2014

<sup>169</sup> Należy jednak zaznaczyć, że pomimo tego ograniczenia analizy wykorzystujące modelowanie ekonometryczne są stosowane również na poziomie regionów. Przykładem może tu być zastosowanie modelu HERMIN do określenia wpływu funduszy unijnych na gospodarki 16 polskich województw.

<sup>170</sup> Chodzi tu o określenie wskaźnika korelacji dynamiki zmian wybranych wskaźników z wielkościami środków pomocy unijnej w ujęciu bezwzględnym lub na jednego mieszkańca.

korelacje pomiędzy pozyskanymi środkami a wzrostem gospodarczym notowanym w gminach i powiatach są pochodną ich ogólnego poziomu rozwoju – jednostki o wyższym poziomie rozwoju, posiadające większe wpływy budżetowe, są w stanie pozyskać większe środki w ramach funduszy (np. w ramach polityki spójności i wspólnej polityki rolnej UE). Badanie korelacji nie może więc stanowić metody stosowanej w ramach ewaluacji wpływu, która co do zasady bazuje na przyczynowości relacji interwencja – rezultat<sup>171</sup>.

Drugim możliwym do zastosowania w ewaluacjach wpływu podejściem badawczym są eksperymenty i quasi-eksperymenty, które pozwalają na ustalenie efektów konkretnej interwencji przez wyznaczenie, a następnie porównanie (w ramach tej samej populacji) jednostek, które różnią się tylko jedną rzeczą – tym, że zostały lub nie zostały poddane interwencji. Co do zasady podejście to cechuje się silną trafnością wewnętrzną (*internal validity*), co oznacza, że zaobserwowane różnice w grupie eksperymentalnej i kontrolnej z dużym prawdopodobieństwem są wynikiem wdrożonej interwencji<sup>172</sup>.

Stosowanie podejścia quasi-eksperymentalnego w ewaluacjach skuteczności funduszy unijnych we wzmacnianiu wzrostu gospodarczego na poziomie krajowym, regionalnym i lokalnym nie jest jednak powszechne<sup>173,174</sup>. Do tych stosunkowo nielicznych badań, których celem jest określenie efektu netto absorpcji funduszy polityki spójności na gospodarki, opierających się na quasi-eksperymentalnym planie badawczym, w których podmiotami badania są jednostki terytorialne, zaliczyć należy m.in. analizy

---

<sup>171</sup> Na gruncie polskim przykładami zastosowania takiej metodyki mogą być badania: D. Wojtowicz, T. Kupiec, „Fundusze unijne szansą na wzrost gospodarczy? Studium przypadku województwa lubelskiego”, *Studia Regionalne i Lokalne*, nr 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*, Wyższa Szkoła Informatyki i Zarządzania, Rzeszów 2013, oraz G. Gorzelak, „Wykorzystanie środków Unii Europejskiej dla rozwoju kraju – wstępne analizy”, *Studia Regionalne i Lokalne*, 3(57)/2014.

<sup>172</sup> 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.

<sup>173</sup> Podejście kontrfaktyczne ma szerokie zastosowanie w badaniach ewaluacyjnych projektów i programów współfinansowanych ze środków unijnych, w których jednostkami „dzielonymi” na grupę eksperymentalną i kontrolną są osoby lub przedsiębiorstwa. Na gruncie polskim przykładem badań wykorzystujących metody kontrfaktyczne mogą być m.in. badania ewaluacyjne PARP (np. *Ewaluacje ex post programów Phare 2002 i 2003 SSG, Badania efektów netto projektów SPO WKP 2004-2006 wspierających wzrost konkurencyjności przedsiębiorstw, Analizy efektów netto wybranych działań POIG 2007-2013 skierowanych bezpośrednio do przedsiębiorstw, Oceny efektów dofinansowania POKL 2007-2013 studiów podyplomowych oraz kształcenia w grupie wiekowej 50 plus*) oraz ewaluacje regionalne (np. *Ocena wsparcia udzielonego w projektach realizowanych przez organizacje pozarządowe w ramach komponentu regionalnego POKL dla województwa pomorskiego w kontekście przyszej perspektywy finansowej, Ewaluacja RPO WiM w kontekście konkurencyjności firm, produktów i usług*).

<sup>174</sup> Por. J. Pieńkowski, P. Berkowitz, *Econometric assessments of Cohesion Policy growth effects: How to make them more relevant for policy makers?*, Regional Working Paper, Komisja Europejska, Dyrekcja Generalna ds. Polityki Regionalnej i Miejskiej 2015.

Hagena i Mohla<sup>175</sup>, Beckera i in.<sup>176</sup>, Pellegriniego i in.<sup>177</sup>, Percoco i Gagliardiego<sup>178</sup> oraz – jako jedyne badanie koncentrujące się na poziomie lokalnym – Crescenziego i Giuy<sup>179</sup>. W tego typu badaniach przy podziale jednostek na grupy poddane interwencji i kontrolne wykorzystuje się najczęściej metody: PSM, *Regression Discontinuity Design* (RDD) lub *Local Average Treatment Effect* (LATE).

Zastosowanie quasi-eksperymentalnego planu badawczego wykorzystującego metodę PSM w przypadku badania wpływu funduszy unijnych na gospodarki lokalne w Polsce ma jednak poważne ograniczenia. Po pierwsze, w klasycznym podejściu metoda PSM bazuje na modelu regresji logistycznej, co oznacza, że warunkiem *sine qua non* jej wykorzystania jest zidentyfikowanie binarnej zmiennej zależnej. W praktyce PSM można wykorzystywać w badaniu, w którym mamy do czynienia ze ściśle określonymi grupami jednostek poddanych i niepoddanych interwencji (w tym przypadku: korzystających i niekorzystających z funduszy pomocowych). Inną jest natomiast sytuacja, gdy interwencji zostały poddane wszystkie interesujące nas jednostki, tylko w różnym stopniu. Wykorzystanie regresji logistycznej nie jest więc możliwe w badaniu wpływu funduszy unijnych w sytuacji, w której każda jednostka terytorialna została poddana interwencji. Obecnie wszystkie polskie gminy realizowały projekty współfinansowane ze środków UE, choć różnił je zakres i wartość pozyskanych dotacji.

## Model SPSM – założenia teoretyczne

Patrząc na praktykę ewaluacji ostatnich lat, efekt netto interwencji otrzymywany jest zazwyczaj przy użyciu metod pochodnych PSM. PSM jako metoda została zaproponowana w 1983 r. w artykułach Rosenbauma i Rubina<sup>180</sup>. Od tamtego czasu metoda zyskała szerokie grono zwolenników oraz krytyków. W ostatnich latach popularność metody PSM wzrosła, co wiązało się przede wszystkim z rekomendacją Komisji Euro-

<sup>175</sup> 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), s. 353-365.

<sup>176</sup> O. Becker, P. Egger P., M. von Ehrlich, R. Feng, „Going NUTS – The effect of EU structural funds on regional performance”, *Journal of Public Economics* 2010, 94 (9-10), s. 578-590.

<sup>177</sup> G. Pellegrini i in., „Measuring the Impact of the European Regional Policy on Economic Growth: a Regression Discontinuity Approach”, *Papers in Regional Science* 2013, vol. 92, no. 1.

<sup>178</sup> M. Percoco, L. Gagliardi, „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> (dostęp 25.05.2017 r.)

<sup>179</sup> 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.

<sup>180</sup> D.B. Rubin, N. Thomas, „Combining propensity score matching with additional adjustments for prognostic covariates”, *Journal of the American Statistical Association* 2000; Imbens G.W., „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, 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.

pejskiej, zalecającą wykorzystanie metod kontrfaktycznych w ewaluacji wpływu funduszy polityki spójności UE<sup>181</sup>. Jednakże, jak zaznaczono we wstępie, metoda PSM nie była do tej pory stosowana w większości przypadków ewaluacji rozwoju terytorialnego i regionalnego, ze szczególnym uwzględnieniem wpływu polityki spójności UE. Brak binarnej zmiennej zależnej (otrzymanie lub nieotrzymanie wsparcia) mógłby być w niektórych przypadkach ominięty przez zastosowanie wielomianowej analizy regresji logistycznej, jednakże w tym przypadku wciąż musielibyśmy mieć do czynienia z dyskretną i mierzalną miarą wsparcia, o niewielkiej liczbie poziomów (wsparcie dla regionów zazwyczaj ma charakter zmiennej ciągłej). Wykorzystanie bardziej zaawansowanych odmian regresji logistycznej możliwe jest więc rzadko. Co więc zrobić w sytuacji, gdy nie mamy do dyspozycji zero-jedynkowej zmiennej, którą można byłoby wykorzystać jako zmienną zależną?

Proponowane rozwiązanie to SPSM – *Stratified Propensity Score Matching*. W tym modelu zmienna interwencji (*treatment variable*) może być zmienną ciągłą. Co więcej, zmienna interwencji nie jest tu zmienną zależną, gdyż model SPSM w ogóle nie potrzebuje zmiennej zależnej do wyliczenia miar podobieństwa (*propensity scores*). Zmienna interwencji nie odpowiada więc na pytanie: „czy moja jednostka została poddana interwencji, czy też nie?”, tylko na pytanie: „w jakim stopniu moja jednostka została poddana interwencji?”. Sytuacja kontrfaktyczna jest więc tu inaczej zdefiniowana: nie stanowi jej podjęcie bądź niepodjęcie interwencji względem danej jednostki, lecz stopień tej interwencji. W tym przypadku tradycyjna zero-jedynkowa definicja kontrfaktyczna staje się de facto szczególnym przypadkiem sytuacji kontrfaktycznej zdefiniowanej powyżej.

W modelu SPSM miary podobieństwa liczone są w zupełnie inny sposób niż w przypadku modelu PSM. Metodą statystyczną, która stoi za tym modelem, jest rozmyte segmentowanie hierarchiczne (*fuzzy hierarchical clustering*). Samo segmentowanie polega na wyłonieniu maksymalnie podobnych grup jednostek. W odróżnieniu od zwykłego segmentowania, w którym każdej jednostce przypisana jest pojedyncza wartość segmentu, w segmentowaniu rozmytym jednostce przypisana jest nie tylko dyskretna wartość segmentu (np. fakt przynależności do segmentu 1. lub segmentu 2.), lecz także wartość prawdopodobieństwa przynależności do segmentu. W tym modelu wartości prawdopodobieństw w obrębie danego segmentu traktowane są jako miary podobieństwa. Dalsza procedura przebiega podobnie jak w przypadku klasycznego modelu PSM – jednostki należące do danego segmentu parowane są na podstawie prawdopodobieństwa przynależności do danego segmentu. Przykłady działania

---

<sup>181</sup> [https://crie.jrc.ec.europa.eu/sites/default/files/documents/CIE\\_Guidance\\_EN.pdf](https://crie.jrc.ec.europa.eu/sites/default/files/documents/CIE_Guidance_EN.pdf) (dostęp 28.03.2017 r.)

tej metody w praktyce opisane są w dalszej części rozdziału, a techniczny opis procedur statystycznych znajduje się w następnym podrozdziale<sup>182</sup>.

Metoda została nazwana Stratyfikowanym PSM (*Stratified PSM*), gdyż w wyniku jej działania powstaje kilka segmentów, które mogą być traktowane jako warstwy stratyfikacyjne. Niejako ubocznym efektem zastosowania tej metody jest możliwość wykonania analiz nie tylko na całym sparowanym zbiorze jednostek poddanych interwencji, lecz również w podziorach. Zastosowanie metody hierarchicznej pozwala na wyłonienie m.in. niewielkich, lecz ściśle wyspecjalizowanych segmentów, dla których efekty netto mogą być inne niż dla całej populacji jednostek poddanych interwencji. Jedną z wad PSM jest „umiejętność” tej metody generowania efektów netto wyłącznie na całej zadanej populacji. Chęć zejścia na niższy poziom (np. ocena efektów netto w innych podziałach czy podgrupach niż cała badana próba) wymaga powtórzenia procedury PSM. W odróżnieniu od niej SPSM wstępnie dzieli próbę (bądź populację) na segmenty, dzięki czemu analiza efektów może odbyć się na różnych poziomach. Praktyczne implikacje tego podejścia można zobaczyć w podanych przykładach w dalszej części rozdziału.

### SPSM - opis techniczny

Zaproponowana metoda opiera się na hierarchicznym algorytmie segmentowania Warda<sup>183</sup>, który jest metodą aglomeracyjną. Oznacza to, że każda obserwacja wchodzi do analizy jako pojedynczy segment, a następnie w kolejnych iteracjach najbliższe segmenty łączone są w jeden. Ostatecznie algorytm kończy pracę po uzyskaniu założonej uprzednio przez badacza liczby segmentów. Określanie optymalnej liczby segmentów w przypadku metod hierarchicznych wykracza poza zakres tego artykułu, jednak istnieje na ten temat bogata literatura<sup>184</sup>.

Algorytm Warda minimalizuje w każdym kroku błędy sumy kwadratów<sup>185</sup> (*error sum of squares* – ESS) w następujący sposób:

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

---

<sup>182</sup> Warto przy okazji zaznaczyć, że opis techniczny zawarty w następnym podrozdziale ma na celu wyłącznie zapewnienie replikowalności metody dla innych badaczy. Nie chcemy, by metoda SPSM była postrzegana jako metoda „czarnej skrzynki”, w której miary podobieństw generowane są w tajemniczy i niejasny sposób. Czytelnik, który nie jest zainteresowany statystycznymi zawiłościami metody, bez żadnej straty może pominąć część techniczną i przejść do przykładów zastosowania metody.

<sup>183</sup> J. H. Ward., „Hierarchical Grouping to Optimize an Objective Function”, *American Statistical Association Journal* 1963.

<sup>184</sup> 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.

<sup>185</sup> Innymi słowy, każdy krok ma za zadanie zmniejszać tzw. funkcję błędów i polepszać dopasowanie modelu do danych.

gdzie  $x_i$  jest wartością dla  $i$ -tej jednostki, zaś  $n$  – wielkością próby. Algorytm powtarzany jest do momentu uzyskania ostatecznej liczby segmentów (lub „zbiorów”, wedle nomenklatury samego Warda).

Działanie każdej iteracji definiowane jest jako:

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

gdzie  $p_{n-1}$  jest mniejszym z dwóch segmentów (podzbiorów), zaś  $q_{n-1}$  jest większym z dwóch segmentów (zbiorów).

Procedura segmentowania hierarchicznego (podobnie jak inne procedury segmentowania, takie jak  $k$ -średnich czy  $k$ -median) oblicza domyślnie tylko przynależność do segmentu. W trakcie obliczania kolejnych kroków powstaje jednak również tablica dystansów – najczęściej w formie tablicy niepodobieństwa (*dissimilarity matrix*) lub podobieństwa (*similarity matrix*) – która może zostać użyta do obliczenia prawdopodobieństw przynależności do segmentów. Dzięki przekształceniom tablicy dystansów uzyskujemy wartości segmentowania rozmytego (*fuzzy clustering*). Mając informację o przynależności do segmentu oraz dysponując odległościami z tablicy dystansów, można obliczyć rozmyte miary segmentowania hierarchicznego (*Fuzzy Hierarchical Clustering Scores* – FHCS), które w następnym kroku posłużą jako miary PSM<sup>186</sup>. FHCS definiowane jest jako:

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

gdzie DS to miara odległości (*Dissimilarity Score*).

Procedura wyliczania miar segmentowania rozmytego powinna być przeprowadzona dla każdego segmentu z osobna. Jak powiedziano powyżej, FHCS w dalszej procedurze traktowane są jako miary PSM, tak więc dalsza procedura uwzględnia różnice między wartościami każdych dwóch jednostek leżących najbliżej pod względem FHCS. Dla każdego segmentu można więc maksymalnie ustalić  $n-1$  par, zakładając, że liczba jednostek w segmencie =  $n$ .

## Metoda szacowania efektu netto

W sytuacji, gdy zostały policzone miary SPSM, można przystąpić do szacowania efektu netto. W przedstawionych poniżej przykładach dla każdej pary w obrębie segmentu liczone były różnice wysokości wsparcia oraz różnice dla poszczególnych wskaźników sukcesu (dla pierwszego badania był tylko jeden wskaźnik sukcesu, dla drugiego – kilkanaście)<sup>187</sup>. Ponieważ miary SPSM pokazują bliskość między jednostkami w ob-

<sup>186</sup> Rozumiane jako miary podobieństw jednostek między sobą, interpretowane podobnie jak w przypadku działania klasycznej metody PSM.

<sup>187</sup> W każdym badaniu ewaluacyjnym należy zawczasu określić wskaźniki sukcesu. Są one niezależne względem opisywanej metody. Przykłady pokazują, że można liczyć np. zarówno efekty netto dochodów własnych gmin, jak i efekty netto napływu kapitału zagranicznego w powiatach.

rębie segmentu, należy najpierw policzyć wartości efektów w segmentach, a dopiero później efekty netto dla całej próby. Metoda nie pozwala na liczenie efektów netto na dwóch lub więcej wskaźnikach sukcesu jednocześnie, umożliwia jednak policzenie całkowitego efektu netto, który szacowany jest na podstawie sumy różnic wsparcia oraz poziomów wskaźników sukcesu, tyle że nie w segmentach, ale na całej sparowanej za pomocą miar PSM populacji.

Podsumowując część techniczną, należy zauważyć, że:

- metoda SPSM sprawdza się w ocenie wpływu tych polityk, które obejmują większość lub wszystkie jednostki z danej populacji, szczególnie dobrze sprawdzając się w ocenie efektów funduszy unijnej polityki spójności na poziomie gmin lub powiatów;
- SPSM umożliwia liczenie efektów netto interwencji nie tylko na poziomie całej populacji (np. powiatów), ale również w podgrupach (np. efekty netto w powiatach miejskich lub efekty netto w dużych terytorialnie powiatach o charakterze przede wszystkim rolniczym);
- w odróżnieniu od metody PSM, której działanie ograniczone jest tylko do sytuacji, gdy istnieją dwie grupy jednostek (wsparci i niewsparci), SPSM działa w sytuacji, gdy wszystkie lub prawie wszystkie jednostki otrzymały wsparcie;
- żeby dało się mówić w tym przypadku o sytuacji kontrfaktycznej, poszerzono jej definicję. W klasycznym rozumieniu sytuacja kontrfaktyczna to odpowiedź na pytanie: „co by było, gdyby omawiane jednostki w ogóle nie otrzymały wsparcia?“, w SPSM odpowiada się na pytanie: „co by było, gdyby omawiane jednostki otrzymały wsparcie w innej wysokości?“. W dalszym ciągu jest to jednak sytuacja kontrfaktyczna.

## **Przykłady zastosowania SPSM w badaniach ewaluacyjnych<sup>188</sup>**

### **Ewaluacja III i IV osi Programu Rozwoju Obszarów Wiejskich 2007-2013**

Pierwszym przykładem zastosowania metody SPSM jest badanie wpływu realizacji Programu Rozwoju Obszarów Wiejskich na lata 2007-2013 (PROW 2007-2013) na zmiany zachodzące na terenach wiejskich w zakresie funkcjonowania gospodarki<sup>189</sup>. Przeprowadzenie analiz poprzedziło dokonanie klasyfikacji gmin ze względu na wybrane cechy społeczne, gospodarcze i przestrzenne. Stworzenie w miarę jednorodnych grup gmin było niezbędne do porównania zmian w dochodach własnych per capita (zmienna wyjściowa) bliźniaczych gmin, które różniły się ze względu na wysokość nakładów PROW 2007-2013, i tym samym określenia wpływu netto tychże funduszy. Dało to także podstawy do udzielenia odpowiedzi na pytanie dotyczące wpływu cha-

<sup>188</sup> Zaprezentowane zostały ewaluacje, w których autorzy niniejszego rozdziału byli członkami zespołów badawczych.

<sup>189</sup> Badanie ewaluacyjne pt. „Ocena wpływu PROW 2007-2013 na jakość życia na obszarach wiejskich z uwzględnieniem podejścia LEADER” zrealizowane na zlecenie Ministerstwa Rolnictwa i Rozwoju Wsi przez EGO – Evaluation for Government Organizations s.c., Warszawa, listopad 2016.



rakterystyki gmin na osiągnięte efekty netto wykorzystania funduszy PROW 2007-2013. Klasyfikację gmin przeprowadzono w oparciu o wybrane cechy społeczne, gospodarcze i przestrzenne, biorąc pod uwagę wybrane dane z roku 2006 (jako roku poprzedzającego interwencję), aby móc wyróżnić te gminy, które były do siebie maksymalnie podobne, zanim zaczęły się nabory wniosków o dofinansowanie w ramach Programu.

Klasyfikacja gmin odbyła się w kilku etapach. Po pierwsze, przygotowane zostały wskaźniki syntetyczne, służące do redukcji wymiarów. W wyniku tego działania uzyskano kilkanaście wskaźników zrelatywizowanych do liczby ludności gminy. W drugim kroku uzyskane wskaźniki zostały wystandaryzowane w taki sposób, by miały szansę zbudować segment z równą siłą. Posłużono się w tym celu klasyczną standaryzacją Z (*Z-scores*)<sup>190</sup>.

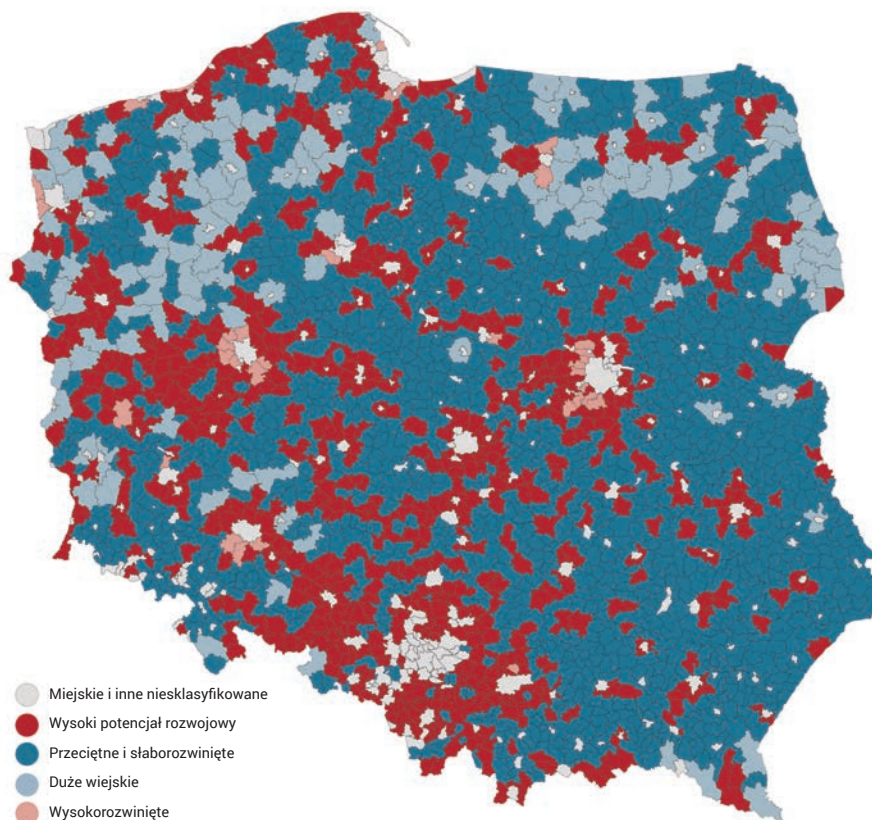
Kolejnym etapem analizy była próba wyłonienia gmin podobnych do siebie pod względem obserwowalnych cech. W tym celu posłużono się metodą bazującą na hierarchicznej analizie segmentacyjnej<sup>191</sup>. Uzyskana w trakcie analizy segmentacyjnej macierzy bliskości (*proximity matrix*) dostarczyła informacji na temat odległości poszczególnych gmin od centrum segmentu. Wykorzystując metodę segmentowania rozmytego, każdej gminie przypisano prawdopodobieństwo przynależności do każdego z czterech powstałych segmentów: (1) gminy o wysokim potencjale rozwojowym, (2) gminy przeciętne i słabo rozwinięte, (3) duże gminy wiejskie, (4) gminy wysoko rozwinięte. W ostatnim kroku posłużono się analizą dyskryminacyjną służącą uzyskaniu ostatecznych prawdopodobieństw przynależności do segmentów. Te ostatnie traktowane były w analizie jako miary PSM – miary bliskości gmin względem siebie – i pozwoliły na wyłonienie par tzw. bliźniaków. „Bliźniaczość” gmin definiowana była jako podobieństwo przed momentem działania Programu, czyli w 2006 r., pod względem takich cech jak: powierzchnia, dochody gminy, liczba ludności, poziom bezrobocia itp.

---

<sup>190</sup> Standaryzacja zmiennych ma zapewnić ich porównywalność i usunąć efekty użycia różnych skal.

<sup>191</sup> Więcej o metodzie: D. Wojtowicz, Ł. Widła-Domaradzki, „Estimating the impact of the EU Cohesion Policy on local level”, w: 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, Bruksela 2016, <http://www.regionalstudies.org/uploads/documents/Lessons4CP.pdf> (dostęp 25.05.2017 r.).

**Rysunek 1. Rozkład przestrzenny gmin wg przyjętej klasyfikacji (2006 r., stan przed rozpoczęciem wdrażania III i IV osi PROW)**



Źródło: Ocena wpływu PROW 2007-2013 na jakość życia na obszarach wiejskich z uwzględnieniem podejścia LEADER

W modelowaniu wpływu netto projektów współfinansowanych ze środków UE wykorzystano wartość funduszy finansujących projekty zrealizowane w ramach PROW, a także programów wdrażanych w ramach polityki spójności UE. Przyjęto założenie, że im większe kwoty zaabsorbowane przez gminę, tym większy będzie ich obserwowany efekt netto<sup>192</sup>. Dla każdego rodzaju interwencji (PROW i inne programy finansowe)

<sup>192</sup> Przyjęcie założenia było konieczne, gdyż w badaniu wykorzystywana była tylko jedna zmienna zależna. Chcąc wyjaśnić wpływ dwóch różnych form finansowania projektów (PROW i fundusze polityki spójności UE) bez przyjęcia takiego założenia, obserwowana byłaby sytuacja, kiedy interwencja o niższym poziomie nakładów finansowych zawsze będzie wywoływać wyższy efekt netto (ponieważ efekt netto jest opisany jako związek między poziomem wydatków i zmianą w dochodach własnych gminy w ujęciu per capita).

wane w ramach polityki spójności UE) stworzono „idealny model”, w którym wszystkie zaabsorbowane środki odpowiedzialne były za całkowity wzrost dochodów. „Idealny model” był traktowany jako punkt odniesienia i testowany pod kątem jego przystawalności do rzeczywistości. W następnym kroku w każdej kontrfaktycznej parze odejmowano od siebie wysokość wsparcia oraz zmiany w dochodach własnych w ujęciu per capita. Suma tych różnic dla wszystkich kontrfaktycznych par stanowiła podstawę oszacowania efektu netto środków, jak to przedstawiono w tabeli 1. Wpływ netto pokazuje, w jakim stopniu zaabsorbowane przez gminę jedno euro przyczyniło się do wzrostu całkowitych dochodów własnych w ujęciu per capita. Na przykład wartość 0,049 oznacza, że każde wykorzystane euro przyniosło zwiększenie dochodów własnych gminy o prawie 5 groszy.

**Tabela 1. Efekt netto funduszy zaabsorbowanych przez gminy w ramach PROW i programów wdrażanych w ramach polityki spójności UE (w zł)**

	Fundusze polityki spójności UE	PROW i fundusze polityki spójności UE	PROW
<b>Wszystkie gminy</b>	<b>0,049</b>	<b>0,083</b>	<b>0,085</b>
Gminy o wysokim potencjale rozwojowym	0,049	0,085	0,095
Gminy przeciętne i słaborozwinięte	0,049	0,080	0,080
Duże gminy wiejskie	0,056	0,094	0,081
Gminy wysokorozwinięte	0,054	0,096	0,090

Źródło: opracowanie własne.

Wyniki modelowania wykazały, że w przypadku gmin wiejskich wyższe efekty netto w postaci wzrostu dochodów własnych per capita generowały projekty współfinansowane ze środków PROW, a nie funduszy polityki spójności UE. Ponadto największe efekty netto wystąpiły w gminach sklasyfikowanych jako jednostki o wysokim potencjale rozwojowym oraz wysokorozwinięte. Jednocześnie największy pozytywny wpływ na wzrost dochodów gmin miał miejsce w przypadku inwestycji finansowanych z PROW i funduszy polityki spójności w wysokorozwiniętych i dużych jednostkach.

## Ewaluacja I osi Programu Operacyjnego Rozwój Polski Wschodniej (2007-2013)

Innym badaniem, w którym zastosowano SPSM, jest „Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej”<sup>193</sup>.

W przypadku tego badania jednostką analizy były powiaty, nie zaś – jak w przypadku badania PROW – gminy. Różnicę tę należy tłumaczyć celem badania: o ile zmiany zachodzące w rozwoju terenów wiejskich powinno się badać na możliwie największym poziomie dezagregacji, o tyle badanie dotyczące rozwoju przedsiębiorczości powinno brać pod uwagę przynajmniej poziom powiatów, ze względu na możliwość występowania efektów *spill-over* (znacząca zmiana wskaźników przedsiębiorczości w jednej gminie będzie miała wpływ na sytuację w gminie ościennej – traktowanie ich jako osobnych i niepowiązanych jednostek znacząco zmniejszyłoby jakość opracowania statystycznego). Istotną również była tu kwestia dostępności niektórych wskaźników w podziale powiatowym, a niedostępnych w podziale na gminy.

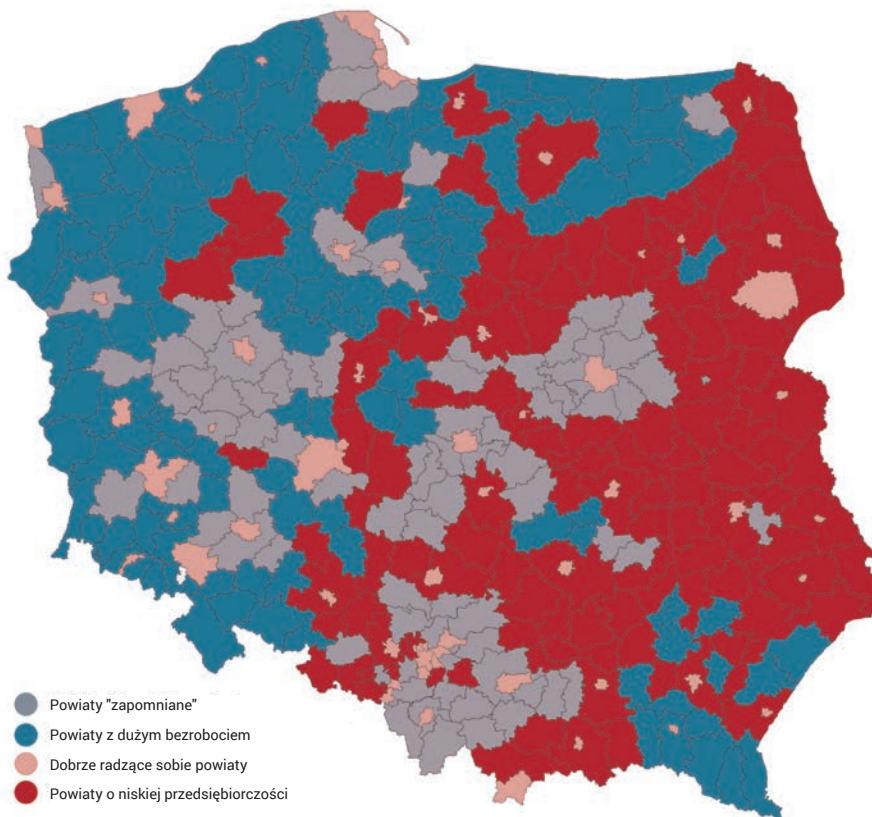
Podobnie jak w przypadku pierwszego z opisanych badań również w przypadku PORPW podzielono badaną populację na stratyfikacyjne warstwy. Mapa zamieszczona poniżej pokazuje rozkład poszczególnych typów powiatów w całej Polsce, jednak do modelowania SPSM zostały użyte tylko powiaty z makroregionu Polski Wschodniej. Typologia powiatów została przeprowadzona w oparciu o wskaźniki BDL GUS za 2006 r. Zmienne segmentacyjne zostały zgrupowane w trzy bloki: przedsiębiorczość, infrastruktura, zmienne uzupełniające<sup>194</sup>. Wyróżnione zostały cztery segmenty powiatów: dobrze radzące sobie powiaty, powiaty o niskiej przedsiębiorczości, powiaty z dużym bezrobociem oraz te, które można było zaklasyfikować do grupy powiatów „zapomnianych”, zgrupowanych przede wszystkim wokół dużych miast i aglomeracji, charakteryzujących się nieco mniejszą niż przeciętna powierzchnią i nieco niższą gęstością zaludnienia niż przeciętny powiat w Polsce (patrz rys. 2).

---

<sup>193</sup> „Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej”, konsorcjum Agrotec Polska sp. z o.o., WYG PSDB sp. z o.o. oraz EGO – Evaluation for Government Organizations s.c., Warszawa, grudzień 2016.

<sup>194</sup> Szczegółowe założenia metodyczne (wybór zmiennych służący do wyłonienia miar podobieństw, szczegółowy opis typów powiatów itp.) znajdują się w aneksie metodologicznym do raportu końcowego z badania.

**Rysunek 2. Rozkład przestrzenny powiatów wg przyjętych segmentów (2006 r., stan przed rozpoczęciem wdrażania PORPW 2007-2013)**



Źródło: Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 (PORPW) na rozwój przedsiębiorczości w Polsce Wschodniej

Podobnie jak w przypadku pierwszego z omawianych badań w niniejszym modelowaniu wykorzystano wartości funduszy finansujących projekty zrealizowane w ramach PORPW, ale także innych programów dotyczących rozwoju przedsiębiorczości wdrażanych w ramach polityki spójności UE w Polsce Wschodniej. Wśród funduszy mających w założeniu działać podobnie do środków z I osi PORPW<sup>195</sup> znalazły się programy: Program Operacyjny Innowacyjna Gospodarka, Program Operacyjny Infrastruktura i Środowisko oraz 16 RPO. W odróżnieniu jednak od badania PROW w przypadku ewaluacji PORPW wyróżniono nie jeden, lecz aż 16 wskaźników sukcesu. Przez wskaźniki sukcesu w tym badaniu rozumiano różne zmienne, które mogą być rozumiane jako wskaźniki poprawy

<sup>195</sup> Przedmiotem przytaczanych badań ewaluacyjnych była I oś PORPW.

konkurencyjności gospodarczej (np. liczba pracujących, liczba nowo zakładanych firm, obecność kapitału zagranicznego itp.). W związku z tym modelowanie kontrfaktyczne za pomocą modelu SPSM musiało zostać wykonane 16 razy. Poniżej przedstawiono wynik działania modelowania na przykładzie jednego wskaźnika sukcesu. Jest nim bilans nowo powstających oraz wykreślanych z rejestru REGON firm. Dodatni bilans oznacza, że więcej firm było zakładanych niż wykreślanych, ujemny bilans oznacza, że w danym okresie więcej firm wykreślono, niż wpisano. Wszystkie wartości zostały zrelatywizowane do wydatkowania jednego miliona złotych w ramach PORPW oraz innych środków przeznaczonych na rozwój przedsiębiorczości. Wskaźnik ujęty w tabeli pokazuje także średnią zmianę roczną. Oznacza to, że każdy milion wydanych złotych w ramach PORPW średnio rocznie powodował wzrost liczby firm w makroregionie o 0,13.

**Tabela 2. Efekt netto funduszy zaabsorbowanych przez gminy w ramach PORPW i programów wdrażanych w ramach polityki spójności UE (w euro)**

	Efekty netto dla bilansu powstających / wykreślonych
<b>Wszystkie powiaty</b>	<b>0,13</b>
Powiaty I typu <b>(Zapomniane)</b>	-0,52
Powiaty II typu <b>(Z dużym bezrobociem)</b>	-0,02
Powiaty III typu <b>(Dobrze sobie radzące - miejskie)</b>	-0,08
Powiaty IV typu <b>(O niskiej przedsiębiorczości)</b>	0,33

Źródło: opracowanie własne.

Jak widać, ogólny efekt netto jest dodatni. Oznacza to, że środki wydatkowane w ramach I osi PORPW działały lepiej niż inne środki skierowane na rozwój przedsiębiorczości, jeśli chodzi o bilans nowo powstających i wykreślanych firm. Analiza w segmentach pokazuje jednak dość duże zróżnicowanie pomiędzy powiatami. Niewielkie, lecz ujemne efekty netto dla powiatów z dużym bezrobociem oraz dobrze sobie radzących miejskich oznaczają, że inne środki (spoza PORPW) lepiej wpływały na omawiany bilans. Wśród powiatów pozostałych dwóch typów obserwujemy sytuację diametralnie różną. W przypadku powiatów „zapomnianych” (na które składają się głównie powiaty podmiejskie, okoliczne względem aglomeracji) bilans działania PORPW jest negatywny: każdy milion wydanych złotych w ramach PORPW działał w tych powiatach gorzej niż milion złotych wydany w ramach innych środków. W tym przypadku efekt równy -0,52 oznacza, że na

każde dwa miliony wydanych złotych w omawianych powiatach wykreślano mniej więcej o jedną firmę więcej, niż zostało ich założonych. Zupełnie inna jest sytuacja w powiatach o niskiej przedsiębiorczości – tu środki PORPW działały lepiej niż inne środki przeznaczone na rozwój przedsiębiorczości. Każde trzy miliony złotych wydane w ramach PORPW skutkowało założeniem o jedną firmę więcej, niż zostało wykreślonych z rejestru REGON.

## **Podsumowanie**

Zaprezentowana w niniejszym rozdziale metoda jest odpowiedzią na wyzwania związane z ograniczeniami zastosowania ewaluacji wpływu do szacowania efektów netto interwencji realizowanych na poziomie lokalnym. Jak wykazano, modele SPSM mają dwie główne zalety w porównaniu z klasyczną metodą PSM. Po pierwsze, modele te nie wymagają określenia binarnej zmiennej zależnej do ustalenia miar skłonności badanych jednostek. Co więcej, w praktyce w ogóle nie jest konieczne identyfikowanie zmiennej zależnej, bowiem w przypadku SPSM do obliczania miar skłonności wykorzystywany jest inny algorytm segmentowania. Po drugie, SPSM daje możliwość oceny efektu netto dla różnych podzbiorów jednostek objętych interwencją, a nie całej ich populacji. Badania ewaluacyjne, w których wykorzystano SPSM, dowodzą, że uzyskane wyniki mogą stanowić kluczową wiedzę dla decydentów – dostarczają informacji o efekcie netto generowanym przez projekty w zależności od poziomu rozwoju społeczno-gospodarczego jednostek, w których są wdrażane, oraz w zależności od rodzaju samych interwencji (np. działań, osi priorytetowych, programów) lub funduszy je finansujących (np. Europejski Fundusz Rolny Rozwoju Obszarów Wiejskich, Europejski Fundusz Rozwoju Regionalnego, środki własne gmin itp.).

Podsumowując, należy zaznaczyć, że modele SPSM nie są modelami pozwalającymi oszacować efekty względne, nie są również modelami makroekonomicznymi. W odróżnieniu od analiz efektów względnych w modelu SPSM można badać wpływ pojedynczej interwencji. Model ten nie jest także modelem makroekonomicznym, gdyż nie potrafi wyjaśnić, czy otrzymane efekty są niskie czy wysokie – do tego celu bardziej właściwe jest zastosowanie właśnie modeli makroekonometrycznych.

Odpowiednio zastosowany model SPSM powinien być oczywiście powiązany z charakterem interwencji. Wciąż w wielu badaniach powinno się wykorzystywać klasyczne metody kontrfaktyczne, a model SPSM nie jest względem nich konkurencyjny, lecz komplementarny. Przedstawiony model SPSM może i – zdaniem autorów – powinien być wykorzystywany w sytuacji, gdy wsparcie ma charakter ciągły, co oznacza, że nie ma możliwości znalezienia jednostek, które nie otrzymały wsparcia finansowego. Przedstawione w tekście przykłady dotyczyły badań o charakterze terytorialnym, jednak zastosowanie tej metody może być oczywiście szersze – mogą to być uczestnicy szkoleń, którzy otrzymują różnego rodzaju wsparcie, czy też bezrobotni, wobec których zastosowano dowolną ciągłą interwencję (np. bony o różnej wysokości).

## O autorach

### **Dominika Wojtowicz**

---

Doktor nauk o zarządzaniu. Absolwentka Akademii Leona Koźmińskiego oraz Stosunków Międzynarodowych Uniwersytetu Warszawskiego. Stypendystka Wydziału Nauk Politycznych *Universita' di Siena* (Włochy) oraz Wydziału Zarządzania *Universidad Politecnica de Cartagena* (Hiszpania). Jej zainteresowania badawcze koncentrują się na szeroko pojętej skuteczności interwencji publicznych, ukierunkowanych na wzmacnianie rozwoju regionalnego i lokalnego. Posiada wieloletnie doświadczenie badawcze – brała udział w szeregu projektów badawczych koncentrujących się na wybranych obszarach polityk publicznych. Jest autorką ekspertyz z zakresu wpływu funduszy unijnych na gospodarki regionalne w Polsce, zleczanych m.in. przez Parlament Europejski i Komisję Europejską. Była koordynatorką badań nad nowoczesnymi metodami ewaluacji specyficznych obszarów interwencji i regulacji publicznych, badań na temat wykorzystania ewaluacji w zarządzaniu regionalnymi programami operacyjnymi w Polsce i Hiszpanii oraz wykorzystania doświadczeń z zakresu ewaluacji w procesie opracowywania oceny skutków regulacji. Jest adiunktem w katedrze ekonomii Akademii Leona Koźmińskiego. Prowadzi zajęcia z metodyki badań ilościowych i jakościowych, stosowanych w analizach polityk na studiach podyplomowych dla pracowników administracji publicznej (Akademia Ewaluacji, Akademia Analityka, Zarządzanie w Administracji Publicznej).

### **Łukasz Widła-Domaradzki**

---

Statystyk i analityk z blisko dwudziestoletnim stażem. W swojej karierze pracował w dużych międzynarodowych firmach badawczych (MillwardBrown, Ipsos), firmach zajmujących się ewaluacjami na skalę europejską (ICF International) jak również w sektorze publicznym (Polska Agencja Rozwoju Przedsiębiorczości, Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy). Biegłe posługuje się wieloma pakietami statystycznymi, takimi jak SPSS, Stata, Statistica czy SAS. Posługuje się również specjalistycznym oprogramowaniem służącym do modelowania strukturalnego (w tym AMOS czy SmartPLS). W swojej karierze pracował zarówno z wielkimi korporacjami, pomagając lepiej zrozumieć ich klientów (Orange Polska, T-mobile, HBO) jak również brał udział w rozmaitych ewaluacyjnych i naukowych projektach badawczych, w których odgrywał rolę eksperta do spraw analiz statystycznych (np. w projekcie MUS – „Ministerstwa Uczące Się”, koordynowanym przez Kancelarię Prezesa Rady Ministrów). Jest autorem lub współautorem kilkunastu pozycji książkowych lub artykułów, ma na swoim koncie również działalność edukacyjną (zajęcia dla studentów na Akademii Leona Koźmińskiego, warsztaty ze stosowania metod ilościowych w ewaluacjach dla słuchaczy studiów podyplomowych czy też szkolenia z wykorzystania pakietu SPSS dla Ministerstwa Obrony Narodowej).



## Bibliografia:

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).
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, nr 2 (36).
7. Hagen T., Mohl P., „Does EU structural funds promote regional growth? New evidence from various panel data approach”, *Regional Science and Urban Economics* 2010, 40 (5).
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*, Wyższa Szkoła Informatyki i Zarządzania, Rzeszów 2013.
10. Pellegrini G. i in., „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> (dostęp 25.05.2017 r.).
12. Polverari L., Bachtler J. i in., *Balance of Competences 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. Ward J. H., „Hierarchical Grouping to Optimize an Objective Function”, *American Statistical Association Journal* 1963.
19. Wojtowicz D., Kupiec T., „Fundusze unijne szansą na wzrost gospodarczy? Studium przypadku województwa lubelskiego”, *Studia Regionalne i Lokalne* 3 (61)/2015.

### Inne:

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”, Konsorcjum EGO – Evaluation for Government Organizations s.c. i Fundacja Idea Rozwoju na zamówienie Urzędu Marszałkowskiego Województwa Pomorskiego, Warszawa 2013.
2. „Ewaluacja RPO WiM w kontekście konkurencyjności firm, produktów i usług. Raport końcowy”, EGO – Evaluation for Government Organizations s.c. na zamówienie Urzędu Marszałkowskiego Województwa Warmińsko-Mazurskiego, Warszawa 2012.
3. „Ewaluacja *ex post* programów Phare 2002 i 2003 SSG, regionalny i sektorowy komponent wsparcia MŚP. Raporty końcowe”, Pentor Research International na zamówienie PARP, Warszawa 2006-2007.
4. Ewaluacje *ex post* programów Phare 2002 i 2003 SSG, regionalny komponent RZL. Raporty końcowe, PBS DGA na zamówienie PARP, Sopot 2006-2007.
5. „Badanie efektów netto projektów wspierających wzrost konkurencyjności przedsiębiorstw. Raport końcowy”, PSDB sp. z o.o. na zamówienie PARP, Warszawa 2009.
6. Trzcziński R., „Impact of direct subsidies on SMEs”, w: Pokorski J. (red.), *Towards Innovative Economy Effects of Grants to Enterprises in Poland*, PARP, Warszawa 2011.
7. „Barometr Innowacyjności – analiza efektów netto wybranych Działań POIG skierowanych do przedsiębiorstw, z wykorzystaniem podejścia *counterfactual impact evaluation*. Raport z ewaluacji wewnętrznej”, PARP, Warszawa 2013.
8. „Efekty (netto) dofinansowanych studiów podyplomowych. Raport z ewaluacji wewnętrznej”, PARP, Warszawa 2014.
9. „Efekty dofinansowanego kształcenia w grupie wiekowej 50 plus. Raport końcowy”, Instytut Społeczno-Ekonomicznych Ekspertyz I-SEE na zamówienie PARP, Warszawa 2014.

10. „Analizy efektów netto wybranych Działań Programu Operacyjnego Innowacyjna Gospodarka 2007-2013 z wykorzystaniem podejścia *counterfactual impact evaluation* w ramach projektu ewaluacyjnego PARP „Barometr Innowacyjności”. Raport końcowy”, konsorcjum Uniwersytet Jagielloński i EGO – Evaluation for Government Organizations s.c. na zamówienie PARP, Warszawa 2015.
11. „Ewaluacja wpływu Programu Operacyjnego Rozwój Polski Wschodniej 2007-2013 na rozwój przedsiębiorczości w Polsce Wschodniej. Raport końcowy”, konsorcjum spółek Evalu, WYG PSDB i EGO – Evaluation for Government Organizations na zamówienie PARP, Warszawa 2016.
12. „Ocena wpływu PROW 2007-2013 na jakość życia na obszarach wiejskich z uwzględnieniem podejścia LEADER. Raport końcowy”, EGO – Evaluation for Government Organizations s.c. na zamówienie Ministerstwa Rolnictwa i Rozwoju Wsi, Warszawa 2016.



## **Ewaluacja wspierana teorią w badaniach systemu wyboru projektów innowacyjnych w programach operacyjnych na lata 2014-2020**

### **Wprowadzenie**

Rozdział prezentuje doświadczenia z realizacji kilku badań, których przedmiotem była ewaluacja systemów wyboru i oceny projektów innowacyjnych w programach operacyjnych, realizowanych w Polsce w ramach polityki spójności Unii Europejskiej na lata 2014-2020. Analizując te systemy, odwoływano się do podejścia ewaluacji opartej na teorii (*theory-based evaluation*), z zamiarem możliwie precyzyjnego odwzorowania logiki interwencji działań wspierających, aby możliwe było sformułowanie rzeczowych wskazań doskonalących systemy wyboru pod kątem ich lepszego dostosowania do specyfiki instrumentów wsparcia<sup>196</sup>. Zebrane w tym zakresie doświadczenia stanowią główną oś prezentowanych tu rozważań. Uzupełniają je wnioski wynikające z kształtowania systemów wyboru projektów w poprzedniej perspektywie finansowej (2007-2013). W przeprowadzonych badaniach, do których odwołuje się niniejszy rozdział, stanowiły one punkt wyjścia do doskonalenia rozwiązań w programach perspektywy 2014-2020. Ponadto, również bazując na doświadczeniach płynących z tych badań, dokonano oceny przydatności ewaluacji opartej na teorii jako formy wspierania procesu programowania. Na koniec sformułowano kilka kluczowych rekomendacji wynikających z ewaluacji dotyczących organizacji procesów wyboru i oceny projektów w ramach programów operacyjnych perspektywy 2014-2020.

<sup>196</sup> W latach 2015-2016 szeroki zespół badaczy, którego członkami byli autorzy niniejszego rozdziału, wykonał dwa badania systemów wyboru i oceny projektów w krajowych programach operacyjnych na lata 2014-2020 (był to Program Operacyjny Inteligentny Rozwój i Program Operacyjny Polska Cyfrowa – por. „Ewaluacja systemu wyboru projektów w POIR – etap I i II”, Konsorcjum IMAPP, Fundacja Idea Rozwoju, PAG Uniconsult i Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju, Warszawa 2016 i 2017, „Ewaluacja systemu wyboru projektów POPC”, Konsorcjum Fundacja Idea Rozwoju, IMAPP, PAG Uniconsult, Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju, Warszawa 2017) oraz badanie w makroregionalnym Programie Operacyjnym Polska Wschodnia na lata 2014-2020 (w tym przypadku ewaluacją systemu wyboru projektów objęto wszystkie instrumenty interwencji – działania i poddziałania – w ramach I osi priorytetowej POPW – por. „Ewaluacja systemu wyboru projektów w Programie Operacyjnym Polska Wschodnia 2014-2020”, Konsorcjum PAG Uniconsult i IMAPP na zamówienie Ministerstwa Rozwoju, Warszawa 2016). Każdorazowo w badaniach tych wykorzystywano podejście oparte na teorii jako punkt wyjścia do pogłębionej oceny w poszczególnych instrumentach interwencji, trafności, skuteczności i efektywności istniejącego systemu wyboru projektów (w szczególności innowacyjnych).

## Znaczenie teorii interwencji w systemie wyboru projektów

Każdy program interwencji publicznej powinien prowadzić do wywołania określonej zmiany w zakresie przedmiotu interwencji, odzwierciedlonej jej celem ogólnym i celami szczegółowymi. Interwencję można postrzegać jako pewną teorię zmiany, obrazującą sposób rozwiązania określonego problemu według założonej logiki postępowania. Podejście to jest szczególnie przydatne w programowaniu polityk wsparcia publicznego, gdyż umożliwia uchwycenie zależności pomiędzy celami i rezultatami przez odtworzenie logiki interwencji z uwzględnieniem rozmaitych założeń (również niewyeksponowanych wprost). Umożliwia ustalenie, dlaczego i jakie mechanizmy prowadzą (lub nie) do oczekiwanych rezultatów w danym kontekście ich realizacji<sup>197</sup>. W podejściu tym znaczenie ma nie tyle to, czy rezultaty interwencji udało się osiągnąć, ale przede wszystkim to, jakie mechanizmy i zależności przyczynowo-skutkowe legły u ich podstaw. Jak pokazują zebrane przez nas doświadczenia, posiadanie wiedzy o tych mechanizmach i zależnościach umożliwia doskonalenie programów interwencji publicznej, w szczególności zaś poprawę konstruowania i modyfikowanie rozwiązań w zakresie wyboru projektów, których dofinansowanie i wdrożenie ma prowadzić do realizacji założonych celów interwencji. Wprowadzane w oparciu o ewaluację zmiany mogą obejmować zarówno regulacje prawne kształtujące system wyboru, jak również zagadnienia merytoryczne i techniczne dotyczące budowy kryteriów oceny projektów.

Określona logika interwencji publicznej jest wynikiem zdiagnozowania konkretnej sytuacji problemowej i wynikających z niej wyzwań (przykładowo może to być niska innowacyjność gospodarki i konsekwencje tego stanu). Kształtują ją wyznaczone, mierzalne cele interwencji (np. wzrost innowacyjności i konkurencyjności gospodarki), określone działania służące osiągnięciu tych celów (np. instrumenty interwencji ukierunkowane na wspomaganie działań podejmowanych przez adresatów wsparcia) oraz przyjęte założenia dotyczące warunków realizacji oczekiwanego poziomu celów. Ujmując rzecz inaczej, logikę interwencji odzwierciedla następujący układ przyczynowo-skutkowy: (1) wdrożenie określonych działań składających się na interwencję publiczną, na które alokowano określone zasoby (np. środki finansowe), doprowadzi do pojawienia się efektów rzeczowych (konkretnego produktu interwencji, np. wyników badań przemysłowych), (2) wykorzystanie tych produktów skutkować będzie zmianą sytuacji beneficjentów interwencji (wystąpi rezultat bezpośredni, np. wzrost konkurencyjności będący skutkiem wdrożenia innowacji opracowanej w oparciu o uzyskane wyniki badań), co ostatecznie (3) wywoła określone efekty w szerszym otoczeniu społeczno-gospodarczym (chodzi tu o rozmaite rezultaty strategiczne odpowiadające celowi głównemu interwencji, np. pojawienie się nowych, wysokiej jakości miejsc pracy). Podejście takie opiera się na założeniu, że określone działanie wspierające, uzasadnione zidentyfikowaną sytuacją problemową lub potrzebą, jest warunkiem koniecz-

---

<sup>197</sup> M. Richie, „Theory Based Evaluation: A wealth of approaches and untapped potential”, Komisja Europejska 2012, s. 1.

nym doprowadzenia do oczekiwanej (pozytywnej) zmiany w gronie beneficjentów wsparcia publicznego, a zmiana ta w ostatecznym rozrachunku wpłynie pozytywnie na całą grupę docelową i szersze otoczenie społeczno-gospodarcze. Przedstawiony układ, uporządkowany według zależności przyczynowo-skutkowych, kształtuje logikę interwencji. W praktyce badawczej musi on jednak podlegać dalszej weryfikacji w celu wyjaśnienia, w jaki sposób i dlaczego poszczególne elementy składowe interwencji przyczynią się (lub nie) do osiągnięcia określonych rezultatów<sup>198</sup>.

Zaprezentowane powyżej rozumowanie uwidacznia się wyraźnie w przypadku interwencji w obszarze innowacyjności. Mianowicie niska innowacyjność sfery gospodarczej ma być redukowana przez dostarczenie przedsiębiorcom wsparcia finansowego zachęcającego do podejmowania aktywności badawczo-rozwojowych i zmniejszającego ryzyko niepowodzenia badań czy gospodarczego wykorzystania innowacyjnych rozwiązań tworzonych w oparciu o wyniki działalności badawczo-rozwojowej. Ostatecznie działania te, dzięki osiągnięciu określonej masy krytycznej, powinny podnieść poziom innowacyjności całej gospodarki. Oczywiście, tak przebiegające rozumowanie znacznie upraszcza cały problem organizacji i przebiegu interwencji, bowiem w praktyce związki przyczynowo-skutkowe wypełniające tę sekwencję okazują się dużo bardziej skomplikowane – zależą od szeregu czynników, które ponadto nie muszą być (i najczęściej nie są) jednoznaczne i trwałe, tj. dające się kontrolować lub przewidywać. Interwencja publiczna realizowana jest w zmiennym otoczeniu społecznym, gospodarczym i prawnym, ponadto sama wpływa na zachodzące w nim zmiany. Rządzące interwencją mechanizmy nie sprowadzają się wyłącznie do sekwencji „wkład – wynik” (*input – output*), ale obejmują także szereg przesłanek charakteryzujących otoczenie interwencji i czynników behawioralnych wpływających na podejmowane działania i dokonywane wybory<sup>199</sup>.

Jak pokazały wyniki prowadzonych przez nas badań nad systemami wyboru i oceny projektów w ramach programów operacyjnych, wykorzystanie podejścia opartego na teorii okazało się przydatne do ewaluacji i formułowania użytecznych wskazań w odniesieniu do konstrukcji systemów wyboru. Nie ulega wątpliwości, że jednym z kluczowych czynników sukcesu, warunkującym powodzenie interwencji publicznej (skuteczność w realizacji celu), jest prawidłowe zaprojektowanie oraz zapewnienie skutecznego działania całego mechanizmu wyboru projektów. Chodzi tu zatem o mechanizm selekcyjny gwarantujący wybór najlepszych projektów operacjonalizujących logikę interwencji. Oczywiście, zdefiniowanie pojęcia „najlepszy projekt” i odpowiedniego zestawu kryteriów selekcji umożliwiającego jego identyfikację nie jest zadaniem pro-

---

<sup>198</sup> Por. P. Stronkowski i in., „Raport końcowy – Ocena wsparcia w obszarze ekonomii społecznej udzielonego ze środków EFS w ramach POKL”, Coffey International Development na zamówienie Ministerstwa Infrastruktury i Rozwoju, Warszawa 2013, s. 15.

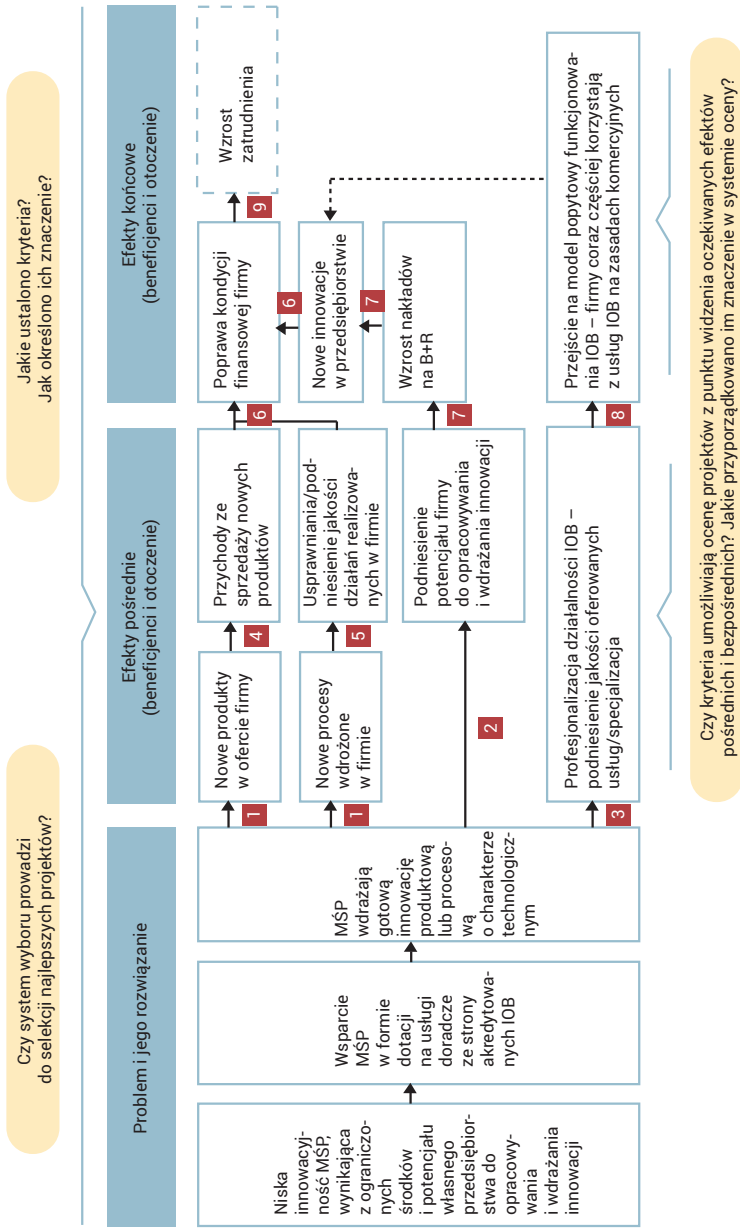
<sup>199</sup> Por. Komisja Europejska, *Evalsed Sourcebook: Method and Techniques*, Bruksela 2013, s. 51 [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/guide/evaluation\\_sourcebook.pdf](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf) (dostęp 24.03.2017 r.).

stym. Analiza systemu wyboru – a w jego ramach szczególnie podsystemu kryteriów oceny projektów – jest szczególnie ważna, bowiem stanowi krytyczny element teorii zmiany odpowiadającej zaprojektowanej interwencji.

W praktyce prowadzonych przez nas badań powyższa konstatacja oznaczała konieczność podjęcia próby odtworzenia teorii interwencji na różnych poziomach – zarówno programu, jak i poszczególnych instrumentów, tj. działań i poddziałań, w ramach których zorganizowany został system wsparcia publicznego (w tym system wyboru i oceny projektów). Było to konieczne, bowiem całościowa logika interwencji zazwyczaj nie bywa jednoznacznie uzewnętrzniona i nie jest komentowana wprost w dokumentach programowych – pozostaje ukryta pod celami interwencji, proponowanymi działaniami i schematami wdrażania. Zwykle bardzo ograniczone okazują się również możliwości pozyskania informacji na ten temat bezpośrednio od autora interwencji, którym najczęściej nie jest jeden ośrodek decyzyjny. Problemy te wynikają również z upływu czasu, a w związku z tym z niedostępności osób, które uczestniczyły w projektowaniu interwencji. Poza tym często okazuje się, iż nie są dostępne pogłębione opisy teorii interwencji w postaci sformalizowanych źródeł. Wszystko to powoduje, że odtworzenie całościowej teorii zmiany wywoływanej interwencją nie jest zadaniem łatwym, ale z drugiej strony jest konieczne, aby możliwe było analizowanie systemu wyboru i oceny projektów jako elementu kształtującego tę teorię oraz doskonalenie systemu pod kątem realizacji celów interwencji. Szczególne znaczenie ma to właśnie w przypadku badania kryteriów wyboru, których system powinien być tak zaprojektowany, aby umożliwiał wspieranie przedsięwzięć w możliwie najwyższym stopniu gwarantujących realizację celów interwencji. Badanie tego systemu powinno koncentrować się na ocenie, czy przyczyni się on (i w jakim stopniu) do realizacji celów działań wspierających. W praktyce może się okazać, że stosowane kryteria prowadzą do wyboru projektów, które nie gwarantują realizacji celów. Dopiero precyzyjne zidentyfikowanie takich sytuacji umożliwia formułowanie rekomendacji doskonalących system wyboru i kryteria oceny. Przykład zastosowanego podejścia, opartego na odtworzeniu teorii interwencji, w kontekście oceny dostosowania kryteriów wyboru projektów do celów interwencji, przedstawia poniższy schemat.



# Schemat 1. Uproszczony model logiczny Poddziałania 2.3.1 POIR – Proinnowacyjne usługi instytucji otoczenia biznesu dla MŚP – ustalony do celów badania systemu wyboru i oceny projektów



Źródło: Opracowanie własne na podstawie dokumentów i prac analitycznych zrealizowanych w ramach projektu „Ewaluacja systemu wyboru projektów w POIR – etap II”, Konsorcjum IMAPP, Fundacja Idea Rozwoju, PAG Uniconsult i Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju, Warszawa 2017

Skoro ewaluacja wspierana teorią zmierza do ustalenia, dlaczego występuje i jaki jest mechanizm zmiany oraz czy prowadzi on do oczekiwanych rezultatów, to system wyboru i oceny projektów stanowi jeden z elementów tego mechanizmu. Zatem analiza kryteriów wyboru i wynikające z niej wnioski mogą stanowić odpowiedź na pytanie, dlaczego interwencja działa (lub nie). W prowadzonych przez nas badaniach podejście oparte na teorii było wykorzystywane wyłącznie do zidentyfikowania założeń i oczekiwanych rezultatów wsparcia na poziomie efektów pośrednich i bezpośrednich (uwiadczenia to powyższy schemat). Był to bowiem minimalny warunek, który musiał być spełniony, aby można było ocenić prawidłowość skonstruowania kryteriów pod kątem ukierunkowania przez nie wyboru projektów w ramach analizowanej interwencji. Badanie głębszych zależności i czynników kształtujących interwencję (poza samymi kryteriami i formalnym kształtem systemu wyboru) nie było możliwe<sup>200</sup>. W konsekwencji odpowiedź na pytanie o znaczenie kryteriów (ich siłę oddziaływania) jako elementu składającego się na mechanizm zmiany, występujący w ramach danej interwencji, może być jedynie cząstkowa. Decydują o tym zarówno kwestie dotyczące realiów badawczych, jak i kompleksowość zagadnienia. Zauważmy, że już w samym systemie wyboru projektów występuje wiele czynników, które powodują, że system ten działa w określony sposób. W sumie mamy tu do czynienia z bardzo skomplikowanym układem różnorodnych czynników i występujących pomiędzy nimi zależności, z których wszystkie są bardzo trudno kontrolowalne<sup>201</sup>. Przykładowo istnienie odpowiednio zaprojektowanych kryteriów to warunek konieczny, lecz niewystarczający do realizacji celów interwencji, bowiem zawsze istotną kwestią jest sposób ich stosowania, np. przez ekspertów oceniających wnioski. W tej sytuacji wiele zależeć będzie od czasu przeznaczanego na ocenę, możliwości doboru ekspertów o określonych kwalifikacjach i doświadczeniu – to te (przykładowe) czynniki ostatecznie będą w znaczącym stopniu decydować o jakości przeprowadzanych ocen.

Ponadto trzeba również podkreślić, że rekomendacje ewaluacyjne wskazujące na potrzebę daleko idących zmian w kryteriach są z trudem akceptowane przez zamawiających badania, bowiem powodują one konieczność zaprojektowania nowych kryteriów, następnie ich formalnego przyjęcia i wdrożenia w proces oceny – w efekcie może to komplikować i opóźniać proces realizacji interwencji. Wyniki badań mogą również prowadzić do konkluzji podważającej sensowność stosowania danego instrumentu. Oczywiście, znaczenie tych kwestii rośnie w miarę zaangażowania we wdrażanie danego programu. Oznacza to konieczność odpowiedniego zaplanowania prac nad ewaluacją systemu wyboru i oceny projektów. Badania tego rodzaju prowadzone być powinny możliwie wcześnie, bowiem w sytuacji, gdy system został

---

<sup>200</sup> Decydowały o tym realia prowadzonych badań, w tym m.in. dotyczące czasu przeznaczanego na ich wykonanie, jak również dostępnego budżetu. Z podobną sytuacją mamy zapewne do czynienia w przypadku innych ewaluacji dotyczących systemów wyboru i oceny projektów.

<sup>201</sup> P. Rogers, „Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions, Evaluation”, SAGE Publications, Los Angeles 2008.

już zaprojektowany, a szczególnie, gdy jest już stosowany, gotowość do jego dalekich zmian jest mniejsza.

Ostatecznie nasze doświadczenia wskazują, że odwoływanie się do ewaluacji opartej na teorii w badaniach systemu wyboru i oceny projektów jest zabiegiem nieco na wyrost. Owszem, jest to przydatne – jak wspomnieliśmy wcześniej – do identyfikacji założeń i efektów interwencji pod kątem badania, czy dana postać systemu wyboru i układ kryteriów zapewnią wybór adekwatnych projektów. Jednocześnie jednak, z uwagi na uwarunkowania prowadzonych badań, analiza teorii interwencji w naszych badaniach musiała być ograniczona do pewnego minimum. Decydowały o tym zarówno cele zamawiających, jak i realia prowadzonych badań. Pozostając przy celach badań systemów wyboru i oceny projektów – nasze doświadczenia wskazują, że w ramach podejścia *theory-based evaluation* wystarczające jest ustalenie głównych elementów logiki interwencji publicznej (np. działań/poddziałań w programie operacyjnym). Stanowi to wystarczającą podstawę do oceny, czy system i kryteria wyboru rzeczywiście zwiększają prawdopodobieństwo wyboru projektów posiadających pożądane właściwości z punktu widzenia celów interwencji. W badaniach poszukiwanie odpowiedzi na to pytanie przekłada się ostatecznie na analizowanie wielu kwestii szczegółowych, obejmujących poszczególne elementy systemu wyboru, począwszy od zawartości dokumentacji konkursowej, a skończywszy na aspektach dotyczących kryteriów: 1) ich odpowiedniej liczby, uwzględniającej wszystkie kluczowe oczekiwania interwencji w sferze produktów i rezultatów, 2) odpowiedniego dopasowania merytorycznego kryteriów do logiki interwencji czy wreszcie 3) aspektów technicznych ich budowy, obejmujących sposób skalibrowania skal oceny i definicje poszczególnych kryteriów.

Ponadto z uwagi na zazwyczaj rozbudowane katalogi kryteriów zawsze istotnym etapem procesu badawczego jest weryfikacja ich spójności – chodzi tu o odpowiedź na pytanie, czy stanowią one zbiór kompletny i odpowiednio się uzupełniający (komplementarny), zapewniając w ten sposób kompleksowość oceny projektów. Jest to jeden z kluczowych elementów doskonalenia zestawów kryteriów, które w sytuacji spełnienia postulatu kompleksowości (*complexity*) stają się kluczowym czynnikiem operacjonalizującym teorię interwencji programu.

Inne zagadnienie stanowi kwestia funkcjonowania systemu wyboru projektów, dotycząca rzeczywistego sposobu działania kryteriów (praktycznych skutków ich zastosowania) – w jakim stopniu skupiają się one na ocenie prawidłowości przygotowania wniosku o wsparcie (np. jego kompletności, przejrzystości prezentacyjnej), a w jakim na (pożądaną) ocenę jakości i perspektywach rozwojowych przedsięwzięcia, które ma stanowić przedmiot wsparcia. Wskazać można również na badanie funkcji dyskryminacyjnej kryteriów, np. faworyzującej lub defaworyzującej niektórych wnioskodawców. Bez wątplenia, co potwierdziły nasze badania, czynniki określające praktyczne

funkcjonowanie kryteriów stają się szczególnie istotnym elementem teorii zmiany zawartej w interwencji publicznej.

## **Doświadczenia w zakresie kształtowania systemu wyboru i oceny projektów w perspektywie finansowej 2007-2013 z uwzględnieniem kwestii innowacyjności**

Pomimo względnie wielu lat wdrażania programów wsparcia publicznego w Polsce, finansowanych w ramach unijnej polityki spójności, wciąż widoczne jest występowanie licznych trudności związanych z kształtowaniem kryteriów wyboru projektów, które dobrze odpowiadałyby ustalonym w dokumentach programowych celom interwencji. Identyfikacja tych trudności ma szczególne znaczenie dla perspektywy wdrażania funduszy unijnych w latach 2014-2020. Funkcjonujące w niej rozwiązania legislacyjne i wytyczne wskazują wyraźnie na potrzebę dalszego silnego zorientowania interwencji na rezultaty. W tej sytuacji wnioski w tym zakresie, płynące z poprzedniego okresu programowania, stają się przydatne do identyfikowania czynników i mechanizmów kształtujących teorię zmiany nowych interwencji publicznych<sup>202</sup>.

Po pierwsze, w perspektywie 2007-2013 wyzwaniem było prawidłowe odzwierciedlenie logiki działań wspierających w kryteriach wyboru projektów<sup>203</sup>. Problemy w tym zakresie wynikały z nie zawsze oczywistych rozstrzygnięć zawieranych w dokumentach programowych, co utrudniało – poza wszystkim innym – precyzyjne odtworzenie logiki interwencji. Jak wynika z realizowanych w Polsce badań, problemy wywoływała często ograniczona precyzja opisu (ogólnikowość) założeń i celów interwencji, w tym brak krótkoterminowych planów działań stanowiących ich konkretyzację i aktualizację, które umożliwiałyby ustalenie logiki działań wspierających<sup>204</sup>. W rezultacie w procesach operacjonalizacji wsparcia pojawiały się tendencje do konkretyzowania tej logiki w ramach tworzonych kryteriów w systemie wyboru projektów, co było dyskusyjne, gdyż w działaniach tych przywiązywano uwagę głównie do zapewnienia zgodności formalnej z zapisami dokumentów programowych, a nie poszukiwano rozwiązań, które gwarantowałyby wybór projektów w największym stopniu realizujących cele programowe.

---

<sup>202</sup> W niniejszym rozdziale prezentujemy wybrane z nich, sformułowane przez zespół badawczy na początkowych etapach poszczególnych badań systemów wyboru i oceny projektów w POIR, POPC oraz POPW z lat 2015-2017 (wskazanych na wstępie rozdziału).

<sup>203</sup> 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, s. 19-20.

<sup>204</sup> K. Keler, S. Krupnik, P. Skórska, A. Strzebońska, „Metaewaluacja badań dotyczących oceny kryteriów wyboru projektów w programach operacyjnych współfinansowanych z funduszy europejskich w Polsce w perspektywie 2007-2013, Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju Regionalnego, Kraków 2011, s. 6.

W sferze wspierania innowacyjności gospodarki problemy w określeniu logiki interwencji wynikały również z braku spójnej polityki innowacyjnej na poziomie kraju<sup>205</sup>.

Po drugie, widoczną słabością działań zmierzających do odzwierciedlenia logiki interwencji w systemie wyboru była koncentracja na zapewnieniu wygody, redukcji obciążeń natury logistycznej i kosztowej w stosowaniu kryteriów, a nie na maksymalizacji ich dopasowania do celów programu. Problem ten przejawiał się szczególnie przy odzwierciedlaniu założeń interwencji dotyczących innowacyjności. System wyboru projektów stosowany do celów Programu Operacyjnego Innowacyjna Gospodarka 2007-2013 (POIG) przywiązywał większą wagę do łatwości oceny. W związku z tym, jak podkreślono to w jednym z opracowań ewaluacyjnych, nie wypracowano skutecznego mechanizmu selekcji bazującego na ocenie poziomu innowacyjności rezultatów wspieranych projektów<sup>206</sup>. Zauważalny był brak równowagi pomiędzy poziomem operacyjnym i strategicznym wdrażania programów interwencji publicznej. Procesy wyborów projektów nierzadko dyktowały przeważające potrzeby perspektywy operacyjnej, a nie potrzeby strategiczne<sup>207</sup>. Było to konsekwencją dążenia do terminowego wydatkowania środków. Poza tym koncentracja na efektach wsparcia jest zawsze bardziej wymagająca niż weryfikacja poprawności wniosku pod kątem poszczególnych kryteriów oceny.

Po trzecie, problemem o znaczeniu horyzontalnym w POIG było nadmierne skoncentrowanie form przekazywania wsparcia na instrumentach bezzwrotnych (dotacjach). Forma ta powinna występować rzadziej i dotyczyć wyłącznie wspierania przedsięwzięć o przełomowej innowacyjności. Natomiast w pozostałych przypadkach należy szerzej stosować mechanizmy zwrotne, które w przypadku finansowania przedsięwzięć innowacyjnych również mogą mieć sens<sup>208</sup>. Oczywiście, wybór formy finansowania (dotacja vs. instrument zwrotny) ma znaczenie dla kształtowania kryteriów oceny projektów. W przypadku instrumentów zwrotnych mogą mieć one inny, bardziej uproszczony charakter.

Po czwarte, często pojawiającym się problemem było formułowanie definicji pojęć stosowanych w kryteriach wyboru projektów w taki sposób, by mogły być one stosowane horyzontalnie (do celów każdego rodzaju interwencji), tj. skutkowały jednolitością interpretacji i poddawały się uniwersalnemu pomiarowi. Szczególnie silnie uwidaczniały się problematyczne kwestie dotyczące pojęcia innowacyjności i definio-

---

<sup>205</sup> T. Klimczak, A. Lis i in., „Ocena wpływu Programu Operacyjnego Innowacyjna Gospodarka na zwiększenie innowacyjności przedsiębiorstw”, WYG PSDB na zamówienie Ministerstwa Infrastruktury i Rozwoju, Warszawa 2014, s. 199.

<sup>206</sup> Tamże, s. 202.

<sup>207</sup> Por. K. Keler, S. Krupnik, P. Skórska, A. Strzebońska, *op.cit.*, s. 7.

<sup>208</sup> 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”, s. 51, w: J. Pokorski (red.), „Towards Innovative Economy – Effects of Grants to Enterprises in Poland”, Warszawa 2011.

wania różnych jej poziomów<sup>209</sup>, a także towarzyszące temu pomiary ryzyka projektów, które często faworyzowały pomysły znane, wymagające tylko dopracowania, a nie ryzykowne przedsięwzięcia nastawione na radykalny rozwój i tworzenie całkowicie nowych rozwiązań<sup>210</sup>. W konsekwencji system wyboru często prowadził do wspierania przedsięwzięć, których poziom innowacyjności nie był wysoki<sup>211</sup>.

Na koniec podkreślić należy, że szereg problemów dotyczył także aspektów technicznych konstruowania kryteriów oceny projektów. Niektóre związane z tym zagadnieniem sprawy mają pierwszorzędne znaczenie dla wyboru projektów, a więc także wpływają na stopień realizacji celów programu. Interesujące kwestie dotyczą tu liczby kryteriów w zestawach (często widoczne było nadmierne rozdrobnienie wprowadzające niespójność całego zestawu, powodujące niekiedy „zerowanie” się pewnych kryteriów), rodzajów i sposobów dokonywanych ocen projektów (ekspercka ocena indywidualna, oceny panelowe), jak i stosowanych sposobów punktowania w ramach kryteriów. Przykładowo wykorzystywane rozmaite skale oceny (dwuelementowe – zero-jedynkowe, wieloelementowe) były w różnym stopniu dostosowane do strony merytorycznej konkretnych kryteriów. Czasami jednak znacznie ograniczały elastyczność oceny, co utrudniało różnicowanie oceny jakości projektów, lub też pozostawiały nadmierne pole do interpretacji, wymagające dodefiniowania znaczenia kryteriów już w trakcie oceny.

Przedstawione tu doświadczenia z minionej perspektywy programowej 2007-2013 wskazują na występowanie wielu rozmaitych czynników kształtujących faktyczną logikę interwencji, najczęściej jednak zniekształcających ją. Szczególnie ważnym czynnikiem okazują się tworzone i stosowane w praktyce systemy wyboru i oceny projektów. Pojawiające się w tym zakresie ograniczenia powodują, że wsparcie często kierunkuje się na projekty, które jednak nie prowadzą do maksymalizacji celów interwencji. Oznacza to, że kształtowaniu systemów wyboru i oceny projektów, jako niezwykle ważnemu i bardzo praktycznemu elementowi operacjonalizującemu każdą interwencję publiczną, należy poświęcić szczególną uwagę, zarówno na etapie ewaluacji *ex ante*, jak i szczególnie na etapie ewaluacji badającej skutki zastosowania kryteriów wyboru (rezultaty wyłonionych na ich podstawie projektów), pod kątem dalszej optymalizacji całego systemu wyboru i oceny wspieranych przedsięwzięć.

---

<sup>209</sup> M. Gajewski, J. Szczucki i in., „Ocena stanu realizacji 3, 4, 5 i 6 Priorytetu POIG w połowie okresu programowania”, PAG Uniconsult na zamówienie Ministerstwa Rozwoju Regionalnego, Warszawa 2011, s. 8.

<sup>210</sup> Tamże, s. 126-127.

<sup>211</sup> P. Gorgol, J. Kotrasiński, A. Weremiuk, „Metaewaluacja wyników badań ewaluacyjnych Programu Operacyjnego Innowacyjna Gospodarka, 2007-2013”, Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju Regionalnego, Warszawa 2012, s. 6.

## **Wnioski z wykorzystania podejścia wspieranego teorią w ewaluacjach systemu wyboru projektów w perspektywie finansowej 2014-2020**

Dobrym przykładem zastosowania podejścia opartego na teorii są ewaluacje systemu wyboru projektów w perspektywie finansowej 2014-2020. W programach operacyjnych tej perspektywy wsparcie innowacyjności i konkurencyjności przedsiębiorstw przypisano do jednego programu krajowego (Program Operacyjny Inteligentny Rozwój), jednego programu makroregionalnego (Program Operacyjny Polska Wschodnia) oraz szesnastu programów regionalnych. Zgodnie z Wytycznymi w zakresie ewaluacji polityki spójności na lata 2014-2020<sup>212</sup> ewaluacja kryteriów i systemu wyboru projektów jest jednym z obligatoryjnych zagadnień, które powinny zostać uwzględnione w planach ewaluacji wszystkich programów operacyjnych. Instytucje zarządzające poszczególnymi programami zostały zobligowane do realizacji tego typu analiz najpóźniej w ciągu dwóch lat od przyjęcia finalnej wersji dokumentu programowego<sup>213</sup>.

Na podstawie dotychczasowego przebiegu i wyników ewaluacji systemów wyboru projektów w programach operacyjnych na lata 2014-2020 można pokusić się o pierwsze podsumowanie samego procesu ich realizacji oraz wypracowanych wniosków. Warto jednak podkreślić, że poniższe opinie formułowane są z perspektywy wykonawcy badań i w przyszłości bardzo cennym ich uzupełnieniem byłoby także poznanie perspektywy zamawiających. W pierwszej kolejności koncentrujemy się na analizie tych uwarunkowań ewaluacji systemu wyboru projektów, które w dotychczas zakończonych badaniach miały decydujące znaczenie dla jakości i użyteczności wypracowanych wniosków i rekomendacji<sup>214</sup>.

1. Położenie nacisku (silniejszego niż w ewaluacjach prowadzonych we wcześniejszych okresach programowania) na teorię interwencji należy ocenić zasadniczo pozytywnie. Warto jednak podkreślić, że ewaluacja systemu wyboru projektów najczęściej okazuje się pomocna w doprecyzowaniu i zoperacjonalizowaniu logiki interwencji danego instrumentu. W perspektywie finansowej 2014-2020 założenia wsparcia na poziomie dokumentów programowych (które były przedmiotem negocjacji z Komisją Europejską) formułowano zazwyczaj w sposób dość ogólny, szczególnie pozostawiając do rozstrzygnięcia w Szczegółowych Opisach Osi Priory-

---

<sup>212</sup> „Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020” (sekcja 3.1.5 *Zakres tematyczny planów ewaluacji*), Ministerstwo Infrastruktury i Rozwoju, Warszawa 2015.

<sup>213</sup> W praktyce ewaluacje POIR i POPW uruchomiono stosunkowo szybko (już w 2015 r.), z kolei w programach regionalnych w większości rozpoczęły się one dopiero w 2016 r. lub są planowane na 2017 r.

<sup>214</sup> Podsumowanie głównie na podstawie następujących opracowań: (1) „Ewaluacja systemu wyboru projektów w POIR – etap I i II”, Konsorcjum IMAPP, Fundacja Idea Rozwoju, PAG Uniconsult i Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju, Warszawa 2016 i 2017, (2) „Ewaluacja systemu wyboru projektów POPC”, Konsorcjum Fundacja Idea Rozwoju, IMAPP, PAG Uniconsult, Uniwersytet Jagielloński na zamówienie Ministerstwa Rozwoju, Warszawa 2017, (3) „Ewaluacja systemu wyboru projektów w Programie Operacyjnym Polska Wschodnia 2014-2020”, Konsorcjum PAG Uniconsult, IMAPP na zamówienie Ministerstwa Rozwoju, Warszawa 2016.

tetowych danego programu operacyjnego. Partycypacyjna formuła ewaluacji obejmowała najczęściej warsztaty, których jednym z głównych celów było odtworzenie logiki danego instrumentu wsparcia – w spotkaniach tego typu brali udział zarówno ewaluatorzy, jak i przedstawiciele instytucji zarządzających i pośredniczących. Dzięki warsztatom możliwe było doprecyzowanie oczekiwanych efektów interwencji oraz sformułowanie postulatów ewentualnych zmian w kryteriach lub innych elementach systemu wyboru projektów. W tym kontekście warto zaznaczyć, że optymalnym rozwiązaniem wydaje się rozpoczęcie ewaluacji systemu wyboru projektów możliwie jak najwcześniej, najlepiej przed uruchomieniem pierwszych konkursów, ponieważ tylko wtedy możliwe są jeszcze istotne korekty analizowanych instrumentów.

2. Niezależnie od wymienionej wcześniej funkcji korekcyjnej ewaluacja systemu wyboru projektów nie pozwala na poprawę kluczowych elementów źle zaprojektowanego instrumentu wsparcia. W niewielkiej liczbie przypadków przeprowadzone dotychczas analizy prowadziły do wniosku, że problemem jest nie tyle pojedyncze kryterium lub inny element systemu wyboru, ile bardziej fundamentalne założenia interwencji (np. źle dobrana forma pomocy, cel interwencji lub wręcz brak uzasadnienia interwencji publicznej). Niestety, radykalne i daleko idące rekomendacje wskazujące na potrzebę zmiany tych założeń nierzadko traktowane są przez instytucje z nadmierną rezerwą.
3. Szczególnie na etapie rozpoczynającym się przed uruchomieniem pierwszych konkursów analizy systemu wyboru projektów polegają nie tyle na realizacji klasycznej ewaluacji, co raczej na opracowaniu specjalistycznej ekspertyzy przy dość napiętym harmonogramie. To w praktyce oznacza, że ilościowe i jakościowe badania społeczne mogą być jedynie wspomagającym źródłem informacji, a podstawą formułowanych wniosków i rekomendacji muszą być wiedza i doświadczenie zespołu eksperckiego. Nie oznacza to jednak, że komponent badawczy powinien być pomijany w ewaluacjach systemu wyboru projektów wykorzystujących podejście oparte na teorii. W praktyce szczególnie użyteczne okazują się badania jakościowe (wywiady pogłębione, grupy fokusowe) z przedstawicielami instytucji zaangażowanych we wdrażanie interwencji oraz ekspertami – członkami komisji oceniających projekty.
4. Dla użyteczności wyników ewaluacji systemu wyboru projektów decydujące znaczenie ma jakość bieżącej współpracy pomiędzy wykonawcą i zamawiającym: faktyczne zaangażowanie, gotowość do wysłuchania argumentów drugiej strony oraz elastyczne dostosowanie prac do realiów wdrożeniowych programów (choć oczywiście wymiar wdrożeniowy nie powinien przysłaniać znacznie bardziej istotnej logiki interwencji przewidzianej w danym instrumencie).
5. W tym miejscu warto też zaznaczyć, że ewaluacje systemu wyboru projektów wykorzystujące podejście oparte na teorii należą do specyficznej grupy badań, w których raporty końcowe są relatywnie mniej istotnym efektem prac. Dużo większe



znaczenie ma wspomniana wcześniej bieżąca współpraca zamawiającego z wykonawcą, obejmująca robocze spotkania, warsztaty oraz przekazywane (formalnie lub nieformalnie) opinie robocze. W ten sposób możliwe jest ścieranie się poglądów oraz wspólne wypracowywanie optymalnych (z punktu widzenia przyjętej logiki interwencji) i możliwych do wdrożenia rozwiązań. Same raporty z badań powstają z założenia ze stosunkowo dużym opóźnieniem, kiedy najważniejsze rozstrzygnięcia dotyczące procesu wyboru projektów zostały już przesądzone. Niemniej wartościowe jest formułowanie względnie dużej liczby wniosków i rekomendacji, oczywiście z zastrzeżeniem ich odpowiedniego uzasadnienia i uwzględnienia realiów wdrożeniowych. Nadmierna selektywność oraz dążenie do stuprocentowej zgodności obu stron może prowadzić do wyeliminowania twórczego myślenia i przedwczesnej rezygnacji z pomysłów mających potencjalnie szanse na poprawę jakości interwencji w przyszłości.

6. Użyteczność ewaluacji systemu wyboru projektów wymaga zaangażowania nie tylko samego zamawiającego (którym najczęściej jest jednostka ewaluacyjna instytucji zarządzającej), ale także wszystkich pozostałych podmiotów odpowiedzialnych za wdrażanie interwencji: osób odpowiedzialnych za programowanie w ramach instytucji zarządzającej oraz odpowiednich komórek instytucji pośredniczących i wdrażających. Zaangażowanie tych podmiotów jest konieczne nie tylko na etapie odbioru wyników prac, ale także w ich trakcie. Wydaje się szczególnie istotne, żeby eksperci biorący udział w ewaluacji mieli systematyczny dostęp do informacji o postępach prac związanych z procedowanymi zmianami ocenianych instrumentów, co ma szczególnie duże znaczenie dla adekwatności formułowanych wniosków i rekomendacji. Idealnym (i możliwym do wdrożenia w praktyce) rozwiązaniem jest także bezpośrednie dopuszczenie ewaluatorów do procesu wyboru projektów jako obserwatorów (np. do pracy paneli eksperckich, które bezpośrednio spotykają się z wnioskodawcami).

W kontekście wsparcia innowacyjności i przedsiębiorczości na podstawie zakończonych już ewaluacji systemu wyboru projektów w krajowych programach operacyjnych na lata 2014-2020 można wstępnie wyróżnić kilka kluczowych wniosków i rekomendacji, które należy traktować jako hipotezy wymagające weryfikacji w ramach ewaluacji *ex post*.

- W przypadku bardziej skomplikowanych instrumentów wsparcia, szczególnie obejmujących realizację prac badawczo-rozwojowych lub wdrażanie ich wyników, optymalną formą oceny wydaje się panel ekspertów dający możliwość bezpośredniego kontaktu z wnioskodawcą. Taka formuła pozwala na weryfikację kluczowych elementów dokumentacji projektowej oraz szczegółową dyskusję na temat technologicznych i biznesowych aspektów projektów.
- Użyteczność paneli oraz ich zdolność do identyfikacji najlepszych przedsięwzięć zależy z kolei od tego, na ile instytucjom odpowiedzialnym za wybór projektów uda

się zaangażować kompetentne osoby oraz dobrze zorganizować ich prace. Praktyka wdrażania programów operacyjnych na lata 2014-2020 pokazuje, że liczba potencjalnych ekspertów jest ograniczona, co powinno być brane pod uwagę przy pracy nad harmonogramami ogłaszania konkursów. Szczególnie istotne jest także usprawnienie wymiany wiedzy pomiędzy ekspertami.

- Praktyka perspektywy finansowej 2007-2013 oraz 2014-2020 pokazuje, że w obszarze wsparcia konkurencyjności i innowacyjności przedsiębiorstw zasadniczo nie sprawdza się nadmierna parametryzacja kryteriów wyboru projektów, szczególnie w odniesieniu do zagadnień z założenia subiektywnych lub wymagających eksperckiej oceny w zakresie danego pomysłu biznesowego (innowacyjność produktów, potencjał rynkowy projektu itp.). Precyzyjne określenie tego, za co przyznawane są punkty, powoduje przygotowywanie wniosków „pod kryteria”, jednocześnie utrudniając ekspertom całościową ocenę przedsięwzięcia z punktu widzenia zamierzonych efektów interwencji.
- Na przestrzeni ostatnich lat udało się znacznie poprawić przyjazność i przejrzystość procedury wyboru projektów z punktu widzenia aplikujących o wsparcie przedsiębiorstw, szczególnie w zakresie dostępności odpowiednich materiałów oraz jakości i funkcjonalności systemów informatycznych. Obecnie wydaje się, że wartym wzmocnienia i dopracowania aspektem jest przededefiniowanie roli i formuły działań informacyjno-promocyjnych, których zadaniem powinno być nie tylko przekazywanie podstawowych informacji o ogłaszanych konkursach, ale także aktywne wspieranie wnioskodawców w przygotowywaniu dobrej jakości wniosków (np. przez umożliwienie przedsiębiorcom odbycia niezobowiązujących konsultacji elementów wniosku o dofinansowanie z ekspertami – członkami komisji oceniających projekty).

## Podsumowanie

Jak zaprezentowano w pierwszej części niniejszego rozdziału, odtworzenie i uwzględnienie różnych aspektów teorii interwencji w ewaluacji systemu wyboru i oceny projektów może podnieść skuteczność działań finansowanych ze środków publicznych. Trzeba jednak pamiętać, że rezultaty analiz bazujących na ewaluacji opartej na teorii, choć niewątpliwie przydatne na etapie identyfikacji założeń i rezultatów interwencji pod kątem dostosowania do nich systemów wyboru i oceny projektów, są często niwelowane ograniczoną gotowością do wprowadzania dalej idących zmian w funkcjonujących już systemach. Ograniczenia wynikają także z realiów prowadzenia ewaluacji systemów i kryteriów oceny (szczególnie dostępnego czasu i budżetu na badania, a w konsekwencji – możliwego ich zakresu). Oznacza to, że praktyczne wykorzystanie tego typu badań wymaga odpowiedniego określenia ich roli oraz właściwego zaprojektowania i powiązania z procesem programowania i wdrażania interwencji. Z przeprowadzonych ewaluacji systemów wyboru projektów dotychczas wynika, że podejście zastosowane w programach perspektywy finansowej 2014-2020 zasadniczo

wychodzi naprzeciw tym oczekiwaniom. Najważniejszym czynnikiem sukcesu tego typu badań jest sprawna komunikacja pomiędzy ekspertami-wykonawcami badania oraz wszystkimi instytucjami odpowiedzialnymi za wdrażanie programu wsparcia. Współpraca obu stron musi polegać na stałej wymianie opinii i pomysłów na temat ocenianych instrumentów, dalece wykraczając tym samym poza standardowe ramy ewaluacji, których głównym obszarem współpracy jest najczęściej jedynie konsultacja raportu końcowego z badania. Ewaluacje systemu wyboru projektów w programach operacyjnych na lata 2014-2020 jednoznacznie wskazują na to, że – nawet przy bardzo ograniczonym czasie na realizację analizy – wskazana jest najsilniejsza ich koordynacja z procesem programowania. Bez niej wykorzystanie podejścia opartego na teorii jest znacznie trudniejsze do zastosowania w praktyce i nie daje szansy na faktyczne przełożenie procesu ewaluacji na poprawę jakości badanych instrumentów wsparcia.

## O autorach

### Maciej Gajewski

---

Doktor nauk ekonomicznych, wykładowca na Wydziale Zarządzania Uniwersytetu Warszawskiego. Posiada ponad 25 letnie doświadczenie w świadczeniu usług badawczych (ewaluacyjnych), doradczych i szkoleniowych. Jest współnikiem i ekspertem w firmie doradczej Policy & Action Group Uniconsult. Specjalizuje się w doradztwie dotyczącym tworzenia i funkcjonowania instrumentów finansowych, służących zwiększaniu dostępności sektora mikro, małych i średnich przedsiębiorstw do zewnętrznych źródeł finansowania, w szczególności instrumentów pożyczkowych, poręczeniowych i kapitałowych. W dziedzinie jego profesjonalnej działalności mieszczą się również zagadnienia wspierania finansowego przedsięwzięć innowacyjnych, we wszystkich fazach ich rozwoju. Posiada wieloletnie doświadczenie w projektowaniu i prowadzeniu badań oraz zarządzaniu zespołami eksperckimi, zajmującymi się ewaluacją interwencji publicznych. Jest ewaluatorem i ekspertem w zakresie tworzenia i wdrażania programów operacyjnych oraz realizowanych w ich ramach działań, finansowanych ze środków UE – zarówno na poziomie ogólnopolskim, jak i regionalnym. Jest autorem lub współautorem wielu publikacji prezentujących wyniki ewaluacji programów interwencji publicznej.

### Andrzej Regulski

---

Od 2013 r. Dyrektor Zarządzający spółki IMAPP, w przeszłości Dyrektor Programu Ewaluacji Polityk Publicznych w Instytucie Badań Strukturalnych. Z wykształcenia ekonomista, absolwent Szkoły Głównej Handlowej w Warszawie. Dotychczas brał udział (w większości jako koordynator) w ponad 50 badaniach ewaluacyjnych i innych projektach analitycznych, realizowanych na zlecenie administracji rządowej i samorządowej. Specjalizuje się w zagadnieniach związanych z systemem instytucjonalnym polityki rozwoju w Polsce, innowacyjnością oraz energetyką. W ostatnich latach koordynował m.in. ewaluacje *ex ante* Umowy Partnerstwa oraz czterech programów operacyjnych na lata 2014-2020 (Program Operacyjny Infrastruktura i Środowisko, regionalne programy operacyjne woj. łódzkiego, podkarpackiego i pomorskiego), badania związane z aktualizacjami strategii rozwoju województw (małopolskiego, podkarpackiego, pomorskiego oraz warmińsko-mazurskiego), a także szereg analiz dotyczących budowy systemu wskaźników w dokumentach strategicznych i operacyjnych na poziomie krajowym i regionalnym.

## **Bibliografia:**

1. Gajewski M., Szczucki J. i in., *Ocena stanu realizacji 3,4,5 i 6 Priorytetu PO IG w połowie okresu programowania*, PAG Uniconsult, Warszawa 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", w: Pokorski J. (red.), *Towards Innovative Economy - Effects of Grants to Enterprises in Poland*, Warszawa 2011.
3. Gorgol P., Kotrański J., Weremiuk A., *Metaewaluacja wyników badań ewaluacyjnych Programu Operacyjnego Innowacyjna Gospodarka, 2007-2013*, Warszawa 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*, Uniwersytet Jagielloński, Kraków 2011.
6. Klimczak T., Lis A. i in., *Ocena wpływu Programu Operacyjnego Innowacyjna Gospodarka na zwiększenie innowacyjności przedsiębiorstw*, WYG PSDB, Warszawa 2014.
7. Richie M., *Theory Based Evaluation: A wealth of approaches and untapped potential*, European Commission 2012.
8. Stronkowski P., Andrzejewska M. i in., *Raport końcowy - Ocena wsparcia w obszarze ekonomii społecznej udzielonego ze środków EFS w ramach PO KL*, Coffey International Development, Warszawa 2013.
9. Rogers P., "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions", *Evaluation*, SAGE Publications, Los Angeles, 2008.

## **Inne:**

1. *Ewaluacja systemu wyboru projektów w POIR - etap I i II*, IMAPP, IDEA of Development Foundation, PAG Uniconsult, Uniwersytet Jagielloński, Warszawa 2016 and 2017,
2. *Ewaluacja systemu wyboru projektów POPC*, IDEA of Development Foundation, PAG Uniconsult, Uniwersytet Jagielloński, IMAPP, Warszawa 2017.
3. *Ewaluacja systemu wyboru projektów w Programie Operacyjnym Polska Wschodnia 2014-2020*, PAG Uniconsult, IMAPP, Warszawa 2016.
4. Komisja Europejska *Evalsed Sourcebook: Method and Techniques*, Bruksela 2013.
5. *Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020*, Ministerstwo Infrastruktury i Rozwoju, Warszawa 2015.



## Doświadczenia w wykorzystaniu ewaluacji opartej na teorii na poziomie regionalnym

### Wprowadzenie

Polska administracja od chwili akcesji do Unii Europejskiej (UE) skonfrontowana została z wymaganiami obowiązującymi w Unii, w tym dotyczącymi organizacji systemu ewaluacji programów współfinansowanych ze środków unijnych. Różnica doświadczeń i umiejętności w tym zakresie sprawiła, że konieczne stało się m.in. przyspieszone nadrobienie luki w potencjale ewaluacyjnym instytucji odpowiadających za redystrybucję środków unijnych. O ile pierwsze dwie perspektywy wykorzystania funduszy europejskich w Polsce (2004-2006, 2007-2013) posłużyły budowaniu odpowiedniego dla nich krajowego systemu ewaluacji (i jego stopniowej decentralizacji), o tyle w perspektywie 2014-2020 do poprzednich wyzwań doszły kolejne – m.in. związane z rygorystycznymi wymaganiami metodologicznymi ewaluacji, które mają „na celu uzyskanie bardziej wiarygodnych i użytecznych informacji o rzeczywistych efektach polityki spójności”, tj. rozwojem i stosowaniem „zaawansowanych metod badawczych, w tym w szczególności metod (...) ewaluacji opartej na teorii<sup>215</sup>”.

Celem niniejszego rozdziału jest przedstawienie pierwszych doświadczeń prowadzenia ewaluacji opartej na teorii oraz wynikających z tego faktu wyzwań, z jakimi musiała sobie poradzić polska administracja samorządowa. Analizę przeprowadzono na podstawie 15 badań *ex ante* regionalnych programów operacyjnych na lata 2014-2020. W artykule wskazano także generalne trudności w budowaniu potencjału ewaluacyjnego w regionach, a także te bezpośrednio wynikające ze stosowania ewaluacji opartej na teorii. W tym też układzie uporządkowane zostały główne wątki niniejszego tekstu, bazującego na przeglądzie literatury i badań ewaluacyjnych prowadzonych na poziomie regionalnym w Polsce.

---

<sup>215</sup> Por. Wytyczne Ministra Infrastruktury i Rozwoju z 22 września 2015 r. w zakresie ewaluacji polityki spójności na lata 2014-2020, podrozdział 1.2, pkt 4) s. 8 (<http://www.funduszeuropejskie.gov.pl/stroiny/o-funduszach/dokumenty/wytyczne-w-zakresie-ewaluacji-polityki-spojnosci-na-lata-2014-2020> (dostęp 28.04.2017 r.).

## Ewaluacja w regionie

Trzy perspektywy programowania funduszy europejskich, w jakich uczestniczą regiony od akcesji Polski do UE, wyznaczają kolejne kamienie milowe rozwoju systemu ewaluacji na szczeblu centralnym i regionalnym<sup>216</sup>. W okresie 2004-2006 cały system ewaluacji zaplanowany został i był wdrażany przez instytucje zarządzające umiejscowione na poziomie administracji centralnej. Od 2007 r. w ramach wprowadzonej decentralizacji systemu zarządzania polityką spójności po raz pierwszy polskie województwa wraz z przygotowaniem regionalnych programów operacyjnych (RPO) stały się odpowiedzialne za budowanie systemu ewaluacji na poziomie regionalnym<sup>217</sup>. O ile można uznać, iż na poziomie centralnym system ewaluacji, w tym szeroko rozumiana kultura ewaluacyjna, w roku 2007 był już ugruntowany, o tyle regiony dopiero wraz z procesem wdrażania RPO 2007-2013 zaczęły się uczyć prowadzenia ewaluacji i rozwijać kulturę ewaluacji na szczeblach administracji samorządowej. Wówczas w szesnastu województwach powstawały wyspecjalizowane zespoły organizacyjne – jednostki ewaluacyjne (JE), powoływano grupy sterujące ewaluacją oraz uruchamiano pierwsze badania ewaluacyjne. Można więc stwierdzić, że podwaliny systemu ewaluacji położono oraz szczególną pracę nad rozwojem potencjału<sup>218</sup> i kultury ewaluacyjnej na poziomie samorządów województw wykonano w okresie programowania funduszy europejskich na lata 2007-2013.

Perspektywa finansowa 2014-2020 przynosi regionom kolejne wyzwania ze względu na nowe wymagania dotyczące sposobu planowania i prowadzenia badań ewaluacyjnych. Komisja Europejska (KE), nakładając na państwa członkowskie bardziej wymagające przepisy dotyczące systemu ewaluacji programów operacyjnych, wymusiła

---

<sup>216</sup> Por. J. Bachtler: „Ewaluacja regionalnej polityki w Europie: kultura, zaangażowanie i potencjał”, w: *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, pod red. K. Olejniczaka, M. Kozaka, B. Ledziona, Wydawnictwa Akademickie i Profesjonalne. Akademia Leona Koźmińskiego, Warszawa 2008, s. 60-63.

<sup>217</sup> Interwencje publiczne, zarządzane i wdrażane w regionach, a w szczególności Regionalne Programy Operacyjne oraz komponent regionalny Programu Operacyjnego Kapitał Ludzki, podlegały bezpośredniej analizie i ocenie przez poszczególne samorzady województw (Por. Wytyczne nr 6 Ministra Rozwoju Regionalnego z 30 maja 2007 r. w zakresie ewaluacji programów operacyjnych na lata 2007-2013).

<sup>218</sup> Potencjał ewaluacji rozumieć można jako zdolność do prowadzenia wysokiej jakości badań ewaluacyjnych, których wyniki są absorbowane i wykorzystywane w celu poprawy jakości zarządzania publicznego. Źródłem tych zdolności są zarówno jednostki ewaluacyjne, wykonawcze, decydenci, odbiorcy, jak i kapitał społeczny. Por. M. Szałaj, *Budowa potencjału ewaluacyjnego*, w: *Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka*. Praca zbiorowa, Urząd Marszałkowski Województwa Opolskiego, Opole 2009, s. 76.



dalszy rozwój potencjału administracyjnego<sup>219</sup> (w tym jednostek ewaluacyjnych na poziomie regionalnym) oraz nowe podejście do planowania i prowadzenia badań ewaluacyjnych w Polsce. KE po raz pierwszy w bardzo szczegółowy sposób określiła zakres i tryb przygotowania oraz zatwierdzania Planów ewaluacyjnych dla programów operacyjnych (PO)<sup>220</sup>. Wzrost roli ewaluacji w systemie zarządzania PO (w tym także Regionalnymi Programami Operacyjnymi) łączy się przede wszystkim z potrzebą poprawy jakości (metodologii) realizowanych badań ewaluacyjnych oraz zwiększenia stopnia wykorzystania ich wyników<sup>221</sup>. Zwiększenie wymagań metodologicznych ma na celu uzyskanie bardziej użytecznych informacji o rzeczywistych efektach polityki. Głównym zadaniem dla ewaluacji w regionie staje się wskazanie jaka część zmiany sytuacji społeczno-gospodarczej (np. zmiany wartości wskaźnika rezultatu) zaistniała dzięki realizacji polityki spójności. Drugim - równie istotnym - jest wyjaśnienie mechanizmów oddziaływania polityki spójności, czyli uzyskanie odpowiedzi na pytanie dlaczego dana zmiana nastąpiła. Kluczowy staje się w tych warunkach wybór odpowiedniego podejścia metodologicznego, które zapewniłoby wysoką jakość procesu ewaluacji. Wytyczne wskazują bezpośrednio na konieczność stosowania zaawansowanych metod badawczych, w tym w szczególności metod kontrfaktycznych oraz wy-

---

<sup>219</sup> Wg stanu na koniec 2015 r. w ramach systemu ewaluacji polityki spójności w Polsce aktywnie działały 34 jednostki ewaluacyjne ulokowane w instytucjach zarządzających, pośredniczących, pośredniczących II st. oraz wdrażających. W ich ramach ewaluacją zajmowało się łącznie 158 osób, z czego 70% to pracownicy jednostek regionalnych. Z badań Ministerstwa Rozwoju wynika, że liczba pracowników jednostek ewaluacyjnych w latach 2013-2015 uległa znacznemu spadkowi (w 2014 r. było to 188 osób, w 2013 – 196). Częściowo wyjaśnia się to zmianami struktury jednostek ewaluacyjnych na poziomie regionalnym i utworzeniem wspólnych komórek odpowiedzialnych za ewaluację wsparcia Europejskiego Funduszu Rozwoju Regionalnego i Europejskiego Funduszu Społecznego, udzielanego w ramach Regionalnych Programów Operacyjnych 2014-2020 (Por. *Rocznik Ewaluacji 2013*, Ministerstwo Rozwoju, Warszawa 2014, s. 8-9; *Rocznik Ewaluacji 2015*, Ministerstwo Rozwoju, Warszawa 2016, s. 6).

<sup>220</sup> Podstawowy dokument wyznaczający na poziomie państw członkowskich kształt systemu ewaluacji wsparcia Europejskiego Funduszu Rozwoju Regionalnego i Funduszu Spójności na lata 2014-2020 to *Guidance Document on Monitoring and Evaluation. European Regional Development Fund and Cohesion Fund*, przygotowany przez Dyрекcyję Generalną ds. Polityki Regionalnej i Miejskiej w 2014 r. [http://ec.europa.eu/regional\\_policy/sources/docoffic/2014/working/wd\\_2014\\_en.pdf](http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf) (dostęp 19.03.2017 r.). Dla pozostałych funduszy (np. Europejskiego Funduszu Społecznego) czy innych mechanizmów finansowania (np. programów pomocowych) dodatkowo powstały odrębne wytyczne horyzontalne poszczególnych dyrekcji generalnych Komisji Europejskiej (np. *Commission Staff Working Document Common methodology for state aid evaluation*, przygotowany przez DG ds. Konkurencji). Na poziomie krajowym uszczegółowienie wskazań KE dotyczących ewaluacji w nowym okresie programowania zawierają Wytyczne Ministra Infrastruktury i Rozwoju z 22 września 2015 r. w zakresie ewaluacji polityki spójności na lata 2014-2020.

<sup>221</sup> W samych Wytycznych Ministra Infrastruktury i Rozwoju w podrozdziale 1.2 zapisano: „W okresie 2014-2020 następuje zwiększenie roli ewaluacji w procesie zarządzania PO wdrażanymi w ramach polityki spójności. Znajduje to odzwierciedlenie w formalnych zapisach aktów prawnych stanowiących podstawę do wdrażania funduszy unijnych oraz wytycznych formułowanych przez KE. Zwiększenie znaczenia ewaluacji w procesie wdrażania PO nakłada także dodatkowe obowiązki na państwa członkowskie UE, a tym samym powoduje konieczność reorganizacji i adaptacji systemów ewaluacji do nowych wymagań” (Wytyczne Ministra Infrastruktury i Rozwoju z 22 września 2015 r. w zakresie ewaluacji polityki spójności na lata 2014-2020).

korzystania podejścia ewaluacji opartej na teorii<sup>222</sup>. Konsekwentne stosowanie tego podejścia pozwala decydom lepiej zrozumieć regionalne uwarunkowania powodzenia projektowanych interwencji, a w dalszej kolejności poprawiać jakość rządzenia. Odwołanie się do teorii opisujących mechanizmy i kierunki rozwoju regionu ułatwia zestawianie w procesie ewaluacji wybranych działań z alternatywnymi, które również mogą być zasadne w świetle aktualnego stanu wiedzy oraz doświadczeń (zwłaszcza regionalnych). Dzięki temu można szybciej identyfikować mechanizmy nieefektywne, precyzyjniej wyjaśniać przyczyny ich niesprawności i trwale je eliminować, zastępując lepszymi rozwiązaniami. Głównym wyzwaniem wydaje się przy tym przejście od ogólnych teorii rozwoju regionalnego do ich regionalnej adaptacji, budowanej poprzez coraz lepsze rozpoznanie lokalnych uwarunkowań i powiązanie ich z zamierzonymi działaniami oraz oczekiwanymi rezultatami. Takie podejście wymaga dobrego przygotowania metodycznego, woli zaangażowania się w szerokie poszukiwania najbardziej adekwatnych teorii wyjaśniających mechanizmy osiągania rezultatów w skali regionalnej, dyscypliny i konsekwencji w realizacji przyjętych interwencji, a przede wszystkim czasu i otwartej, partnerskiej postawy wszystkich uczestników procesów projektowania i ewaluacji programów publicznych.

## **Znaczenie logiki programu w kontekście ewaluacji zorientowanych na teorię**

Dzięki promowaniu przez KE w perspektywie 2014-2020 ewaluacji opartej na teorii decydenci są mobilizowani do skupienia się na zależnościach przyczynowo-skutkowych uruchamianych w trakcie realizowania interwencji publicznych. Można je opisywać, używając metafory „czarnej skrzynki” programu, kryjącej mechanizm zmiany sytuacji problemowej zidentyfikowanej „na wejściu” w rezultat osiągnany „na wyjściu”.

Kluczowym etapem takich ewaluacji jest identyfikacja przez ewaluatora założeń leżących u podstaw zmiany, którą zamierza się osiągnąć poprzez daną interwencję publiczną<sup>223</sup>. Zależnie od koncepcji ewaluacji zorientowanej na teorię<sup>224</sup> różne są przekonania odnośnie do zawartości „czarnej skrzynki”, a w efekcie także zadania ewaluatora

---

<sup>222</sup> Por. Wytyczne Ministra Infrastruktury i Rozwoju z 22 września 2015 r. w zakresie ewaluacji polityki spójności na lata 2014-2020, podrozdział 1.2 pkt 4, s. 8.

<sup>223</sup> A. Markiewicz, I. Patrick, *Developing Monitoring and Evaluation Frameworks*, Sage Publications Inc. 2016, s. 35.

<sup>224</sup> Więcej w niniejszej publikacji, w rozdziale pt. „Przegląd systematyczny koncepcji ewaluacji opartej na teorii” autorstwa Katarzyny Hermann-Pawłowskiej i Pauliny Skórskiej.

w procesie identyfikacji podstaw koncepcyjnych logiki programu<sup>225</sup>. Co do zasady proces ten wymaga od ewaluatora pracy z interesariuszami nad lepszym zrozumieniem tego, w jaki sposób program ma doprowadzić do rozwiązania problemu. Ewaluator ułatwia decydom dookreślenie wątpliwych obszarów teorii zmiany (niepotwierdzonych związków przyczynowo-skutkowych) oraz dobranie odpowiednich metod badawczych, które pozwolą na ich wypełnienie.

Przydatne jest rozróżnienie w TBE pojęcia teorii programu (*programmatic theory*) o charakterze koncepcyjnym oraz bardziej operacyjnej logiki interwencji (teorii zmiany – *implementation theory*), które stanowią wzajemnie powiązane, lecz odmienne fazy budowy polityk publicznych<sup>226</sup>. Prezentację graficzną teorii programu i logiki interwencji stanowi zazwyczaj diagram obrazujący powiązania przyczynowo-skutkowe mechanizmów prowadzących do zmiany (zob. rys. 1).

Podczas ewaluacji programu badaniem powinny zostać objęte zarówno teoria programu, jak i logika interwencji. Dla teorii programu należy zweryfikować adekwatność (*plausibility*) projektu interwencji i ustalić jego składowe, ponieważ w TBE zakłada

---

<sup>225</sup> Przykładowo założeniem *theory-driven evaluation* (TDE) wg N. Stame jest brak teorii towarzyszący zazwyczaj interwencjom publicznym (pusta „czarna skrzynka”). Zadaniem ewaluatora jest zatem wypracowanie zestawu teorii, jakie mogą wyjaśniać osiągnięcie zakładanych celów, czyli odpowiedź, co może zadziałać w danej interwencji, czerpiąca z dorobku nauk społecznych. Realizuje się to poprzez analizę zaproponowanych mechanizmów interwencji, wymianę poglądów nt. oczekiwanych rezultatów między interesariuszami a ewaluatorami oraz badanie, dlaczego i jak program działa, w nawiązaniu do istniejących teorii normatywnych i przyczynowo-skutkowych. W podejściu *realistic evaluation* teorie prowadzą do założonych rezultatów dzięki zaangażowaniu interesariuszy w ich wdrożenie. „Czarna skrzynka” jest w tym przypadku wypełniona ludźmi, którzy wykorzystują szanse tworzone przez daną interwencję, osiągając rezultaty (O – *outcomes*) poprzez uruchomienie mechanizmu zmiany (M), który powinien zadziałać w danym kontekście (C – *context*). Podstawą ewaluacji jest zatem odtworzenie układu CMO, czyli ustalenie, jak dany mechanizm może działać w danym kontekście, a także uzyskanie od interesariuszy interwencji odpowiedzi na pytanie, jak on zadziałał z ich perspektywy oraz jakich dowodów na potwierdzenie jego uruchomienia mogą dostarczyć. Otoczenie programu jest rozpatrywane w tym podejściu przede wszystkim w kontekście społecznym, a proces ewaluacji interwencji publicznej sprowadza się do oceny dowodów (*adjudication*) potwierdzających jej faktyczny przebieg i osiągnięte rezultaty. W przypadku *theory-based evaluation* (TBE) „czarna skrzynka” jest wypełniona teoriami nazywanymi teoriami zmiany, sformułowanymi niezależnie przez każdego z interesariuszy, zatem często więcej niż jedną dla danego programu, które miewają charakter założeń, milcząco przyjmowanych „oczywistości”. W trakcie TBE należy ujawnić mechanizmy zmiany leżące u podstaw osiągnięcia oczekiwanych rezultatów interwencji. Z jednej strony wymaga to porozumienia się odnośnie tego, która z przyjętych przez poszczególnych interesariuszy teorii zmiany zasługuje na faktyczne testowanie poprzez włączenie do programu interwencji (ewaluacja *ex ante*). Z drugiej strony TBE pozwala na sprawdzenie skuteczności uruchomionych mechanizmów dzięki ocenie przeprowadzonych interwencji (ewaluacja *ex post*), co przyczynia się do budowania nowych teorii zmiany. Por. N. Stame, *Theory-based Evaluation and Types of Complexity*, Evaluation, 2004, vol. 10 (1), s. 58-76, s. 61, 63.

<sup>226</sup> Teorię programu możemy określić wyjaśnienie, w jaki sposób i z jakich powodów uznano, iż uruchomienie programu doprowadzi do zamierzonych rezultatów. Logika interwencji natomiast obejmuje celowy i sekwencyjny proces przechodzenia od działań programu do jego rezultatów przebiegający w ustalonym czasie. Składają się na nią zasoby, planowane działania, ich produkty i rezultaty identyfikowane jako kamienie milowe modelu logicznego interwencji. Por. A. Markiewicz, I. Patrick, *Developing Monitoring and Evaluation Frameworks*, Sage Publications Inc. 2016, s. 74.

Rysunek 1. Teoria programu i logika interwencji – przykład dla regionalnego programu zwiększenia udziału dzieci w zajęciach lekcyjnych



Źródło: A. Markiewicz, I. Patrick, Developing Monitoring and Evaluation Frameworks, Sage Publications Inc. 2016, s. 86.

się, iż projekt interwencji jest wypadkową przekonań i wpływu poszczególnych osób uczestniczących w jego formułowaniu – także tych, które nie zostały wyartykułowane w trakcie jego budowania, natomiast znalazły w nim swój faktyczny wyraz. Z kolei ewaluacja logiki interwencji polega na ustaleniu z jednej strony przebiegu procesu realizacji programu, a z drugiej prawidłowości związków przyczynowo-skutkowych uwzględnionych w projekcie interwencji<sup>227</sup>.

W ten sposób, badając logikę programu, wspiera się budowanie doskonalszych programów – opartych na faktycznie działających mechanizmach zmiany oraz jej wdrażania – i poprawia się zarządzanie nimi. Takie podejście zgodne jest jednocześnie z zasadą *evidence-based policy*.

Syntezę korzyści z TBE dla ewaluacji na poziomie regionalnym przedstawiono w poniższej tabeli.

**Tabela 1. Znaczenie TBE dla ewaluacji na poziomie regionalnym**

Rodzaj ewaluacji	Składowa TBE	
	Teoria programu	Teoria zmiany (logika interwencji)
Ewaluacja <i>ex ante</i>	<p>Odtworzenie teorii programu</p> <ul style="list-style-type: none"> <li>• Identyfikacja założeń (w tym ukrytych) programu</li> <li>• Ocena prawidłowości przyjętych założeń w świetle stanu wiedzy odnoszącej się do danego obszaru interwencji, z uwzględnieniem teorii naukowych, doświadczenia innych regionów (benchmarków) oraz specyfiki regionalnej (kontekstu)</li> <li>• Weryfikacja spójności założeń w kontekście celów i uwarunkowań programu</li> </ul>	<p>Odtworzenie teorii zmiany</p> <ul style="list-style-type: none"> <li>• Identyfikacja związków przyczynowskutkowych pomiędzy składowymi programu interwencji (mechanizmów), które mają prowadzić do osiągnięcia założonych rezultatów</li> <li>• Ocena prawidłowości przyjętych założeń w świetle wiedzy, doświadczeń i regionalnego kontekstu</li> <li>• Weryfikacja wzajemnej spójności wszystkich planowanych do uruchomienia mechanizmów interwencji</li> <li>• Weryfikacja adekwatności planowanych działań instytucji wdrażających interwencję względem projektowanych rezultatów oraz kontekstu regionalnego</li> </ul>

<sup>227</sup> Tamże.

Rodzaj ewaluacji	Składowa TBE	
	Teoria programu	Teoria zmiany (logika interwencji)
Ewaluacja <i>on-going</i> / <i>ex post</i>	<p>Powiązanie teorii programu z osiągniętymi wynikami</p> <ul style="list-style-type: none"> <li>• Identyfikacja teorii programu oraz ścieżki jej (ewentualnych) zmian w trakcie realizacji interwencji</li> <li>• Ocena stopnia potwierdzenia teorii programu w świetle uzyskanych wyników (produkty, rezultaty, wpływ)</li> <li>• Identyfikacja uwarunkowań skuteczności teorii programu z uwzględnieniem wiedzy, umiejętności i zachowań wszystkich podmiotów powiązanych z interwencją na każdym jej etapie (projektowanie, realizacja, ocena) oraz kontekstu regionalnego</li> <li>• Ocena przydatności składowych teorii programu dla kolejnych interwencji oraz zasadności kontynuowania i/lub modyfikacji interwencji (ewaluacja <i>on-going</i>) bądź jej powtarzania (ewaluacja <i>ex post</i>).</li> </ul>	<p>Powiązanie teorii zmiany (logiki interwencji) z osiągniętymi wynikami</p> <ul style="list-style-type: none"> <li>• Identyfikacja logiki interwencji oraz ścieżki jej (ewentualnych) zmian w trakcie realizacji</li> <li>• Ocena stopnia potwierdzenia skuteczności uruchomionych mechanizmów interwencji</li> <li>• Identyfikacja uwarunkowań skuteczności mechanizmów interwencji z uwzględnieniem świadomości, postaw i zachowań wszystkich podmiotów powiązanych z realizacją interwencji oraz kontekstu regionalnego</li> <li>• Ocena przydatności poszczególnych mechanizmów dla kolejnych interwencji oraz zasadności kontynuowania i/lub modyfikacji interwencji (ewaluacja <i>on-going</i>) bądź jej powtarzania (ewaluacja <i>ex post</i>).</li> </ul>

Źródło: Opracowanie własne.

W myśl realistycznego podejścia do ewaluacji opartej na teorii (*realistic evaluation*) do optyki dotychczasowego projektowania interwencji publicznych wprowadza się kontekst jako istotny czynnik wpływający na skuteczność uruchamianych teorii zmiany, obejmujący otoczenie, kulturę i zachowania administracji publicznej. Właśnie uwzględnienie kontekstu interwencji publicznych wydaje się szczególnie cennym wkładem tego podejścia w doskonalenie jakości Regionalnych Programów Operacyjnych (RPO). Zwrócenie uwagi na specyficzne dla danego regionu uwarunkowania (np. kulturę lokalną, kapitał społeczny, kulturę organizacyjną administracji), które mogą w istotny sposób wpłynąć na rezultaty osiąmane poprzez uruchomiony mechanizm zmiany, pozwala trafniej oceniać wyniki interwencji.

Związany z istotą ewaluacji zorientowanych na teorię brak możliwości pełnego porównania wyników polityk publicznych realizowanych w różnych sytuacjach (ograniczony potencjał uogólnień) może stać się ich główną zaletą. Wymuszają one porównywanie różnych sytuacji interwencji (np. identycznych instrumentów wsparcia realizowanych w różnych regionach), tak by zrozumieć, co i dlaczego pasuje do jednej sytuacji, ale nie sprawdza się wszędzie. Prowadzi to do głębszego zrozumienia uwarunkowań mechanizmów interwencji (zwłaszcza czynników lokalnych) oraz w efekcie do zwiększenia elastyczności adaptacji polityk budowanych na podstawie tzw. dobrych praktyk i większej ich adekwatności względem potrzeb.

Zarazem jednak wykorzystanie możliwości ewaluacji zorientowanych na teorię może stanowić wyzwanie dla regionów w Polsce: ze względu na wciąż niedostateczną wiedzę na temat TBE, stosunkowo ograniczone doświadczenia w jej prowadzeniu, zarówno po stronie zamawiających, jak i wykonawców, oraz ograniczenia organizacyjne i finansowe jednostek ewaluacyjnych umiejscowionych w urzędach marszałkowskich. Zbudowanie u decydentów świadomości specyfiki regionalnej i koniecznych adaptacji projektowanych interwencji wymaga czasu zarówno po stronie wykonawcy, jak też zamawiającego. Istotne są tu również czynniki związane z kulturą organizacyjną zamawiającego i jego wolą docierania do faktycznych założeń, często nieświadomych, które stoją za wyborami konkretnych działań do projektowanych interwencji. TBE ujawni swój potencjał i będzie rzeczywiście przydatna, gdy posłuży do weryfikowania kierunków działania oraz identyfikowania najskuteczniejszych w kontekście danego regionu mechanizmów rozwiązywania problemów. Wymaga to także rzeczywistego zainteresowania wynikami ewaluacji ze strony wszystkich uczestników interwencji – od decydentów na najwyższym szczeblu poczynając, na beneficjentach kończąc. Zainteresowania, które wynika z chęci pogłębienia wiedzy także na swój temat oraz woli eliminowania ograniczeń indywidualnych i zbiorowych procesów decyzyjnych.

### **Praktyka ewaluacji opartej na teorii na poziomie regionalnym (ze szczególnym uwzględnieniem badań *ex ante* RPO 2014-2020)**

Włączenie podejścia TBE do ewaluacji *ex ante* pozwoliło zgodnie z założeniami na doskonalenie programu operacyjnego na etapie formułowania jego koncepcji. Poprawiony układ związków przyczynowo-skutkowych pomiędzy potrzebami, celami, zasobami oraz działaniami podejmowanymi w ramach interwencji publicznej może zwiększyć szanse na jej efektywną realizację. Zarazem uporządkowanie logiki programu ułatwia prowadzenie ewaluacji *on-going* oraz *ex post*, ponieważ wyraźnie wskazuje, jakie elementy uznane zostały w procesie programowania za kluczowe dla osiągnięcia zamierzonych celów. Ocena zmian tych elementów w trakcie i po wdrażaniu interwencji może pozwolić na zweryfikowanie prawidłowości całej koncepcji działania i potwierdzić bądź wykluczyć zasadność jej stosowania w przyszłości.

Z racji wprowadzonego wymogu realizowania ewaluacji *ex ante* dla projektów regionalnych programów operacyjnych na lata 2014-2020 (art. 55 Rozporządzenia Rady nr 1083/2006<sup>228</sup>) w taki sposób, by dla wszystkich RPO odtworzyć i ocenić ich modele logiczne (teorie zmiany), właśnie raporty z ewaluacji programów regionalnych ujawniają wiedzę o gotowości do stosowania TBE w regionach. Głównym celem badań *ex ante* projektów programów operacyjnych była kompleksowa ocena trafności i spójności (wewnętrznej i zewnętrznej) RPO oraz systemu realizacji programu, w tym potencjału ludzkiego i administracyjnego, a także systemu monitoringu. Weryfikacja spójności zewnętrznej wymagała ustalenia stopnia zgodności RPO z europejskimi, krajowymi i regionalnymi strategiami oraz z politykami horyzontalnymi. Przed wykonawcami postawiono również zadanie wymagające wiedzy i doświadczenia, bowiem mieli sformułować wnioski i rekomendacje w zakresie zmian w programie operacyjnym.

Istotną trudnością przy przeprowadzaniu ewaluacji *ex ante* programów przygotowywanych na lata 2014-2020 były trwające negocjacje ram prawnych nowej perspektywy, m.in. umowy partnerstwa, na poziomie krajowym. Powodowało to w wielu przypadkach ewaluowanie projektów RPO, a nie ostatecznych dokumentów, i zwiększało niepewność decyzyjną, zarówno po stronie zamawiającego, jak i wykonawcy. W efekcie logika interwencji w toku prac nad programami była poddawana istotnym zmianom polegającym na przeformułowaniu celów szczegółowych, a ewaluacje były realizowane jako proces partycypacyjny, w którym uczestniczyły zarówno zespoły jednostek ewaluacyjnych, eksperci zewnętrzni, jak również pracownicy Instytucji Zarządzającej przygotowujący projekt RPO 2014-2020. Partycypacja oznaczała równoległe prowadzone procesy programowania i ewaluacji z wykorzystaniem wiedzy eksperckiej zespołu wykonawców, jednak zmiany modelu (logiki) interwencji mogły być wprowadzane pod wpływem innych czynników, a nie tylko wniosków płynących z badania ewaluacyjnego.

Ocena logiki interwencji i proponowanych form wsparcia dokonywana była według osi priorytetowych, natomiast w żadnej ewaluacji *ex ante* dla poszczególnych RPO nie dokonano całościowego odtworzenia logiki programu.

Dokonując szczegółowego przeglądu 15 raportów końcowych ewaluacji *ex ante* RPO, można sformułować kilka najważniejszych wniosków:

- Część wykonawców ewaluacji jest przygotowana do realizacji formalnie poprawnej ewaluacji opartej na teorii. Oznacza to, że zespoły ewaluatorów prawidłowo

---

<sup>228</sup> Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1303/2013 z dnia 17 grudnia 2013 r. ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności, Europejskiego Funduszu Rolnego na rzecz Rozwoju Obszarów Wiejskich oraz Europejskiego Funduszu Morskiego i Rybackiego oraz ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności i Europejskiego Funduszu Morskiego i Rybackiego oraz uchylające rozporządzenie Rady (WE) nr 1083/2006.



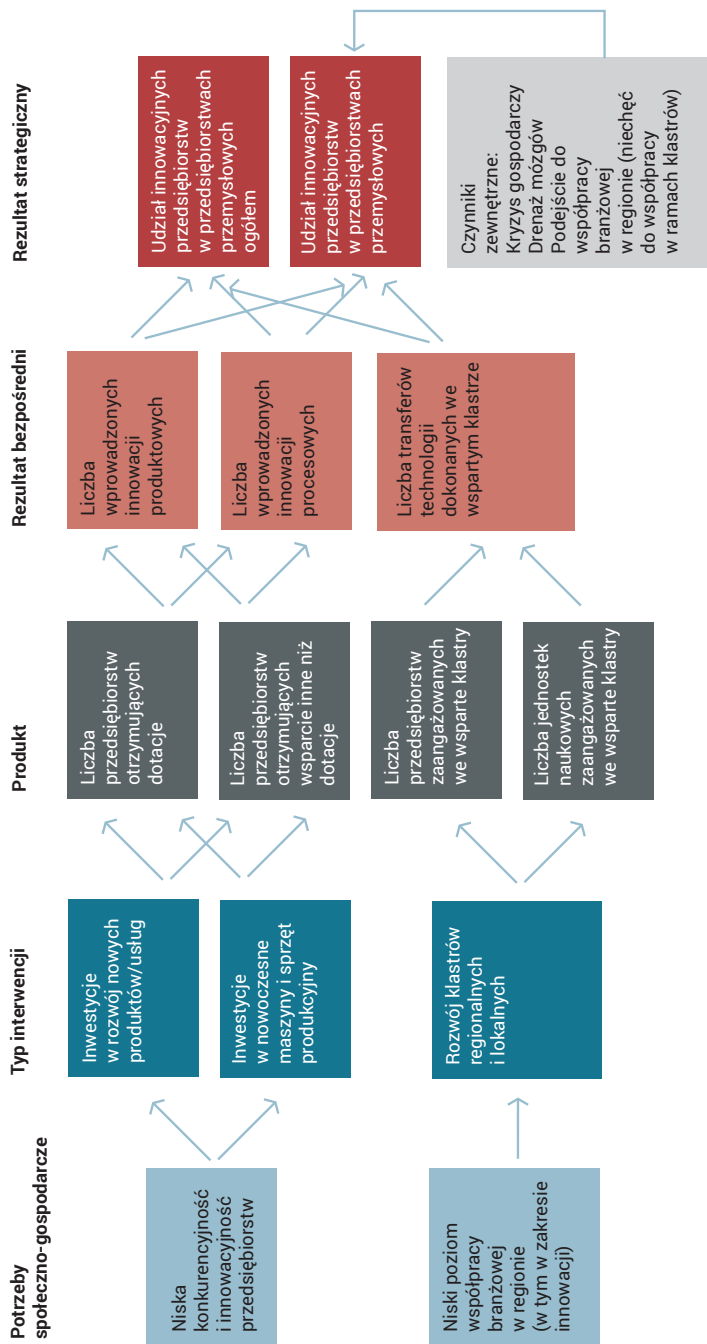
odtworzą zapisane w dokumentach programowych (RPO) związki przyczynowo-skutkowe oraz dokonują ich szerszej lub węższej oceny, jednak głównie w kontekście problemów zapisanych w diagnozie – bez weryfikacji przyczyn dokonywanej przez decydentów hierarchizacji i doboru problemów do rozwiązania.

- Występują dwa rozbieżne podejścia do zakresu zadań realizowanych przez ewaluatora: z jednej strony wykonawcy wykazują się tendencją do przejmowania ciężaru ustalenia teorii bazowych, zasilenia zasobów informacyjnych oraz budowania modeli logiki interwencji dla ocenianego RPO, z drugiej pojawiają się tendencje do realizacji zadań w trybie czysto formalnej weryfikacji braków występujących w ocenianym materiale. Wówczas ewaluator ogranicza się do stwierdzenia, iż niemożliwa jest ocena logiki interwencji w przypadku którejś z analizowanych osi. Zarówno jedno, jak i drugie podejście ogranicza formatywny charakter każdej ewaluacji, od którego zależy uruchomienie w regionie procesów pogłębiania wiedzy o mechanizmach jego rozwoju.
- Raporty końcowe z ewaluacji RPO świadczą o sporych różnicach w rozumieniu istoty teorii programu oraz logiki interwencji. Tam, gdzie ta ostatnia jest odtwarzana, jako uzasadnienie spójności programu podaje się zgodność z diagnozą regionalną. W niektórych ewaluacjach rekonstrukcję logiki interwencji sprowadzono do powielenia zapisów RPO z pominięciem faktycznych związków przyczynowo-skutkowych. Inne badania ujawniają brak powiązania celów i działań z poszczególnymi typami rezultatów oraz uwarunkowaniami RPO. Jeśli uwzględnia się w modelach logiki interwencji ich uwarunkowania, to zazwyczaj chodzi o czynniki zewnętrzne z pominięciem kontekstu wewnętrznego. Wreszcie logika interwencji bywa sprowadzana do trafności i spójności planowanych działań.
- Logika interwencji zazwyczaj ilustrowana jest diagramami przyczynowo-skutkowymi. Obok tego jednak pojawia się opis w tekście lub w formie tabeli, co jest mniej czytelne od prezentacji graficznej i może prowadzić do niezrozumienia się stron.
- Obok tego jednak dostępne ewaluacje *ex ante* wskazują na dobre przygotowanie rzeczowe wykonawców, którzy są w stanie zidentyfikować usterki ocenianych RPO i sformułować rekomendacje pozwalające na ich poprawienie.
- Powtarzanie się relatywnie niewielkiej liczby wykonawców pozwala także dostrzec procesy upowszechniania się wiedzy o niezbędnych elementach TBE (np. o przydatności prezentowania logiki interwencji w formie diagramów), a także rosnącą kreatywność zespołów ewaluacyjnych. Tworzą oni interesujące zestawienia, które zwiększają ładunek informacyjny raportów. Należy jednak zaznaczyć, że tabele i schematy blokowe nie są najlepszą formą opisu odtworzonej logiki programu, bowiem nie ujawniają związków przyczynowo-skutkowych prowadzących do określonych rezultatów.
- Stosunkowo słabo na tym tle prezentuje się także sposób uzasadniania doboru metod do badania. Oczywiście zalety niektórych z nich (IDI, FGI) nie są eksploatawalne do zrozumienia przesłanek kierujących decyzjami twórców interwencji. General-

nie wydaje się, że aspekt ludzki w badaniu ewaluacyjnym *ex ante* RPO 2014-2020 w dużym stopniu został pominięty, co stoi w sprzeczności z istotą TBE. Pominięcie obok aspektu ludzkiego także teorii wdrożenia należy uznać również za istotne ograniczenie.

Analizując szczegółowo raport z badania *ex ante* RPO województwa opolskiego na lata 2014-2020, dostrzec można, że zastosowanie TBE na poziomie regionalnym napotykało na pewne trudności. Na poniższym rysunku zaprezentowano jeden z grafów przygotowanych dla osi priorytetowej PI 3.3 „Wspieranie tworzenia i rozszerzania zaawansowanych zdolności w zakresie rozwoju produktów i usług”, wspartej w ramach RPO WO 2014-2020, który w sposób bardzo ogólny opisuje rekomendowaną logikę interwencji (zob. rys. 2.).

**Rysunek 2. Diagram przyczynowo-skutkowy opisujący rekomendowaną logikę interwencji dla osi priorytetowej PI 3.3 „Wspieranie tworzenia i rozszerzania zaawansowanych zdolności w zakresie rozwoju produktów i usług”, uwzględnionej w ramach RPO województwa opolskiego 2014-2020.**



Źródło: Załącznik nr 3 do raportu końcowego „Ewaluacja ex ante programu operacyjnego województwa opolskiego na lata 2014-2020”, Urząd Marszałkowski Województwa Opolskiego, Opole 2014 r.

Należy zaznaczyć, iż logika interwencji (teoria zmiany) przedstawiona na diagramie wskazuje na przyjęte przez projektodawcę mechanizmy osiągnięcia rezultatów w istniejących, rozpoznanych przed planowaniem interwencji warunkach otoczenia. Wydaje się jednak, że zadaniem ewaluatora powinno być – zwłaszcza na etapie oceny *ex ante* programu, prowadzonej metodą partycypacyjną – wsparcie procesu planowania interwencji poprzez konstruktywne wskazania odnoszące się do ewentualnych niespójności lub niezgodności logiki interwencji (np. uspoźnienie i uzupełnienie luk w logice interwencji). Rekomendacje odnoszące się do błędów formalnych w przygotowanych dokumentach, niespójności z dokumentami wyższego rzędu są najprostsze do sformułowania i najczęściej stosowane. Natomiast propozycje zmian odnoszące się do modyfikacji zaprojektowanych instrumentów (działań), które mają uruchomić efektywny mechanizm zmiany, stanowią większe wyzwanie dla ewaluatora, bowiem wymagają wiedzy na temat skuteczności różnych rozwiązań i ich uwarunkowań. Biorąc pod uwagę relatywnie krótkie terminy realizacji badań ewaluacyjnych i równoległy proces programowania, podporządkowany odrębnemu reżimowi terminów, w analizowanych ewaluacjach RPO nadal w sferze wyzwań pozostawała funkcja doradcza ewaluatora. Jej przejawem mogłyby być formułowane na bieżąco eksperckie propozycje zmian (robocze rekomendacje na potrzeby procesu programowania) odnoszące się do wad i braków wykrytych w badanym programie. Do najtrudniejszych należy rekomendowanie korekt w obszarze zaprojektowanych założeń teorii zmiany programu. Ich formułowanie jest bowiem możliwe w oparciu o rozległą wiedzę ekspercką. Z kolei ograniczeniem wielu zespołów wykonujących ewaluację na poziomie regionalnym pozostaje dostęp do tej wiedzy (tj. brak woli lub możliwości pozyskiwania odpowiednich ekspertów).

Należy zatem stwierdzić, że praktyka polskiej ewaluacji na poziomie regionalnym pokazuje stosunkowo rzadkie przypadki rzetelnego odtwarzania logiki projektu programu na etapie przygotowywania interwencji publicznych. W wielu przypadkach, gdy na etapie tworzenia koncepcji interwencji formułowana jest logika programu, w procesie decyzyjnym podlega ona bardzo często modyfikacjom (szczególnie na etapie formułowania systemu i kryteriów wyboru projektów) i ostatecznie może ulec rozmyciu. W konsekwencji utrudnia to późniejszą prawidłową realizację procesu ewaluacyjnego, dla którego głównym celem jest zweryfikowanie sprawności mechanizmu prowadzącego do osiągnięcia rezultatów.

Poza zestawem ewaluacji *ex ante* RPO wśród najnowszych raportów końcowych dostępnych na stronach internetowych Ministerstwa Rozwoju są jednak badania warte wskazania jako odpowiadające na istotne wymogi TBE. Pod względem dopracowania koncepcji badania ewaluacyjnego, a także spełnienia kryteriów identyfikacji i opisu teorii bazowych ocenianej interwencji, można na przykład wskazać badanie

pn. „Ocena stosowania zasady równości szans na wszystkich etapach realizacji RPO WP na lata 2007-2013”<sup>229</sup>.

Niemniej należy podkreślić, że nowe podejście do ewaluacji pozostaje dużym wyzwaniem dla wszystkich regionalnych jednostek ewaluacyjnych. Jednym z ograniczeń w uzyskaniu wysokiej jakości wyników badań, m.in. stosujących podejście TBE, są stosunkowo niskie budżety zaplanowane w regionach na realizację badań ewaluacyjnych. Te województwa, w których środki w ramach pomocy technicznej RPO 2014-2020 są stosunkowo skromne, posiadają także relatywnie niewielkie budżety na ewaluację. Regionom tym trudno jest zaplanować, a następnie realizować badania ewaluacyjne uwzględniające „rygorystyczne metodologicznie” wytyczne KE oraz zalecenia krajowe.

Kolejnym bardzo ważnym aspektem, który ma znaczenie dla ewaluacji w polskich regionach, jest kultura organizacyjna administracji samorządowej. Problemy organizacyjne wewnątrz instytucji (np. w zakresie obowiązków osób zajmujących się ewaluacją znajdują się także liczne inne zadania) utrudniają znacząco wprowadzanie innowacji (korzystanie z ewaluacji jako narzędzia wspierającego projektowanie polityk publicznych w regionie nadal pozostaje w sferze innowacji). W regionach z poważnymi wyzwaniami spotykają się te JE, w których wystąpiła duża fluktuacja kadr lub znaczące ograniczenie liczby osób zajmujących się ewaluacją. Dodatkowym elementem wpływającym na jakość wyników badań ewaluacyjnych realizowanych z wykorzystaniem podejścia TBE jest ograniczona współpraca pomiędzy JE a działami odpowiedzialnymi za projektowanie interwencji publicznych. TBE wymaga bowiem ścisłej współpracy JE i ewaluatora z osobami odpowiedzialnymi za tworzenie programów, uwzględnienia specyfiki kontekstu, poszukiwania dowodów na to, że związki przyczynowo-skutkowe zaistniały lub mają realną szansę zaistnieć.

Istotnym wyzwaniem jest również sfera metodyczna ewaluacji, bowiem przy całym zapleczu informacyjnym i edukacyjnym oferowanym przez Krajową Jednostkę Ewaluacji, PARP, Polskie Towarzystwo Ewaluacyjne i uczelnie wyższe, które rozwijają potencjał ewaluacyjny w Polsce (np. Centrum Europejskich Studiów Regionalnych i Lokalnych EUROREG, Uniwersytet Jagielloński, Akademia Leona Koźmińskiego), podejście TBE jest stosunkowo rzadko analizowane w krajowej literaturze przedmiotu (jak dotąd wyjątek stanowi niniejsza publikacja) i uwzględniane w procesie dydaktycznym dotyczącym ewaluacji. Skutkuje to również względnie rzadką praktyką TBE na poziomie regionalnym, a jeśli ona już występuje (np. w ewaluacji *ex ante* RPO 2014-2020 w wyniku odgórnych zaleceń), jej zakres i podejście metodyczne oraz realna pomoc w procesie programowania pozostawiają wiele do życzenia.

---

<sup>229</sup> [https://www.wrotapodlasia.pl/pl/fundusze\\_europejskie/fundusze\\_europejskie/ocena-stosowania-zasady-rownosci-szans-na-wszystkich-etapach-realizacji-rpowa-na-lata-2007-2013.html](https://www.wrotapodlasia.pl/pl/fundusze_europejskie/fundusze_europejskie/ocena-stosowania-zasady-rownosci-szans-na-wszystkich-etapach-realizacji-rpowa-na-lata-2007-2013.html) (dostęp 10.05.2017 r.).

Systematyczne i przekrojowe badania ewaluacyjne, uwzględniające nie tylko stopień rozwiązania problemów regionu, ale również uzupełnianie luk wiedzy o kontekście i jego wpływie na mechanizmy zmian, mogą w dłuższym okresie pogłębić rozpoznanie krytycznych czynników rozwojowych regionu i parametrów interwencji, a w efekcie pozwolić na bardziej adekwatne dopasowywanie programów do warunków lokalnych. Identyfikacja zarówno skutecznych, jak i nieskutecznych interwencji z pogłębionym wyjaśnieniem przyczyn sukcesów i niepowodzeń może posłużyć wyeliminowaniu tych ostatnich z puli działań podejmowanych przez regionalne administracje<sup>230</sup>. Ważny w tym kontekście jest jednak sposób realizacji badania ewaluacyjnego. Dla ewaluacji *ex ante*, gdy wykorzystywane jest TBE, szczególnie przydatne jest podejście partycypacyjne. Dzięki partycypacji, czyli włączeniu w proces identyfikowania logiki interwencji podmiotów i osób planujących polityki publiczne, ewaluator ma możliwość głębszego zrozumienia lokalnego kontekstu interwencji, założeń formułowanych przez poszczególnych aktorów oraz bieżącego wyjaśniania związków przyczynowo-skutkowych, jakie zostały wpisane w badany program. W efekcie możliwe jest rozpoznanie, które elementy interwencji wymagają uzasadnienia lub modyfikacji (brakujące lub słabe ogniwa logiki interwencji), co może przełożyć się na doskonalenie planu, metodyki i efektów zarówno interwencji, jak i samej ewaluacji. Zastosowanie partycypacji pozwala również na skrócenie czasu programowania interwencji poprzez równoległe prowadzenie badań ewaluacyjnych *ex ante* i włączanie rekomendowanych zmian wynikających z tych badań w kolejne obszary projektowania logiki interwencji.

## Podsumowanie

W perspektywie finansowej 2007-2013 zbudowano w województwach podwaliny systemu ewaluacji polityki spójności UE realizowanej w Polsce na poziomie regionalnym. Pozostaje jednak pytanie, czy w kontekście nowych wyzwań oraz zadań, jakie są stawiane przed województwami (m.in. realizacja badań typu TBE), potencjał ten jest wystarczający. Wymagania KE miały doprowadzić do wzrostu praktyki wykorzystywania ewaluacji opartej na teorii. Wydaje się, że dostosowanie się do tych wymagań będzie jednak dość trudne.

Mając świadomość, jak dużym wyzwaniem dla regionów jest dostosowanie się do nowych potrzeb ewaluacyjnych i że proces budowy potencjału ewaluacyjnego w Polsce wchodzi na kolejny, bardziej zaawansowany poziom rozwoju, należy pamiętać, że największym atutem w tym wszystkim powinno być partycypacyjne prowadzenie ewaluacji. Bieżąca współpraca wykonawcy (ewaluatora) z zamawiającym, w tym z osobami, które odpowiadają (odpowiadały) za programowanie danej interwencji, przynieść może bowiem największą wartość dodaną z badania (szczególnie w przypadku prowadzenia ewaluacji w podejściu opartym na teorii).

---

<sup>230</sup> Szerzej na temat trzech podstawowych składników budujących założenia interwencji publicznych: teorii bazowych, teorii zmiany oraz teorii wdrażania, w: K. Olejniczak, „Ewaluacja w sektorze publicznym”, w: *Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka*, praca zbiorowa, Urząd Marszałkowski Województwa Opolskiego, Opole 2009, s. 50.

## O autorkach

### **Anna Bruska**

---

Doktor nauk ekonomicznych, pracuje w Zakładzie Logistyki i Marketingu na Wydziale Ekonomicznym Uniwersytetu Opolskiego. Certyfikowany tutor *Collegium Wratislaviense* oraz trener biznesu. Współpracuje z jednostkami samorządu terytorialnego, m.in. jako ekspert zewnętrzny przy realizacji badań ewaluacyjnych finansowanych ze środków unijnych i przy opracowywaniu planu ewaluacji. Zainteresowania naukowe obejmują m.in. uwarunkowania poznawcze i metodyczne jakości badań, w tym również badań ewaluacyjnych.

### **Katarzyna Lotko-Czech**

---

Doktor nauk ekonomicznych, absolwentka Uniwersytetu Opolskiego oraz podyplomowych studiów z ewaluacji programów społecznych na Uniwersytecie Warszawskim. W latach 1999- 2014 pracowała w Urzędzie Marszałkowskim Województwa Opolskiego, w którym w 2007 r. od podstaw tworzyła Jednostkę Ewaluacyjną, odpowiedzialną za koordynowanie procesu ewaluacji programów operacyjnych na poziomie regionalnym. Od 2015 r. jest ekspertem Fundacji IDEA Rozwoju, gdzie projektuje i prowadzi badania ewaluacyjne, w szczególności dla administracji publicznej. Autorka artykułów naukowych m.in. z zakresu rozwoju regionalnego, polityki spójności Unii Europejskiej, finansów publicznych. Prelegentka konferencji i seminariów naukowych, dotyczących polityki regionalnej oraz budowy kultury ewaluacyjnej na poziomie regionalnym. Prowadzi wykłady z zakresu polityki regionalnej w Wyższej Szkole Bankowej we Wrocławiu (Wydział Ekonomiczny w Opolu). Członek Polskiego Towarzystwa Ekonomicznego, założycielka i prezes zarządu spółki CG2.

## Bibliografia:

1. Bachtler J., „Ewaluacja regionalnej polityki w Europie: kultura, zaangażowanie i potencjał”, w: K. Olejniczak, M. Kozak, B. Ledzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, Wydawnictwa Akademickie i Profesjonalne, Akademia Leona Koźmińskiego, Warszawa 2008.
2. Bober J., „Formalnoprawne podstawy Unii Europejskiej w zakresie ewaluacji”, w: S. Mazur (red.), *Ewaluacja funduszy strukturalnych – perspektywa regionalna*, Uniwersytet Ekonomiczny w Krakowie, Małopolska Szkoła Administracji Publicznej, Kraków 2007.
3. Górniak J., „Szacowanie przyczynowego wpływu interwencji: funkcje, walory, ograniczenia oraz relacje z innymi podejściami ewaluacji”, w: *Ewaluacja RPO WK-P. Nowe wyzwania dla ewaluacji programów operacyjnych w perspektywie 2014-2020*, Urząd Marszałkowski Województwa Kujawsko-Pomorskiego, 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*, Centrum Europejskich Studiów Regionalnych i Lokalnych UW, Wydawnictwo Naukowe „Scholar”, Warszawa 2008.
6. Olejniczak K., „Teoretyczne podstawy ewaluacji ex post”, w: *Ewaluacja ex post. Teoria i praktyka badawcza*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa 2007.
7. Stame N., „Theory-based Evaluation and Types of Complexity”, *Evaluation* 2004, vol. 10 (1).
8. Szałaj M., „Budowa potencjału ewaluacyjnego”, w: *Ewaluacja Programów Operacyjnych na poziomie regionalnym – teoria i praktyka*, Urząd Marszałkowski Województwa Opolskiego, Opole 2009.
9. Żuber P., Bienias S., „System ewaluacji w Polsce – dotychczasowe doświadczenia i wyzwania na przyszłość”, w: K. Olejniczak, M. Kozak, B. Ledzion (red.), *Teoria i praktyka ewaluacji interwencji publicznych. Podręcznik akademicki*, Wydawnictwa Akademickie i Profesjonalne, Akademia Leona Koźmińskiego, Warszawa 2008.

## Inne:

1. Wytyczne nr 6 w zakresie ewaluacji programów operacyjnych na lata 2007-2013, Ministerstwo Rozwoju Regionalnego, Warszawa, 30 maja 2007 r.
2. *Rocznik Ewaluacji 2013*, Ministerstwo Rozwoju, Warszawa 2014.
3. *Rocznik Ewaluacji 2015*, Ministerstwo Rozwoju, Warszawa 2016
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](http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf) (dostęp 19.03.2017 r.).



5. *Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020*, Minister Infrastruktury i Rozwoju, Warszawa, 22 września 2015, podrozdział 1.2 pkt 4. [http://www.funduszeuropejskie.gov.pl/media/9318/wytyczne\\_090915\\_final.pdf](http://www.funduszeuropejskie.gov.pl/media/9318/wytyczne_090915_final.pdf) (dostęp 19.03.2017 r.).
6. Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1303/2013 z dnia 17 grudnia 2013 r. ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności, Europejskiego Funduszu Rolnego na rzecz Rozwoju Obszarów Wiejskich oraz Europejskiego Funduszu Morskiego i Rybackiego oraz ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności i Europejskiego Funduszu Morskiego i Rybackiego oraz uchylające rozporządzenie Rady (WE) nr 1083/2006.
7. Zalecenia w zakresie ewaluacji *ex ante* programów operacyjnych na lata 2014-2020, Departament Koordynacji Polityki Strukturalnej, Ministerstwo Rozwoju Regionalnego, Warszawa, listopad 2012 r.

**Polska Agencja  
Rozwoju Przedsiębiorczości**

ul. Pańska 81/83  
00-834 Warszawa

telefon: 22 432 80 80  
fax: 22 432 86 20  
e-mail: [biuro@parp.gov.pl](mailto:biuro@parp.gov.pl)

**Informatorium PARP**

telefon: 22 432 89 91-93  
e-mail: [info@parp.gov.pl](mailto:info@parp.gov.pl)

[www.parp.gov.pl](http://www.parp.gov.pl)



---

[www.parp.gov.pl](http://www.parp.gov.pl)